UNIVERSIDADE FEDERAL DE PELOTAS

POSTGRADUATE PROGRAM IN EPIDEMIOLOGY

Geospatial analyses of health indicators using national health surveys from low- and middle-income countries:

Analyses of reproductive, maternal, newborn and child health

PhD Thesis

LEONARDO ZANINI FERREIRA

Supervisor: Aluísio J D Barros

Co-supervisor: Fernando Hartwig

2022

UNIVERSIDADE FEDERAL DE PELOTAS

POSTGRADUATE PROGRAM IN EPIDEMIOLOGY

Geospatial analyses of health indicators using national health surveys from low- and middle-income countries:

Analyses of reproductive, maternal, newborn and child health

LEONARDO ZANINI FERREIRA

Supervisor: Aluísio J D Barros

Co-supervisor: Fernando Hartwig

The thesis was presented to the Postgraduate Program in Epidemiology, Federal University of Pelotas, Brazil to complete the requirements of a PhD in Epidemiology.

2022.

F383g Ferreira, Leonardo Zanini

Geospatial analyses of health indicators using national health surveys from low- and income countries: Analyses of reproductive, maternal, newborn and child health. / Leonardo Zanini Ferreira ; Aluísio Jardim Dornellas de Barros, Orientador; Fernando Hartwig, Coorientador. – Pelotas : Universidade Federal de Pelotas, 2022.

131 f. : il.

Tese (Doutorado) – Programa de Pós-Graduação em Epidemiologia, Faculdade de Medicina, Universidade Federal de Pelotas, 2022.

1. Epidemiologia. 2. Saúde materno-infantil. 3. Sistemas de informação geográfica I. Barros, Aluísio Jardim Dornellas, orient. II. Hartwig, Fernando, coorient. III. Título.

CDD 614.4

Ficha catalográfica: M. Fátima S. Maia CRB 10/1347

Leonardo Zanini Ferreira

Geospatial analyses of health indicators using national health surveys from low- and middle-income countries:

Analyses of reproductive, maternal, newborn and child health

Examiner's board:

Professor Aluísio J D Barros (supervisor)

Universidade Federal de Pelotas

Associate Professor Fernando César Wehrmeister (examiner)

Universidade Federal de Pelotas

Assistant Professor Inácio Crochemore Mohnsam da Silva (examiner)

Universidade Federal de Pelotas

Christovam de Castro Barcellos Neto, PhD (examiner)

Fundação Oswaldo Cruz

Acknowledgements

I would like to begin this section thanking and dedicating this accomplishment to my **family**, in particular to my **parents**. I have only gotten this far because I always received full support to seek any objectives I have set and for that I am very privileged and grateful. In special, I want to dedicate this degree to the one person I miss the most and would really like to have in this moment with me: grandma **Ilza**, you will always be in my heart.

My unexpected journey began in 2012, thanks to my auntie **Roberta Zanini,** which referred me as a web developer for a project that needed a website. No one would believe if I said that 10 years later that intern kid would be defending his PhD as the eldest member of the group – a group that I am proud and owe so much. I had the privilege to work with incredible people but first and foremost, a huge thanks to professors **Cesar Victora** and **Aluísio Barros.** Two award-winning researchers of the highest caliber that not only shared their tremendous experience, but were always incredibly supportive, and gave me so many opportunities in all these years. Thank you **Andrea Dâmaso**, the first person who took care of me professionally and helped me settle in the group. I want to thank **Maria Clara**, my big sis, you always trusted my work and told me not to wear flip flops. Thank you to all these amazing people I had the pleasure of sharing an office: **Giovanny**, **Fernanda**, **Inácio**, **Fran**, **Carol**, **Jana**. Thanks to dad **Fernando**, since 2013 hearing me whine. Thank you, **Thiago,** for the valuable discussions and assistance with R. And a huge thanks to **Cauane,** who helped me a lot as part of two of my three articles, always optimistic and very committed.

Teacher, supervisor, and boss **Aluísio Barros** deserves a separate acknowledgment as well. A man who always knows something about everything. I am grateful for being mentored by someone I admire and trust, and that have put so much faith in me in all those years.

Thank you to all teachers, colleagues, and staff from the Postgraduate Program in Epidemiology. And to the brothers this place has given me - **Fernando Hartwig** and **Pedro Crespo**. Well, **Fernando**, you not only willingly helped me in all those coffee breaks but also agreed to become my co-supervisor. Thanks, you are someone I always look up to.

Thank you to **Edson Utazi**, **Andrew Tatem** and **Kristine Nilsen** from the University of Southampton for kindly agreeing to teach me such complex methodology. In particular, a huge thanks to **Edson** for going patiently through the codes and guiding me in the development of my analysis.

Thank you to all my friends, especially **Leo Ribeiro** and **Ísis Gonçalves**. I could not have asked for better best friends than you guys.

Thank you to my girlfriend, **Alessandra Lopes**, for everything, being here, comforting me in my despair and making this last (and more painful) year happier.

And to everyone who was somehow part of this journey, cheered for me, helped me along the way or was there to make it easier.

Resumo

Ferreira, Leonardo Z. **Geospatial analyses of health indicators using national health surveys from low- and middle-income countries: Analyses of reproductive, maternal, newborn and child health**. Tese de doutorado. Programa de Pós-graduação em Epidemiologia. Universidade Federal de Pelotas; 2022

Entre os diversos desafios para melhorar a qualidade de vida de mulheres e crianças em países de baixa e média renda, aumentar a qualidade e a disponibilidade de dados é crucial para monitorar o progresso e garantir que os países estejam comprometidos com uma agenda equitativa. De acordo com a meta 17.18 dos Objetivos de Desenvolvimento Sustentável, os países devem fornecer dados confiáveis desagregados por dimensões de desigualdade fundamentais, como localização geográfica. Nesta tese, buscamos investigar o potencial uso de técnicas de modelagem geoespacial como ferramenta para a produção de dados desagregados geograficamente além do que está disponível no desenho amostral de inquéritos domiciliares nacionais. No primeiro artigo, foi realizada uma revisão sistemática para descrever os principais aspectos metodológicos das abordagens geoespaciais em estudos com foco especial em desfechos de saúde reprodutiva, materna, neonatal e infantil (RMNCH). Esse artigo também buscou emponderar leitores não especialistas para que melhor interpretassem os resultados de tais estudos. Identificamos 82 estudos que geraram estimativas para indicadores de RMNCH em resoluções superiores às obtidas diretamente nos inquéritos. A validação do modelo e a incerteza foram significativamente subnotificadas na literatura e a apresentação da incerteza continua sendo um desafio. O segundo artigo implementou as técnicas de modelagem geoespacial para produzir estimativas para o índice de composto cobertura (CCI) no Peru. Estas estimativas foram apresentadas a nível provincial, a segunda divisão administrativa do país, e em malhas de 5 x 5 km ilustrando como as desigualdades geográficas podem ser mascaradas quando se avalia apenas grandes áreas agregadas. O uso do CCI permite uma perspectiva integrada de como está o progresso em direção à cobertura universal de saúde em todo o país. Nós observamos um padrão claro de maior cobertura nas áreas da costa e baixa cobertura no norte e leste do país. As estimativas para as províncias parecem ser suficientes para descrever os padrões de cobertura na maior parte do Peru, mas grandes províncias em áreas de selva podem se beneficiar de estimativas de alta resolução. O último artigo aborda um fenômeno conhecido chamado de problema da unidade de área modificável. Essa questão implica que a interpretação das análises com dados geográficos pode mudar de acordo com a escala ou delimitação das unidades geográficas. Nós realizamos um estudo empírico para quantificar o impacto desse efeito na avaliação das desigualdades geográficas ao longo do tempo. Para isso, geramos quatro medidas complexas de desigualdade em múltiplas resoluções usando modelos geoespaciais em duas pesquisas peruanas como estudo de caso. Descobrimos que a magnitude das desigualdades ao longo do tempo não foi afetada ao comparar anos na mesma resolução, independente da medida de desigualdade utilizada. Além disso, as medidas de desigualdade ponderadas pela população foram menos suscetíveis ao efeito de agregação e apresentaram resultados consistentemente mais estáveis em todas as resoluções avaliadas. No geral, nossas descobertas sugerem que os modelos geoespaciais são recursos úteis para monitorar e rastrear o progresso dos desfechos de RMNCH e de desigualdades a partir de uma perspectiva geográfica e podem ser de grande ajuda para gestores locais e planejadores de políticas para identificar e agir nas áreas mais desfavorecidas de seus países.

Palavras-chave: Saúde da criança; Saúde da mulher, Pesquisas domiciliares, Análise geoespacial, Desigualdades geográficas, Medidas de desigualdade

Abstract

Ferreira, Leonardo Z. **Geospatial analyses of health indicators using national health surveys from low- and middle-income countries: Analyses of reproductive, maternal, newborn and child health**. PhD Thesis. Postgraduate Program in Epidemiology. Universidade Federal de Pelotas; 2022

Among the many challenges for improving the quality of life of women and children in low- and middle-income countries, increasing data quality and availability is crucial to monitor the progress and ensure countries are committed to an equitable agenda. As stated in the target 17.18 of the Sustainable Development Goals, countries must supply reliable data disaggregated by key inequality dimensions such as geographic location. In this thesis, we aimed to investigate the potential use of geospatial modeling techniques as a tool for producing geographically disaggregated data beyond what is available in the sample design of national household surveys. In the first article, a systematic review was carried out to describe key methodological aspects of the geospatial approaches in studies with a special focus on reproductive, maternal, newborn and child health (RMNCH) outcomes. This study also sought to empower non-specialist readers to better interpret the results of such studies. We identified 82 studies that generated estimates for RMNCH indicators at resolutions higher than obtainable directly from the surveys. Model validation and uncertainty were significantly underreported in the literature and the presentation of uncertainty remains a challenge. The second article implemented the geospatial modeling techniques to produce estimates for the composite coverage index (CCI) in Peru. These estimates were presented at provincial level, the second administrative division of the country, and in 5 x 5 km grid-cells describing how geographical inequalities can be masked when looking only at large, aggregated areas. The use of the CCI allows for an integrated perspective on how the progress towards universal health coverage stands throughout the country. We observed a clear pattern of higher coverage in the coastal areas and low coverage in the north and east of the country. Estimates for the provinces seems to be sufficient to describe coverage patterns in the majority of Peru but large provinces in jungle areas can benefit from high resolution estimates. The last article addresses a well-known phenomenon called the modifiable areal unit problem. This issue implies that the interpretation of analyses using geographical data may change according to the scale or delimitation of the geographical units. We carried out an empirical study to quantify the impact of this effect on the assessment of geographic inequalities over time. To do so, we generated four complex measures of inequality at multiple resolutions using geospatial models in two Peruvian surveys as a case study. We found that the magnitude of inequalities over time was not affected when comparing years at the same resolution, regardless of the inequality measure. Furthermore, the population-weighted inequality measures were less susceptible to the aggregation effect and presented consistently more stable results at all estimated resolutions. Overall, our findings suggest that geospatial models are useful resources to monitor and track progress on RMNCH outcomes from a geographical perspective and can be of great assistance to local managers and policy planners to identify and act on the most disadvantaged areas of their countries.

Keywords: Child health; Woman's health, Household surveys, Geospatial analysis, Geographic inequalities, Inequality measures

Table of contents

Presentation

This PhD thesis is one of the requirements to obtain a PhD degree in Epidemiology from the Postgraduate Program in Epidemiology from the Universidade Federal de Pelotas. It was developed under the supervision of Professor Aluísio JD Barros and the co-supervision of Fernando Pires Hartwig, in collaboration with the University of Southampton. Due to the COVID-19 pandemic, the planned period of study abroad was cancelled, and the collaboration was maintained virtually.

The document is structured following the Postgraduate Program in Epidemiology guidelines and is divided in:

1. Research project presented and approved on September 16th of 2019 which includes the suggestions proposed by the examiners Fernando César Wehrmeister and Cesar Gomes Victora.
2. Adjustments that were necessary to the original project during the development of the proposed articles.
3. Report of activities performed during the PhD including the experience in the International Center for Equity in Health, capacity building events, complementary courses and scientific collaborations.
4. Proposed article entitled “Geospatial estimation of reproductive, maternal, newborn and child health indicators: a systematic review of methodological aspects of studies based on household surveys”, published at International Journal of Health Geographics.
5. Proposed article entitled “Geospatial modeling of the composite coverage index in Peru”, to be submitted to International Journal of Epidemiology.
6. Proposed article entitled “Measuring time trends in geographic health inequalities at different resolutions: the scale effect”, to be submitted to International Journal of Equity in Health.
7. Press release in English and Portuguese.

PhD research project

UNIVERSIDADE FEDERAL DE PELOTAS

POSTGRADUATE PROGRAM IN EPIDEMIOLOGY

GEOSPATIAL ANALYSES OF HEALTH INDICATORS USING NATIONAL HEALTH SURVEYS FROM LOW- AND MIDDLE-INCOME COUNTRIES:

Analyses of reproductive, maternal, newborn and child health

PhD Research Project

LEONARDO ZANINI FERREIRA

Supervisor: Aluísio J D Barros

Co-supervisor: Fernando Hartwig

August 2019.

UNIVERSIDADE FEDERAL DE PELOTAS

POSTGRADUATE PROGRAM IN EPIDEMIOLOGY

GEOSPATIAL ANALYSES OF HEALTH INDICATORS USING NATIONAL HEALTH SURVEYS FROM LOW- AND MIDDLE-INCOME COUNTRIES:

Analyses of reproductive, maternal, newborn and child health

LEONARDO ZANINI FERREIRA

Supervisor: Aluísio J D Barros

Co-supervisor: Fernando Hartwig

The presentation of this PhD research project is a partial requirement for obtaining a doctoral degree from the Postgraduate Program in Epidemiology, Federal University of Pelotas, Brazil.

August 2019.

Table of Contents

List of Tables

Glossary of terms and abbreviations

AIC	Akaike Information Criterion
ANN	Artificial Neural Network
CCI	Composite Coverage Index
DHS	Demographic and Health Surveys
GIS	Geographic Information Systems
GPS	Global Positioning System
HIB	*Haemophilus Influenzae* type B
INLA	Integrated Nested Laplace Approximation
LISA	Local Indicators of Spatial Autocorrelation
LMIC	Low- and Middle-Income Country
MeSH	Medical Subject Headings
MICS	Multiple Indicator Cluster Survey
RMNCH	Reproductive, Maternal, Newborn and Child Health
SDG	Sustainable Development Goals
UHC	Universal Health Coverage
WHO	World Health Organization

Abstract

In light of the agenda defined by the Sustainable Development Goals, improving data quality and availability is crucial to ensure countries are in the right direction towards reducing preventable maternal and newborn deaths. Assessing geographic inequalities is challenging since the main sources of information on low- and middle-income countries are not designed to be analyzed spatially. Recent improvements in computer processing power and the increased availability of ancillary data sources have contributed to the development of modeling approaches that aim to provide estimates at finer resolutions than provided by the administrative divisions commonly used in surveys. This project involves using these techniques of geospatial modeling to investigate geographic inequalities by generating high resolution estimates of health indicators. The estimates will be evaluated in two different scenarios: i) aligned with the most recent surveys, map the composite coverage index, a measure of universal health coverage based on essential reproductive, maternal and child health interventions, at second administrative level (e.g the equivalent to district or counties in different countries) and at pixel level (e.g 5 x 5 km or 10 x 10km); and ii) assess time trends using a geographic perspective to identify spatial patterns of increase in coverage of two vaccines since their introduction in Peru, in the early 2000s.

1 Planned articles

1. Spatial modeling on RMNCH in low- and middle-income countries: a systematic review

This review paper will focus on presenting an overview of the existing applications of spatial modeling on reproductive, maternal, newborn and child health (RMNCH) in low- and middle-income countries (LMICs). Also, we aim to understand and discuss the suitability of these methodologies in RMNCH by evaluating implementations and pointing out limitations. This review will also attempt to identify gaps and opportunities in the areas where spatial modeling is promising and unexplored.

2. High resolution mapping of the Composite Coverage Index in West and Central Africa

Spatial modelling of standardized RMNCH indicators is a novel and promising approach for producing more granular data, which will enhance locating vulnerable geographies and guiding interventions. This article will use these techniques to model the Composite Coverage Index (CCI) in selected countries from West and Central Africa with Demographic and Health Surveys (DHS) and ancillary data sources.

3. Spatio-temporal changes of *Haemophilus Influenzae* type B and Rotavirus vaccines

Whenever a new vaccine is introduced to the calendar, increasing the coverage rapidly and equitably is a challenge and concern to the health system. This paper seeks to understand the spatial patterns behind the expansion of vaccination coverage in newly introduced vaccines such as Haemophilus Influenzae type B (Hib) and Rotavirus in Peru and an additional country.

2 Background

As we move towards a more sustainable and equitable world, as recently praised by world leaders when establishing goals for the next 15 years, preventable deaths remain a massive public health concern in low and-middle income countries due to basic health interventions being inaccessible for far too many (Boerma *et al.*, 2018). It is estimated that 800 women and 7700 newborns die daily during or shortly after pregnancy and childbirth (Chou *et al.*, 2015), in which most of these deaths are preventable or treatable. Those unfair maternal and under-five deaths must be addressed while also taking into account that vulnerable and neglected subgroups of the population are usually much worse (Barros and Victora, 2013). The Sustainable Development Goals (SDGs) have emphasized the importance of reducing within-country inequalities, so no one is left behind. For that, health estimates that are more granular, at subgroup level, are essential, as national estimates often mask heterogeneities at finer resolutions (United Nations, 2015).

Epidemiology, in its essence, focus on studying subgroups of the population that have at least one characteristic in common. Other than blood ties, place of residence is a useful way of grouping people based on their similarities. At national level, residents of a given country share several cultural, political and environmental aspects, as well as specific beliefs and exposure to propaganda when compared to different nations. Yet, there is tremendous within-country heterogeneity - across regions, districts and municipalities. Each geographic division carries a great deal of social reflexes, and as we descend to finer resolutions, we are able to look at more specific contexts and their influences (Cummins *et al.*, 2007).

LMICs heavily rely on survey data to monitor and report maternal and child health progress due to weak or lacking health information and vital registration systems. Capturing geographic inequalities, though, is challenging as these surveys usually provide representativeness for large geographic areas. Since these surveys are not designed to be analyzed on a geographic perspective, information obtained through geospatial analysis, particularly spatial modeling, can be helpful to fill existing gaps of estimates on lower levels of disaggregation. As many countries are structured with decentralized health systems, health managers greatly benefit from local estimates as their autonomy in terms of financing and organization is endorsed by data.

Geospatial information has proven to be effective in supporting governments and organizations to plan and allocate policies and resources (Greene, 2000; Folger, 2010). Maps of high granularity can play an important role in advocacy as they are easy to interpret and are able to provide detailed information on where the need for interventions is greater. The use of Geographic Information Systems (GIS) is not limited to painting maps and pinpointing areas to intervene, though. It can contribute by creating or extracting knowledge when suitable data is available, and location influences the relationships under investigation. Summarizing information for a large number of settings and managing information from different data sources and types are also among the differentials of a geospatial application.

2.1 Spatial analysis

Spatial analysis may be relevant whenever the geographical space is believed to be associated with the subject under study. Within the health geography field of application, spatial analysis can be divided into descriptive or analytic (Litva and Eyles, 1995). The most widely known and possibly one of the first documented uses of spatial epidemiology was John Snow's map of cholera in London (Snow, 1855). Snow used the spatial distribution of the disease to test his hypothesis that the cases were clustered around one of the water pumps. Describing patterns in space is a key strategy to identify priority areas and guide the investigation of factors associated with high prevalence of a disease or low coverage of an intervention. On the other hand, analytical approaches begin when the focus becomes creating or extracting new information. Whether the objective is understanding the relationship of the phenomena with its determinants, or predicting estimates for unmeasured areas, spatial approaches are often applied when: a) territory can function as a proxy for unmeasured contextual, political or environmental influences, and b) data integration is required, especially when the information is available in pixels, satellite images or buffers.

2.2 Spatial autocorrelation

Tobler's first law of geography states that "everything is related to everything else, but near things are more related than distant things." (Tobler, 1970). This concept is the base of spatial autocorrelation, which aims to measure the inter-dependency of observations in a geographic space. Spatial correlation exists when the distribution of settings in the study area is not random. Several statistics, such as Moran, Geary and Getis-Ord, try to quantify the degree of correlation by analyzing the estimates according to the distance between its neighbors and the length of the borders they share (De Smith, Goodchild and Longley, 2018). Positive values indicate higher or lower estimates are more clustered than it would be expected if the distribution was random. Negative values indicate the estimates are more dispersed, similar to what can be observed with the black and white squares in a chess board. The null hypothesis is that the estimates are randomly distributed across the space, thus resulting in a value close to zero.

2.3 Spatial interpolation

In general, data points are observed and collected in order to describe a given phenomenon in a particular area. Most of the times, measured data is incomplete or inaccurate, and insufficient to provide values for the entire area under study. There is a variety of methods to impute missing data based on similar characteristics and patterns found in observed data. In spatial statistics, the act of impute values based on the distance to observed points is defined as spatial interpolation (De Smith, Goodchild and Longley, 2018). Basic methods, such as inverse distance weighting, defines a smooth gradient of values based on distance to closest neighbors alone.

2.4 Data availability for spatial analyses of health attributes

International surveys such as the DHS and the Multiple Indicator Cluster Survey (MICS) are carried out every three to five years, providing accurate and reliable data on maternal and child health indicators. The DHS program is carried out by the United States Agency for International Development (USAID) since 1984, designed as a continuation for the World Fertility Survey in monitoring the progress of women in reproductive age and children up to five years old (Corsi *et al.*, 2012). In the last decades, DHS have also started collecting Global Positioning System (GPS) coordinates for their surveys. Each cluster, a grouping of 20 to 30 households similar to census tracts, has its center georeferenced, allowing them to be linked with different sources of information, such as health facility or environmental data. Due to confidentiality concerns, each cluster is randomly displaced for up to 2km in urban areas and up to 5 km in rural areas, while 1% of them are displaced for up to 10km.

National health surveys are key data sources for LMICs and provide extensive and comparable information on health outcomes and sociodemographic characteristics that can be applied to a variety of research and policy questions at national and subnational levels (Carvajal-Aguirre *et al.*, 2017). Still, determining the mechanisms that lead to coverage or prevalence changes is often a complex task, of which may require combining different domains such as cultural, environmental and political to the available health information. Integrating and standardizing these data, though, is a time-consuming and complex task. Publicly available data sources come in different formats, resolutions and quality, of which many of them are uncleaned. Detailed information at high resolution is generally owned by governments or private institutions and obtaining it, when possible, is costly or depends on collaborating with the holders.

If access is the first barrier to data integration, putting everything together may be an even bigger challenge. Linking georeferenced datasets through GPS coordinates is the simplest approach when working with health facilities or specific events. Yet, individual level data, for ethical reasons, is not released with their precise coordinates (Gething *et al.*, 2015). Noise is added by scrambling the coordinates to ensure residents remain untraceable. Attaching this information to the closest territory it is contained, like the neighborhood, municipality or district is frequently the only possible option, yet precision is lost in the process. The amount of detail and precision that can be lost without jeopardizing a given study will be determined by the scope of the research question.

Assuming all necessary information on point coordinates and territorial areas is available, layers will often need to interact among themselves. Doing so requires human resources with the appropriate skillset, which may be yet another obstacle to overcome. As complexity increases, and many projects aim to go beyond developing enhanced thematic maps, the need for capacity building and collaboration networks with experienced analysts rises.

3 Rationale

To achieve the desired progress proposed by the 2030 agenda, information on the current situation of the countries and on how they are changing over time is key to ensure policies and investments are being aimed properly. Monitoring is best when data on effective and simple indicators are available. Most LMICs lack adequate health information systems and rely on survey data as their primary data source.

Universal health coverage (UHC) is a broad definition covering equitable, efficient and financially secure access to health services, comprising the World Health Organization (WHO) vision for health as a human right (World Health Organization, 2014). In RMNCH studies, UHC is often employed as a collection of preventive, curative and cost-effective essential interventions delivered throughout the entire continuum of care (Kerber et al., 2007).

Monitoring UHC in LMICs is frequently based on survey data collected by international organizations such as the DHS and MICS programs. They conduct interviews with a standardized questionnaire which allows for comparability of internationally accepted health indicators. Estimates for these indicators are provided at national and subnational level. The subnational units are generally representative of the first administrative level of the country. Though relevant patterns can be derived from these units, finer and more specific estimates allow for more focused interventions and better discrimination of geographic inequalities.

Current computer processing power and the continued improvement of geospatial techniques allied to the increase in reliable, updated and accurate data open up opportunities to study maternal, newborn and child health outcomes from different angles. As survey data is still the leading source of information, it is the academic community's duty to extract and report the maximum amount of knowledge available on these data sources to assist governments and stakeholders. The application of geospatial analyses can contribute to better understanding of the spatial patterns while also producing highly granular data for more focused interventions.

This project aims to explore the potential benefits of geospatial analysis in RMNCH through three different approaches:

a) Compiling knowledge from applications of geospatial analysis in the literature to uncover gaps and opportunities for better intervention planning and reduction of inequalities.

b) Generating more granular data on UHC to improve detection, visualization and action on geographic disparities.

c) Examining how vaccination coverage changes over time by studying the spatial patterns to identify possible delays and geographic barriers leading to unequal coverage.

4 Literature review

To summarize the existing literature, we carried out a systematic review by searching the PubMed (https://www.ncbi.nlm.nih.gov/pubmed) database for studies that used GIS and their primary data source was national health surveys in LMICs. The search strategy considered MeSH (Medical Subject Headings) terms and important keywords related to the scope of the study, which were also present in previously identified relevant articles.

("Health Surveys"[Mesh] OR "Demographic and Health Surveys" OR DHS OR MICS OR "Multiple Indicator Cluster Survey" OR "Health Survey") AND ("Geographic Information Systems"[Mesh] OR GIS OR Spatial Analysis[Mesh])

Most of the articles relevant to the composition of this project were open source. The university proxy access through *Periódicos CAPES* (http://www.periodicos.capes.gov.br) allowed us to obtain the remainder. The few exceptions were retrieved from collaborators or directly from the authors.

The initial search, on 18 May 2019, resulted in 1060 articles, of which 121 were pre-selected after screening through titles and abstracts. After full reading, another 67 were removed due to not meeting our eligibility criteria, yielding 54 selected studies.

From the 56 selected studies, 46 included African countries. Asia and Latin America and the Caribbean were included in only 10 and 6 studies, respectively. 80% of the studies were carried out after 2013, corroborating the field has grown in the most recent years. All selected studies were cross-sectional. The studies are described in Table 1.

The nutritional status of children was the most commonly studied outcome, being part of 13 studies. Sociodemographic characteristics were assessed as determinants of the given outcome in most of the studies. Based on Ebener *et. Al* (Ebener *et al.*, 2015), the selected studies were classified into three categories: thematic mapping, spatial analysis and spatial modelling.

4.1 Thematic mapping

Timely and targeted interventions require geographic patterns in countries, subnational regions or smaller territorial units to be properly described and documented. As defined by Ebener *et. Al* (2015), thematic mapping studies are those whose objective lays on the “creation of maps to convey information about a topic or theme”, thus providing augmented visualization details to help surveillance, policy and investments planning.

The spatial patterns are often presented as choropleth or heat maps. While choropleth maps respect the geographic boundaries and illustrate each unit of analysis (e.g., regions or districts of a country) with a single color, heat maps use the density of the available points to interpolate values for the entire surface, presenting a smoothed gradient. Regardless of the mapping strategy, most studies identified in our literature review have focused on identifying statistically clustered areas of low or high coverage. For this purpose, we identified two similar, but slightly different, approaches. The first is scanning the study space to identify areas where the outcome is spatially concentrated (Cuadros, Awad and Abu-Raddad, 2013; Cuadros and Abu-Raddad, 2014; Alemu *et al.*, 2016; Wong *et al.*, 2018). Kulldorff was the most known and used implementation in the studies following this approach (Kulldorff, 1997). The second approach uses measures of spatial autocorrelation or Local Indicators of Spatial Autocorrelation (LISA) to evaluate the existence of spatial patterns (Anselin, 1995). Although LISA detect clustered areas, their primary goal is to quantify the spatial dependence (or autocorrelation), highlighting where it is greater (Adekanmbi, Uthman and Mudasiru, 2013; Lopez-Cevallos, Chi and Ortega, 2014; Bogale *et al.*, 2017; Brownwright, Dodson and van Panhuis, 2017; Hasan *et al.*, 2018; Khan and Mohanty, 2018; Tewara *et al.*, 2018; Yourkavitch *et al.*, 2018). Additionally, a few studies attempted to describe geographic changes in coverage over time (Cuadros and Abu-Raddad, 2014; Barankanira *et al.*, 2017; Hasan *et al.*, 2018).

4.2 Spatial analysis

Spatial analysis covers a wide range of concepts and techniques where space is a central element. In our literature review, we have defined spatial analysis as the collection of methods aimed at extracting or creating new information from spatial data (Ebener *et al.*, 2015). Studies classified as spatial modeling are presented in the next section.

Among the studies identified in our literature review, investigating determinants of maternal and child health outcomes was the most frequent goal of spatial analysis. In addition to estimating the effects social and demographic health determinants, several studies considered a spatial correlation component by applying spatial models (Thuilliez, 2010; Owoo and Lambon-Quayefio, 2013; Chirwa *et al.*, 2014; Gayawan, 2014; Gayawan, Adebayo and Chitekwe, 2014; Gayawan, Arogundade and Adebayo, 2014; Kandala *et al.*, 2015; Mtambo, Masangwi and Kazembe, 2015; Ngwira and Stanley, 2015; Barankanira *et al.*, 2016; Chitunhu and Musenge, 2016; Haile *et al.*, 2016; Helova, Hearld and Budhwani, 2017; Ejigu, Wencheko and Berhane, 2018; Habyarimana and Ramroop, 2018). Both frequentist and Bayesian models were used in similar fashion for estimating the spatial effects throughout the studies. In a slightly different approach, a few studies attempted to estimate the independent effect of environmental factors on health outcomes. We identified studies assessing the influence of air pollution on neonatal and infant mortality (Goyal, Karra and Canning, 2019) and stunting (Goyal and Canning, 2017), proximity to vegetation areas and forest loss cover on child diet and nutrition (Johnson, Jacob and Brown, 2013; Galway, Acharya and Jones, 2018), proximity to conflict (Ostby *et al.*, 2018), vegetation and temperature on schistosomiasis (Yang *et al.*, 2005) and how earthquakes can affect child growth (Rydberg *et al.*, 2015). Giardina *et. al* (2014) examined the effect of interventions and environmental variables in malaria risk changes over times in a few sub-Saharan countries.

Geographic access to health is one of the most common and promising applications of GIS capabilities as mobility is associated with several environmental factors and trajectories could be observed or estimated. Six studies assessed how health care utilization is affected by distance and quality provided by the service (Heard, Larsen and Hozumi, 2004; Hong, Montana and Mishra, 2006; Choi *et al.*, 2010; Gabrysch *et al.*, 2011; McKinnon *et al.*, 2014; Skiles *et al.*, 2015; Tansley *et al.*, 2015; Gao and Kelley, 2019). Studies in which modeling of travel times or distance is estimated while considering associated covariates were classified as spatial modeling and are described in the next section.

4.3 Spatial modeling

A subtle but important difference distinguishes spatial modeling as a separate class of studies: incorporation of ancillary data in mathematical models to provide more precise and reliable estimates. Small area estimation and travel time calculation stand out in the literature as the most popular, yet promising line of works. To concentrate efforts and resources on the most needed, along with characterizing the vulnerable shares of the population, the location of these groups must be uncovered and described using solid and detailed data. National health surveys offer a vast amount of information for policy planners at subnational geographic units – commonly the first administrative level of the country. These units are often extensive and heterogeneous. In our literature review, spatial modeling studies are presented below according to their modeling purposes: a) small area estimation and b) calculation of travel times.

Small area estimation studies vary from descending to finer geographic division (e.g. from regions to districts or counties) to estimation of high-resolution units such as grids of $5km^2$ or $10km^2$. District-level estimates were generated for anthropometric indicators (Akseer *et al.*, 2018), under-5 mortality (Dwyer-Lindgren *et al.*, 2014), maternal and health care service utilization (Ruktanonchai *et al.*, 2016) and adolescence births (Neal *et al.*, 2016). Three studies provided maternal and child health estimates for smaller areas of up to $10km^2$ using associated covariates to improve precision (Acheson, Plowright and Kerr, 2015; Burke, Heft-Neal and Bendavid, 2016; Ruktanonchai *et al.*, 2016). Jia *et. al* (2016) generated spatially-smoothed coverage estimates disaggregated by socioeconomic position but no covariates were considered. Although Bayesian spatial models were the prevailing methodology for small area estimation in the studies identified in our review, Bosco *et al* (2017) compared them to machine learning and generalized linear models using a few outcomes in four different countries and found no strong evidence for preferring one method over the others.

Travel time models attempt to provide more accurate measures of access to healthcare. Geographic obstacles are known to affect the access to health facilities (Khan and Bhardwaj, 1994), especially in poorer and less structured areas (Gething *et al.*, 2012). Euclidean distances tend to underestimate travel times where terrain conditions reduce mobility. Among the studies identified in this literature review, different approaches were observed including the comparison of Euclidean distances to modelled travel times (Noor *et al.*, 2006) and examining the impact of sociodemographic and economic characteristics to enhance travel times models (Ouma *et al.*, 2017). Masters *et al* (2013) attempted to estimate the effect of travel times in health facility delivery in Ghana.

We recognize that a few relevant studies were not comprised in this literature review. This is due to the restriction of the search strategy to studies that were explicitly based on health surveys. Those missed studies, despites using survey data, lacked clarification on data sources in their MeSH terms, title or abstract. The search strategy will be expanded and improved for the proposed review paper.

Table 1. Results of the literature review

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
Goyal *et al* 2019	Outcomes: Neonatal and infant mortality Exposures: Air pollution ($PM_{2.5}$) 43 low and-middle income countries	Generalized linear models	Spatial analysis	To estimate the association of air pollution on neonatal and infant mortality.
Johnson *et al* 2013	Outcomes: Stunting, vitamin A, dietary diversity, episodes of diarrhea Exposures: Vegetation indexes Malawi	Generalized linear models	Spatial analysis	To estimate the association of proximity to vegetation areas with child nutrition indicators
Rydberg *et al* 2015	Outcomes: Stunting Exposures: Seismic activity intensity Peru	Multilevel models	Spatial analysis	To estimate the impact of an earthquake in public health outcomes using stunting as a proxy
Akseer *et al* 2018	Outcomes: Anthropometric indicators Exposures: Sociodemographic characteristics Afghanistan	Bayesian spatial models	Spatial modelling	To assess the geographical differences at district-level in anthropometric indicators
Ouma *et al* 2017	Outcomes: Health facility utilization Exposures: Travel time to health facility Three Kenya's counties	Bayesian spatial models	Spatial modelling	To compare the accuracy of travel time estimation models
Alemu *et al* 2016	Outcomes: Child malnutrition Ethiopia	Cluster analysis	Thematic mapping	To identify clusters of concentration of malnutrition in Ethiopia
Bosco *et al* 2017	Outcomes: Modern contraceptives, literacy and stunting Exposures: Travel time, distance, climate, demographic, environmental, among others Nigeria, Kenya, Tanzania and Bangladesh	Bayesian spatial models and machine learning	Spatial modelling	To test the accuracy of spatial predictive methods on modern contraceptives use, literacy and child stunting

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
Neal *et al* 2016	Outcomes: Adolescent first birth Tanzania, Kenya and Uganda	Bayesian spatial models	Spatial modelling	To assess the geographical distribution of adolescent first births in three countries
Jia *et al* 2017	Outcomes: Improved sanitation Kenya	Bayesian spatial models, cluster analysis	Spatial modelling	To estimate the coverage of access to sanitation in small areas and identify spatial clusters
Ruktanonchai *et al* 2017	Outcomes: Antenatal care, Skilled birth attendance and postnatal care Burundi, Kenya, Tanzania, Rwanda and Uganda	Bayesian spatial models	Spatial modelling	To produce a high-resolution inaccessibility score and district-level estimates to assess how spatial inequalities changed over time
Hong *et al* 2006	Outcomes: Contraceptives (IUD) Exposures: Health facility quality Egypt	Multilevel models	Spatial analysis	To estimate the effect of health facility quality on IUD coverage
Gabrysch *et al* 2011	Outcomes: Skilled birth attendance Exposures: Distance to facility and level of care provided Zambia	Multilevel models	Spatial analysis	To estimate the effects of distance and level of care on health facility delivery
Chirwa *et al* 2014	Outcomes: birth intervals Exposures: Sociodemographic characteristics DR Congo	Bayesian spatial models	Spatial analysis	To investigate the spatial heterogeneity of birth intervals in young Congolese women
Gayawan 2014	Outcomes: Institutional delivery Exposures: Sociodemographic characteristics Nigeria	Generalized linear models, Bayesian spatial models	Spatial analysis	To identify the determinants of place of delivery and the geographical location variation
Ngwira *et al* 2015	Outcomes: Birth weight Exposures: sociodemographic characteristics Malawi	Bayesian spatial models	Spatial analysis	To identify the determinants of birth weight and the spatial patterns

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
Tansley *et al* 2015	Outcomes: Access to emergency health facilities Exposures: Population, roads, Namibia and Haiti	Network analysis	Spatial analysis	To estimate the share of the population within 50km of health facilities with different levels of care
Wong *et al* 2018	Outcomes: Private facility birth delivery Nigeria	Cluster analysis, generalized linear models	Thematic mapping	To identify cluster of low and high private facility birth delivery
Acheson *et al* 2015	Outcomes: Insecticide treated nets Exposures: Population, land cover, elevation, vegetation and temperature Tanzania	Generalized linear models, Species distribution model	Spatial modelling	To compare the distribution of insecticide treated nets with concentration of malaria and estimate the areas at risk
Yourkavitch *et al* 2018	Outcomes: Exclusive breastfeeding, vaccination, care seeking, stunting and under-5 mortality 27 African countries	Cluster analysis	Thematic mapping	To identify areas of low coverage or high need for intervention
Gao and Kelley 2019	Outcomes: Health facility utilization Exposures: Distance to facility and quality of care Haiti and Kenya	Interpolation methods	Spatial analysis	To determine the effects of distance and quality of care on maternal health services
Masters *et al* 2013	Outcomes: Antenatal care and institutional delivery Exposures: Distance to health facility Ghana	Multilevel models	Spatial modelling	To estimate and map the travel time to health facilities and the effect on antenatal care and in-facility deliveries
Skiles *et al* 2015	Outcomes: Use of injectables and unmet need for contraceptives Exposures: Distance to health facility	Interpolation methods	Spatial analysis	To test the association between use of contraceptives and distance to health facilities

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
	Malawi			
Barankanira *et al* 2017	Outcomes: Stunting Côte d'Ivoire	Multilevel models	Thematic mapping	To investigate the spatial heterogeneity of stunting over time and the civil war effect
Hasan *et al* 2018	Outcomes: Malnutrition Bangladesh	Cluster analysis	Thematic mapping	To examine changes over time in spatial clustering of malnutrition
Cuadros *et al* 2014	Outcomes: HIV Tanzania, Malawi, Kenya and Zimbabwe	Cluster analysis	Thematic mapping	To identify how changes over time affected the clusters of high prevalence of HIV
Mtambo *et al* 2015	Outcomes: Overweight Exposures: Sociodemographic characteristics Malawi	Bayesian spatial models	Spatial analysis	To identify overweight determinants based on a quantile-based Bayesian regression
Bogale *et al* 2017	Outcomes: Diarrhea Ethiopia	Interpolation methods	Thematic mapping	To explore the spatial patterns of diarrhea in Ethiopia
Ejigu *et al* 2018	Outcomes: Anemia Exposures: Sociodemographic characteristics and malaria Ethiopia	Multilevel models	Spatial analysis	To identify the determinants of anemia and its spatial pattern
Khan and Mohanty 2018	Outcomes: Malnutrition India	Generalized linear models	Thematic mapping	To examine the spatial heterogeneity among districts and meso-scale correlates
Brownwright *et al* 2017	Outcomes: Measles 10 Sub-Saharan countries	Generalized linear models	Thematic mapping	To examine the spatial heterogeneity of measles coverage
Chitunhu and Musenge 2016	Outcomes: Malaria Exposures: Sociodemographic characteristics, vegetation, precipitation Malawi	Bayesian spatial models, Generalized linear models	Spatial analysis	To compare different methods for predicting malaria in Malawi

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
Habyarimana and Ramroop 2018	Outcomes: Contraceptives use Exposures: Sociodemographic characteristics Rwanda	Bayesian spatial models	Spatial analysis	To identify the determinants of contraceptive use in Rwanda
Barankanira *et al* 2016	Outcomes: HIV Exposures: Sociodemographic characteristics Burundi	Generalized linear models	Spatial analysis	To examine the spatial heterogeneity and the determinants of HIV
Burke *et al* 2016	Outcomes: Under-5 mortality Exposures: Malaria, conflict, temperature 28 Sub-Saharan countries	Interpolation methods, Generalized linear models	Spatial modelling	To provide high resolution estimates of under-5 mortality, assess the changes over time and compare differences between countries' borders
Tewara *et al* 2018	Outcomes: Malaria Cameroon	Interpolation methods	Thematic mapping	To provide hot-spot maps of malaria clustering in Cameroon
Gayawan *et al* 2014a	Outcomes: Anemia Exposures: Sociodemographic characteristics Nigeria	Bayesian spatial models	Spatial analysis	To identify the determinants of anemia in Nigeria
Otsby *et al* 2018	Outcomes: Institutional delivery Exposures: Organized violence 31 Sub-Saharan countries	Generalized linear models	Spatial analysis	To examine the relationship between proximity to conflict and institutional delivery
Owoo and Lambon-Quayefio 2013	Outcomes: Antenatal care Exposures: Sociodemographic characteristics Ghana	Generalized linear models	Spatial analysis	To investigate the effects of social influence and health insurance in antenatal care utilization
Noor *et al* 2006	Outcomes: Access to health facilities Exposures: Travel time Kenya	Naismith–Langmuir rule	Spatial modelling	To compare travel time models of access to health facilities

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
Cuadros *et al* 2013	Outcomes: HIV 20 Sub-Saharan countries	Cluster analysis	Thematic mapping	To identify clusters of high and low prevalence of HIV
Heard *et al* 2004	Outcomes: Modern contraception Malawi	Generalized linear models	Spatial analysis	To determine the effect of distance to health facilities on use of modern contraceptives
Thuilliez *et al* 2010	Outcomes: Fever and school failure Exposures: Sociodemographic characteristics and malaria Mali	Generalized linear models	Spatial analysis	To assess the impact of fever and malaria on school performance and discuss using fever as a proxy for malaria
Goyal and Canning 2017	Outcomes: stunting Exposures: Air pollution Bangladesh	Generalized linear models	Spatial analysis	To estimate the effect of air pollution on stunting
Adekanmbi *et al* 2013	Outcomes: Stunting Nigeria	Cluster analysis	Thematic mapping	To examine the spatial heterogeneity of stunting in Nigeria
Haile *et al* 2016	Outcomes: Stunting Exposures: Sociodemographic characteristics Ethiopia	Mutilevel models	Spatial analysis	To investigate spatial variation and factors associated with stunting
Gayawan *et al* 2014b	Outcomes: Exclusive breastfeeding Exposures: Sociodemographic characteristics Nigeria	Bayesian spatial models	Spatial analysis	To investigate the determinants and spatial heterogeneity in exclusive breastfeeding
Dwyer-Lindgren *et al* 2014	Outcomes: Under-5 mortality Zambia	Generalized linear models, Non-linear models	Spatial modelling	To compare methods of mortality estimation at district-level and evaluate spatio-temporal changes
Giardina *et al* 2014	Outcomes: Malaria 6 Sub-Saharan countries	Bayesian spatial models	Spatial analysis	To estimate the changes in risk of malaria and the effect of intervention over time

Author/Year	Study characteristics	Statistical methods	GIS coverage	Study objective
McKinnon *et al* 2014	Outcomes: Neonatal mortality Exposures: Distance to health facility Ethiopia	Generalized linear models	Spatial analysis	To examine the effect of distance to health facility on neonatal mortality
Kandala *et al* 2015	Outcomes: Modern contraceptives Exposures: Sociodemographic characteristics DR Congo	Bayesian spatial models, Generalized linear models	Spatial analysis	To investigate inequalities in modern contraceptive use in DR Congo
Galway *et al* 2018	Outcomes: Child diet diversity Exposures: Forest cover, aridity, sociodemographic characteristics 15 Sub-Saharan countries	Bayesian spatial models	Spatial analysis	To examine the association between diet diversity and deforestation
Helova *et al* 2017	Outcomes: Child mortality Exposures: Sociodemographic characteristics Pakistan	Multilevel models	Spatial analysis	To identify the determinants of child mortality at individual and community level
López-Cevallos *et al* 2014	Outcomes: Health service utilization Exposures: Sociodemographic characteristics Ecuador	Multilevel models	Thematic mapping	To examine the spatial distribution and inequalities in health service utilization
Yang *et al* 2005	Outcomes: Schistosomiasis Exposures: Vegetation, temperature China	Bayesian spatial models	Spatial analysis	To examine the relationship between schistosomiasis and environmental factors

5 Objectives

5.1 General objective

Investigate how geographic inequalities can be assessed through geospatial methodology

5.2 Specific objectives

1. Based on findings of the literature review, we aim to:
    a) Compare the spatial modelling methodologies utilized for small area estimation.
    b) Discuss spatial modeling suitability for given outcomes.
    c) Identify gaps and opportunities for further studies in RMNCH.
    d) Discuss how health policy makers can benefit from spatial modeling.
2. Model universal health coverage through the composite coverage index using DHS and ancillary data sources:
    a) Determine the best available analytical approach for the task.
    b) Generate gridded surface maps for environmental associated factors.
    c) Explore the possibilities for data imputation on missing CCI interventions to improve data availability.
3. Understand the spatial patterns of vaccination coverage and how they change over time:
    a) Explore how coverage of Hib and rotavirus vaccines increased geographically over the years in both countries given different implementation strategies.
    b) Identify geographical barriers delaying expansion of vaccination coverage.

6 Hypotheses

Since approaches introduced by this proposal point out to more descriptive rather than analytical scenarios, there are no clear hypotheses to be tested. Nonetheless, several challenges and patterns are expected.

1. We expect uncertainty to be inversely proportional to population density when modelling CCI interventions. In Africa, we expect CCI coverage to be higher in urban and capital city areas, and where richer populations live. We also expect to find higher coverage in areas with more convenient environmental situations, like proximity to permanent water sources, nighttime lights and remoteness.

a. We expect INLA to be the most appropriate modeling approach given several studies are using it to model RMNCH outcomes

2. After the implementation of Hib and rotavirus vaccines, we expect more urbanized and richer areas to receive and achieve higher coverage first when compared to more rural and poorer areas. We also expect lesser and late coverage to be found in remote areas with more difficult levels of accessibility.

7 Methods

7.1 Paper 1: systematic review

The review aims at summarizing the current applications of spatial modeling on maternal and child health in LMICs while discussing consolidated lines of work, challenges and unexplored paths. A systematic search strategy will be used, in conjunction with searching references from the selected papers. Also, the search strategy presented in Section 4, will be broadened so that papers not directly referring to surveys are not left out and to include new databases.

Spatial modeling is becoming increasingly popular in the field of epidemiology in recent years, including RMNCH, as observed in our literature review. Not only processing power enables more complex analyses, but methodologies have evolved considerably in a short time span. The review will attempt to summarize the growth of spatial modeling in the literature, pointing out to possible reasons for such gain in popularity. This has also raised concerns regarding inadequate use of these methodologies – e.g., for outcomes where the geographic correlation is not evident, therefore making use of inappropriate predictive modeling approaches.

7.2 Paper 2: high resolution mapping of the CCI

The CCI utilizes eight essential interventions from four phases of the RMNCH continuum of care and can be interpreted as a proxy measure of UHC for mothers and children (Wehrmeister *et al.*, 2016). Each of its components, their definitions and the CCI formulation are presented in Table 2. Estimates for the CCI and its components at national and subnational level are provided by the International Center for Equity in Health (International Center for Equity in Health, 2019), which reanalyzes DHS and MICS surveys to ensure comparability between and within the countries. For this analysis we will recalculate the CCI at the primary sampling unit level.

We anticipate a few challenges for generating CCI estimates for each cluster. Sample sizes for some of its components may be very low to provide reliable – or even any – estimates for some locations. Also, the standard error of the index is calculated using resampling strategies and the low level of disaggregation may invalidate this procedure. Alternatives for assessing variability and filling data gaps are likely to be necessary.

Given the CCI is a composite of several interventions, in light of varied performances observed in the literature (Bosco *et al.*, 2017), there is no consensus on what is the best modelling approach for the task. Regarding the vaccination component of the index, Bayesian spatial models have been tested in one West African country and results were promising (Utazi *et al.*, 2018). Bayesian models, through Integrated Nested Laplace Approximations (INLA), slightly outperformed machine learning techniques in modern contraceptives use estimation (Bosco *et al.*, 2017). However, they highlight, that Artificial Neural Networks (ANN) generally perform better on non-Gaussian distributions.

Two modeling exercises will assist in the selection of the best analytical approach. The first consist of comparing the predictive power of different techniques. Ensemble models (i.e combining different models to increase performance on the output) are also an option. The second lies on determining whether fitting several country-specific models will outperform a single generic model for all countries. Model's performance will be internally assessed based on statistics such as Akaike Information Criterion (AIC) and through n-fold cross-validation. A portion of the data will be left-out for final assessment of external validity, though properly selecting geographically representative training and validation sets is a challenge.

Ancillary data will be obtained from varied data sources to generate gridded surface estimates of associated contextual variables. Those may include population (Tatem *et al.*, 2015), poverty (Steele *et al.*, 2017), satellite imagery data (Tucker, Grant and Dykstra, 2004), among others. The selection of ancillary data and the choice of a specific analytic tool will be part of the activities to be developed during the sandwich PhD period.

Table 2. The composite coverage index and their components definitions

Indicator	Acronym	Numerator	Denominator
Demand for family planning satisfied by modern methods	DFPSm	Who is using (or whose partner is using) a modern contraceptive method	Women aged 15-49 years either married or in union in need of contraception
Antenatal care 4 or more visits	ANC4	Attended at least four antenatal care (ANC) visits with any provider	Women aged 15-49 years who had a birth in the last 2/3 years
Skilled attendant at delivery	SBA	Delivered by a skilled attendant (based on each country's definition of skilled attendant)	Women aged 15-49 years who had a birth in the last 2/3 years
BCG vaccination	BCG	Received Bacillus Calmette-Guérin (BCG) vaccine	All live-children, 12-23/18-29/15-26 months (according to country's calendar)

DPT3 vaccination	DPT3	Received 3 doses of Diphteria, Pertussis, Tetanus (DPT) vaccine	All live-children, 12-23/18-29/15-26 months (according to country's calendar)
Measles vaccination	MSL	Received measles vaccine	All live-children, 12-23/18-29/15-26 months (according to country's calendar)
Treatment for diarrhea	ORS	Received oral rehydration salts (ORS)	All live children aged 0-59 months with diarrhea in the last 2 weeks
Care-seeking for pneumonia	CAREP	Sought treatment from an appropriate health facility or provider.	Live children, 0-59 months, suspected pneumonia in the last 2 weeks
Composite coverage index	$CCI = \frac{1}{4}\left(DFPSm + \frac{SBA + ANC4}{2} + \frac{2(DPT3) + BCG + MSL}{4} + \frac{ORS + CAREP}{2}\right)$		

7.3 Paper 3: spatio-temporal vaccination patterns

Identifying determinants of immunization coverage is essential to implement successfully new vaccines into the routine immunization calendar. Equity-oriented implementation of new vaccines can dramatically shift the traditional patterns of coverage increase, as it is commonly observed in the inverse equity hypothesis (Victora *et al.*, 2018). One way to assess spatial differences associated with equity-oriented implementation is by evaluating coverage expansion in two opposite scenarios. In this paper, we will study Peru and an additional country as examples of implementation with and without equity-oriented approaches.

Peru holds a unique geography by having highlands, coast and forest areas within its territory. Additionally, Peru is located at the boundary of two tectonic plates, which causes the country to suffer with occasional earthquakes. In 2004, the country was the pioneer of a series of continuous surveys conducted by DHS yearly until 2012. Then, they decided to carry on through their own national statistics institute, with 2018 being the most recent survey to this date. With the introduction of new vaccines in the early 2000s, all this makes Peru a great case for evaluating how vaccination coverage expanded spatially over time and how environmental factors can interfere with it.

As part of the pentavalent vaccine, Hib was firstly introduced to Peru in 1998 in areas of extreme poverty (Padilla *et al.*, 2017). The pentavalent vaccine was adopted as part of the national immunization calendar only in 2004. Currently, as of 2016, 83.2% of children 18-29 months received Hib immunization in Peru. The rotavirus vaccine was first introduced in Peru in 2009. In the first year, the coverage was estimated in 41% and rapidly increased to 75% in 2010 (de Oliveira *et al.*, 2013). Information on both indicators is collected in DHS surveys from vaccinations cards and reports from the mother. The standard indicator for tracking vaccine coverage considers both sources and is defined as the number of children aged 18-29 that received the vaccines (3 doses of Hib or 2 doses of rotavirus vaccines) divided by the total number of children 18-29 months.

Senegal rose as a potential candidate for the secondary country as it is also part of a continuous DHS series and it includes both Hib and Rotavirus in their routine immunization calendar. Unfortunately, they only collected information on those vaccines in the 2017's survey, which prevents us from studying changes over. Yet to be decided, the additional country should ideally have at least three data points in a span of at least 10 years between extremes, have considerably improved coverage in the period and have introduced these vaccines in the routine immunization without equity-oriented approaches.

High-resolution estimates will be modeled yearly from 2004 to 2017 using DHS surveys and ancillary data sources. Absolute and relative changes will be examined overlaying maps from earlier years compared to most recent years. Coverage thresholds will be established to determine which areas have reached a desired coverage, which areas have not and how fast these changes took place in different locations.

8 Ethical considerations

The primary data sources for RMNCH information is DHS surveys and ethical approval was already obtained when the surveys were carried out. To preserve the confidentiality of the participants, the georeferenced coordinates of DHS clusters are jittered in up to 5km. Ancillary data sources, if required, will guarantee anonymity of the population under study.

9 Relevant and impact

Extending the application of geospatial analyses for monitoring RMNCH indicators has implications in both the academy and at decision making. A common question asked by local health managers is where and who is the population in need for intervention. These granular estimates allow for a closer look into the specificities of country's subnational divisions, which is particularly valuable in settings where decentralized health systems prevail. Optimally, they can assist targeting policies and investments to the most vulnerable areas.

Aside from the direct impact of these estimates for governments and policy makers, the academy builds knowledge upon small steps towards an asserted direction. Demonstrating a set of techniques is useful instigates further studies to push the field of study forward. Geospatial analysis has already proven itself as a promising method for monitoring UHC and unveiling geographic inequalities, despites several challenges such as dealing with uncertainty. Any step towards improving quality and availability of data must be addressed and spatial modeling is a strong candidate for succeeding in such necessary goal.

10 Timeline of activities

The activities will follow the timeline presented in Table 3, below. We have established a collaboration with University of Southampton and a sandwich period is planned for the second half of 2020. The PhD sandwich supervisor is yet to be decided. We expect the thesis defense to be held in the second half of 2021, achieving all the objectives in approximately 42 to 46 months.

Table 3. Timeline of activities

Year	2018				2019				2020				2021			
Activities	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
ICEH activities																
Literature Review																
PhD Work Planning																
Literature review																
Paper 1 writing																
Data analysis – paper 2																
Paper 2 writing																
PhD sandwich period																
Data analysis – paper 3																
Paper 3 writing																
Thesis defense																

11 Dissemination of results

The main results of the thesis will be presented in scientific events and published in indexed academic journals that we consider appropriate for the papers. In addition, these results will be sent to the press to communicate the community about the findings.

12 Funding

This work is supported by the Bill & Melinda Gates Foundation, the Wellcome Trust and the International Center for Equity in Health.

13 References

Acheson, E. S., Plowright, A. A. and Kerr, J. T. (2015) 'Where have all the mosquito nets gone? Spatial modelling reveals mosquito net distributions across Tanzania do not target optimal Anopheles mosquito habitats.', *Malaria journal*. England, 14, p. 322. doi: 10.1186/s12936-015-0841-x.

Adekanmbi, V. T., Uthman, O. A. and Mudasiru, O. M. (2013) 'Exploring variations in childhood stunting in Nigeria using league table, control chart and spatial analysis.', *BMC public health*. England, 13, p. 361. doi: 10.1186/1471-2458-13-361.

Akseer, N. *et al.* (2018) 'Geospatial inequalities and determinants of nutritional status among women and children in Afghanistan: an observational study', *The Lancet Global Health*, 6(4), pp. e447–e459. doi: 10.1016/S2214-109X(18)30025-1.

Alemu, Z. A. *et al.* (2016) 'Non random distribution of child undernutrition in Ethiopia: Spatial analysis from the 2011 Ethiopia demographic and health survey', *International Journal for Equity in Health*. International Journal for Equity in Health, 15(1), pp. 1–10. doi: 10.1186/s12939-016-0480-z.

Anselin, L. (1995) 'Local indicators of spatial association - LISA', *Geographical analysis*. Wiley Online Library, 27(2), pp. 93–115.

Barankanira, E. *et al.* (2016) 'Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention.', *BMC public health*. England, 16, p. 118. doi: 10.1186/s12889-016-2760-3.

Barankanira, E. *et al.* (2017) 'Stunting among children under 3 years of age in Cote d'Ivoire: spatial and temporal variations between 1994 and 2011.', *Public health nutrition*. England, 20(9), pp. 1627–1639. doi: 10.1017/S1368980017000544.

Barros, A. J. D. and Victora, C. G. (2013) 'Measuring Coverage in MNCH: Determining and Interpreting Inequalities in Coverage of Maternal, Newborn, and Child Health Interventions', *PLoS Medicine*, 10(5). doi: 10.1371/journal.pmed.1001390.

Boerma, T. *et al.* (2018) 'Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health', *The Lancet*, 391(10129), pp. 1538–1548. doi: 10.1016/S0140-6736(18)30104-1.

Bogale, G. G. *et al.* (2017) 'Spatial patterns of childhood diarrhea in Ethiopia: data from Ethiopian demographic and health surveys (2000, 2005, and 2011).', *BMC infectious diseases*. England, 17(1), p. 426. doi: 10.1186/s12879-017-2504-8.

Bosco, C. *et al.* (2017) 'Exploring the high-resolution mapping of gender-disaggregated development indicators.', *Journal of the Royal Society, Interface*. England, 14(129). doi: 10.1098/rsif.2016.0825.

Brownwright, T. K., Dodson, Z. M. and van Panhuis, W. G. (2017) 'Spatial clustering of measles vaccination coverage among children in sub-Saharan Africa.', *BMC public health*. England, 17(1), p. 957. doi: 10.1186/s12889-017-4961-9.

Burke, M., Heft-Neal, S. and Bendavid, E. (2016) 'Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis.', *The Lancet. Global health*. England, 4(12), pp. e936–e945. doi: 10.1016/S2214-109X(16)30212-1.

Carvajal-Aguirre, L. *et al.* (2017) 'Measuring coverage of essential maternal and newborn care interventions: An unfinished agenda', *Journal of Global Health*, 7(2), pp. 1–5. doi: 10.7189/jogh.07.020101.

Chirwa, T. F. *et al.* (2014) 'An exploratory spatial analysis of geographical inequalities of birth intervals among young women in the Democratic Republic of Congo (DRC): a cross-sectional study.', *BMC pregnancy and childbirth*. England, 14, p. 271. doi: 10.1186/1471-2393-14-271.

Chitunhu, S. and Musenge, E. (2016) 'Spatial and socio-economic effects on malaria morbidity in children under 5 years in Malawi in 2012.', *Spatial and spatio-temporal epidemiology*. Netherlands, 16, pp. 21–33. doi: 10.1016/j.sste.2015.11.001.

Choi, K. *et al.* (2010) 'Surveillance of febrile patients in a district and evaluation of their spatiotemporal associations: a pilot study.', *BMC public health*. England, 10, p. 84. doi: 10.1186/1471-2458-10-84.

Chou, D. *et al.* (2015) 'Ending preventable maternal and newborn mortality and stillbirths', *BMJ (Online)*, 351, pp. 19–22. doi: 10.1136/bmj.h4255.

Corsi, D. J. *et al.* (2012) 'Demographic and health surveys: A profile', *International Journal of Epidemiology*, 41(6), pp. 1602–1613. doi: 10.1093/ije/dys184.

Cuadros, D. F. and Abu-Raddad, L. J. (2014) 'Spatial variability in HIV prevalence declines in several countries in sub-Saharan Africa.', *Health & place*. England, 28, pp. 45–49. doi: 10.1016/j.healthplace.2014.03.007.

Cuadros, D. F., Awad, S. F. and Abu-Raddad, L. J. (2013) 'Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa.', *International journal of health geographics*. England, 12, p. 28. doi: 10.1186/1476-072X-12-28.

Cummins, S. *et al.* (2007) 'Understanding and representing "place" in health research: A relational approach', *Social Science and Medicine*, 65(9), pp. 1825–1838. doi: 10.1016/j.socscimed.2007.05.036.

Dwyer-Lindgren, L. *et al.* (2014) 'Estimation of district-level under-5 mortality in Zambia using birth history data, 1980-2010.', *Spatial and spatio-temporal epidemiology*. Netherlands, 11, pp. 89–107. doi: 10.1016/j.sste.2014.09.002.

Ebener, S. *et al.* (2015) 'The geography of maternal and newborn health: The state of the art', *International Journal of Health Geographics*, 14(1). doi: 10.1186/s12942-015-0012-x.

Ejigu, B. A., Wencheko, E. and Berhane, K. (2018) 'Spatial pattern and determinants of anaemia in Ethiopia.', *PloS one*. United States, 13(5), p. e0197171. doi: 10.1371/journal.pone.0197171.

Folger, P. (2010) *Geospatial information and geographic information systems (GIS): Current issues* *and future challenges*. DIANE Publishing.

Gabrysch, S. *et al.* (2011) ‘The influence of distance and level of care on delivery place in rural Zambia: A study of linked national data in a geographic information system’, *PLoS Medicine*, 8(1). doi: 10.1371/journal.pmed.1000394.

Galway, L. P., Acharya, Y. and Jones, A. D. (2018) ‘Deforestation and child diet diversity: A geospatial analysis of 15 Sub-Saharan African countries.’, *Health & place*. England, 51, pp. 78–88. doi: 10.1016/j.healthplace.2018.03.002.

Gao, X. and Kelley, D. W. (2019) ‘Understanding how distance to facility and quality of care affect maternal health service utilization in Kenya and Haiti: A comparative geographic information system study.’, *Geospatial health*. Italy, 14(1). doi: 10.4081/gh.2019.690.

Gayawan, E. (2014) ‘Spatial analysis of choice of place of delivery in Nigeria.’, *Sexual & reproductive healthcare : official journal of the Swedish Association of Midwives*. Netherlands, 5(2), pp. 59–67. doi: 10.1016/j.srhc.2014.01.004.

Gayawan, E., Adebayo, S. B. and Chitekwe, S. (2014) ‘Exclusive breastfeeding practice in Nigeria: a bayesian stepwise regression analysis.’, *Maternal and child health journal*. United States, 18(9), pp. 2148–2157. doi: 10.1007/s10995-014-1463-6.

Gayawan, E., Arogundade, E. D. and Adebayo, S. B. (2014) ‘Possible determinants and spatial patterns of anaemia among young children in Nigeria: a Bayesian semi-parametric modelling.’, *International health*. England, 6(1), pp. 35–45. doi: 10.1093/inthealth/iht034.

Gething, P. *et al.* (2015) *Creating spatial interpolation surfaces with DHS data*, *DHS Spatial Analysis Reports No. 11*. USAID. doi: 10.13140/RG.2.1.1938.5362.

Gething, P. W. *et al.* (2012) ‘Geographical access to care at birth in Ghana: a barrier to safe motherhood’, *BMC Public Health*. BMC Public Health, 12(1), p. 1. doi: 10.1186/1471-2458-12-991.

Giardina, F. *et al.* (2014) ‘Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: a spatial and temporal analysis.’, *The Lancet. Global health*. England, 2(10), pp. e601-15. doi: 10.1016/S2214-109X(14)70300-6.

Goyal, N. and Canning, D. (2017) ‘Exposure to Ambient Fine Particulate Air Pollution in Utero as a Risk Factor for Child Stunting in Bangladesh.’, *International journal of environmental research and public health*. Switzerland, 15(1). doi: 10.3390/ijerph15010022.

Goyal, N., Karra, M. and Canning, D. (2019) ‘Early-life exposure to ambient fine particulate air pollution and infant mortality: pooled evidence from 43 low- and middle-income countries.’, *International journal of epidemiology*. England. doi: 10.1093/ije/dyz090.

Greene, R. W. (2000) *GIS in public policy: Using geographic information for more effective government*. ESRI, Inc.

Habyarimana, F. and Ramroop, S. (2018) ‘Spatial Analysis of Socio-Economic and Demographic Factors Associated with Contraceptive Use among Women of Childbearing Age in Rwanda.’, *International journal of environmental research and public health*. Switzerland, 15(11). doi: 10.3390/ijerph15112383.

Haile, D. *et al.* (2016) ‘Exploring spatial variations and factors associated with childhood stunting in Ethiopia: spatial and multilevel analysis.’, *BMC pediatrics*. England, 16, p. 49. doi: 10.1186/s12887-016-0587-9.

Hasan, M. T. *et al.* (2018) ‘Spatiotemporal heterogeneity of malnutrition indicators in children under 5 years of age in Bangladesh, 1999-2011.’, *Public health nutrition*. England, 21(5), pp. 857–867. doi: 10.1017/S136898001700341X.

Heard, N. J., Larsen, U. and Hozumi, D. (2004) ‘Investigating access to reproductive health services using GIS: proximity to services and the use of modern contraceptives in Malawi.’, *African journal of reproductive health*. Nigeria, 8(2), pp. 164–179.

Helova, A., Hearld, K. R. and Budhwani, H. (2017) ‘Associates of Neonatal, Infant and Child Mortality in the Islamic Republic of Pakistan: A Multilevel Analysis Using the 2012-2013 Demographic and Health Surveys.’, *Maternal and child health journal*. United States, 21(2), pp. 367–375. doi: 10.1007/s10995-016-2121-y.

Hong, R., Montana, L. and Mishra, V. (2006) ‘Family planning services quality as a determinant of use of IUD in Egypt.’, *BMC health services research*. England, 6, p. 79. doi: 10.1186/1472-6963-6-79.

Huicho, L. *et al.* (2016) ‘Child health and nutrition in Peru within an antipoverty political agenda: A Countdown to 2015 country case study’, *The Lancet Global Health*, 4(6), pp. e414–e426. doi: 10.1016/S2214-109X(16)00085-1.

International Center for Equity in Health (2019) *International Center for Equity in Health*. Available at: www.equidade.org.

Jia, P. *et al.* (2016) ‘High-resolution spatial distribution and estimation of access to improved sanitation in Kenya’, *PLoS ONE*, 11(7), pp. 1–21. doi: 10.1371/journal.pone.0158490.

Johnson, K. B., Jacob, A. and Brown, M. E. (2013) ‘Forest cover associated with improved child health and nutrition: evidence from the Malawi Demographic and Health Survey and satellite data’, *Global Health: Science and Practice*, 1(2), pp. 237–248. doi: 10.9745/ghsp-d-13-00055.

Kandala, N.-B. *et al.* (2015) ‘Disparities in modern contraception use among women in the Democratic Republic of Congo: a cross-sectional spatial analysis of provincial variations based on household survey data.’, *Journal of biosocial science*. England, 47(3), pp. 345–362. doi: 10.1017/S0021932014000212.

Kerber, K. J. *et al.* (2007) ‘Continuum of care for maternal, newborn, and child health: from slogan to service delivery’, *The Lancet*, 370(9595), pp. 1358–1369. doi: 10.1016/S0140-6736(07)61578-5.

Khan, A. A. and Bhardwaj, S. M. (1994) ‘Access to health care: a conceptual framework and its relevance to health care planning’, *Evaluation & the health professions*. SAGE PUBLICATIONS, INC. 2455 Teller Road, Thousand Oaks, CA 91320, 17(1), pp. 60–76.

Khan, J. and Mohanty, S. K. (2018) ‘Spatial heterogeneity and correlates of child malnutrition in districts of India.’, *BMC public health*. England, 18(1), p. 1027. doi: 10.1186/s12889-018-5873-z.

Kulldorff, M. (1997) ‘A spatial scan statistic’, *Communications in Statistics-Theory and methods*. Taylor & Francis, 26(6), pp. 1481–1496.

Litva, A. and Eyles, J. (1995) ‘Coming out: exposing social theory in medical geography’, *Health & Place*. Elsevier, 1(1), pp. 5–14.

Lopez-Cevallos, D., Chi, C. and Ortega, F. (2014) ‘[Equity-based considerations for transforming the Ecuadorian health system].’, *Revista de salud publica (Bogota, Colombia)*. Colombia, 16(3), pp. 347–359.

Masters, S. H. *et al.* (2013) ‘Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis.’, *Social science & medicine (1982)*. England, 93, pp. 147–154. doi: 10.1016/j.socscimed.2013.06.012.

McKinnon, B. *et al.* (2014) ‘Distance to emergency obstetric services and early neonatal mortality in Ethiopia.’, *Tropical medicine & international health : TM & IH*. England, 19(7), pp. 780–790. doi: 10.1111/tmi.12323.

Mtambo, O. P. L., Masangwi, S. J. and Kazembe, L. N. M. (2015) ‘Spatial quantile regression using INLA with applications to childhood overweight in Malawi.’, *Spatial and spatio-temporal epidemiology*. Netherlands, 13, pp. 7–14. doi: 10.1016/j.sste.2015.04.001.

Neal, S. *et al.* (2016) ‘Mapping adolescent first births within three east African countries using data from Demographic and Health Surveys: Exploring geospatial methods to inform policy’, *Reproductive Health*. Reproductive Health, 13(1), pp. 1–29. doi: 10.1186/s12978-016-0205-1.

Ngwira, A. and Stanley, C. C. (2015) ‘Determinants of Low Birth Weight in Malawi: Bayesian Geo-Additive Modelling.’, *PloS one*. United States, 10(6), p. e0130057. doi: 10.1371/journal.pone.0130057.

Noor, A. M. *et al.* (2006) ‘Modelling distances travelled to government health services in Kenya.’, *Tropical medicine & international health : TM & IH*. England, 11(2), pp. 188–196. doi: 10.1111/j.1365-3156.2005.01555.x.

de Oliveira, L. H. *et al.* (2013) ‘Systematic documentation of new vaccine introduction in selected countries of the Latin American Region’, *Vaccine*. Elsevier Ltd, 31(SUPPL.3), pp. C114–C122. doi: 10.1016/j.vaccine.2013.05.032.

Ostby, G. *et al.* (2018) ‘Organized Violence and Institutional Child Delivery: Micro-Level Evidence From Sub-Saharan Africa, 1989-2014.’, *Demography*. United States, 55(4), pp. 1295–1316. doi: 10.1007/s13524-018-0685-4.

Ouma, P. O. *et al.* (2017) ‘Univariate and multivariate spatial models of health facility utilisation for childhood fevers in an area on the coast of Kenya’, *International Journal of Health Geographics*. BioMed Central, 16(1), pp. 1–8. doi: 10.1186/s12942-017-0107-7.

Owoo, N. S. and Lambon-Quayefio, M. P. (2013) ‘National health insurance, social influence and antenatal care use in Ghana.’, *Health economics review*. Germany, 3(1), p. 19. doi: 10.1186/2191-1991-3-19.

Padilla, J. *et al.* (2017) ‘Neumonías En Niños En El Perú: Tendencias Epidemiológicas, Intervenciones Y Avances’, *Revista Médica Clínica Las Condes*. Clínica Las Condes, 28(1), pp. 97–103. doi: 10.1016/j.rmclc.2017.01.007.

Ruktanonchai, C. W. *et al.* (2016) 'Equality in maternal and newborn health: Modelling geographic disparities in utilisation of care in five East African countries', *PLoS ONE*, 11(8), p. e0162006. doi: 10.1371/journal.pone.0162006.

Rydberg, H. *et al.* (2015) 'A Promising Tool to Assess Long Term Public Health Effects of Natural Disasters: Combining Routine Health Survey Data and Geographic Information Systems to Assess Stunting after the 2001 Earthquake in Peru.', *PloS one*. United States, 10(6), p. e0130889. doi: 10.1371/journal.pone.0130889.

Skiles, M. P. *et al.* (2015) 'The effect of access to contraceptive services on injectable use and demand for family planning in Malawi.', *International perspectives on sexual and reproductive health*. United States, 41(1), pp. 20–30. doi: 10.1363/4102015.

De Smith, M. J., Goodchild, M. F. and Longley, P. (2018) *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Sixth edit. Troubador publishing ltd.

Snow, J. (1855) *On the mode of communication of cholera*. John Churchill.

Steele, J. E. *et al.* (2017) 'Mapping poverty using mobile phone and satellite data', *Journal of the Royal Society Interface*, 14(127). doi: 10.1098/rsif.2016.0690.

Tansley, G. *et al.* (2015) 'Spatial access to emergency services in low- and middle-income countries: A GIS-based analysis', *PLoS ONE*, 10(11), pp. 1–12. doi: 10.1371/journal.pone.0141113.

Tatem, A. J. *et al.* (2015) 'Fine resolution mapping of population age-structures for health and development applications', *Journal of The Royal Society Interface*, 12(105), pp. 20150073–20150073. doi: 10.1098/rsif.2015.0073.

Tewara, M. A. *et al.* (2018) 'Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon;2000-2015.', *BMC infectious diseases*. England, 18(1), p. 636. doi: 10.1186/s12879-018-3534-6.

Thuilliez, J. (2010) 'Fever, malaria and primary repetition rates amongst school children in Mali: combining demographic and health surveys (DHS) with spatial malariological measures.', *Social science & medicine (1982)*. England, 71(2), pp. 314–323. doi: 10.1016/j.socscimed.2010.03.034.

Tobler, W. R. (1970) 'A computer movie simulating urban growth in the Detroit region', *Economic geography*. Taylor & Francis, 46(sup1), pp. 234–240.

Tucker, C. J., Grant, D. M. and Dykstra, J. D. (2004) 'NASA's global orthorectified Landsat data set', *Photogrammetric Engineering & Remote Sensing*. American Society for Photogrammetry and Remote Sensing, 70(3), pp. 313–322.

United Nations (2015) 'Transforming our world: the 2030 Agenda for Sustainable Development', *General Assembley 70 session*, pp. 1–35. doi: 10.1007/s13398-014-0173-7.2.

Utazi, C. E. *et al.* (2018) 'High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries', *Vaccine*. The Author(s), 36(12), pp. 1583–1591. doi: 10.1016/j.vaccine.2018.02.020.

Victora, C. G. *et al.* (2018) 'The inverse equity hypothesis: Analyses of institutional deliveries in 286 national surveys', *American Journal of Public Health*, 108(4), pp. 464–471. doi:

10.2105/AJPH.2017.304277.

Wehrmeister, F. C. *et al.* (2016) 'Summary indices for monitoring universal coverage in maternal and child health care', *Bulletin of the World Health Organization*, 94(12), pp. 903–912. doi: 10.2471/BLT.16.173138.

Wong, K. L. M. *et al.* (2018) 'Why not? Understanding the spatial clustering of private facility-based delivery and financial reasons for homebirths in Nigeria.', *BMC health services research*. England, 18(1), p. 397. doi: 10.1186/s12913-018-3225-4.

World Health Organization (2014) *Making Fair Choices on the Path to Universal Health Coverage: Final Report of the WHO Consultative Group on Equity and Universal Health Coverage*. World Health Organization.

Yang, G.-J. *et al.* (2005) 'A Bayesian-based approach for spatio-temporal modeling of county level prevalence of Schistosoma japonicum infection in Jiangsu province, China.', *International journal for parasitology*. England, 35(2), pp. 155–162. doi: 10.1016/j.ijpara.2004.11.002.

Yourkavitch, J. *et al.* (2018) 'Using geographical analysis to identify child health inequality in sub-Saharan Africa.', *PloS one*. United States, 13(8), p. e0201870. doi: 10.1371/journal.pone.0201870.

Project adjustments along the course of the work

Project adjustments

The initial project proposed an article to analyze the spatio-temporal vaccination patterns and compare the progress on geographical inequalities over time using Peru, a country that adopted equity-oriented policies, and an additional country that introduced new vaccines to the immunization routine in a regular fashion. This article was replaced due to data availability issues including the lack of information on the vaccines of interest and the absence of surveys with collected cluster coordinates. As an alternative, we developed an article that evaluated the empirical impact of an aggregation issue known as the modifiable areal unit problem in complex measures of inequality, and how different resolutions may hinder the assessment of time trends from a geographic perspective.

As a result of the review article, using a Bayesian framework with the integrated nested Laplace approximation was chosen as the most appropriate technique for modeling the composite coverage index (CCI). This approach was the most extensively used modeling strategy in the literature for modeling reproductive, maternal, newborn and child health indicators. We also opted for a policy-oriented article that focused on describing the geographic disparities of the CCI coverage at different levels of aggregation rather than highlighting the importance of the geospatial modeling methodology as a study that was a pioneer at modeling a composite index. The setting changed from West and Central African countries to Peru mainly for two reasons: the lack of geospatial modeling studies in Latin America, and the possibility of collaborating with a local researcher with expertise on health inequalities in Peru.

Activities report

Activities report

This section describes a series of activities that the student was involved during the PhD that directly or indirectly contributed to the development as a researcher. This thesis was produced within the International Center for Equity in Health (ICEH), one of the many research projects conducted at the Postgraduate Program in Epidemiology of the Universidade Federal de Pelotas. The ICEH mission is to monitor inequalities in health and nutrition, with a special focus on women at childbearing age and children under-five years old, in low- and middle-income countries using data from national health surveys.

International Center for Equity in Health

Founded in 2009 under the leadership of professors Cesar Victora and Aluísio Barros, the initial goal of the center was to analyze a series of health indicators disaggregated by a few inequality dimensions at large scale. At the time guided by the Millennium Development Goals, this process has allowed the group to contribute to the study of health equity as much as the monitoring and accountability of inequalities at global scale.

The ICEH started building a database that as of January 2022 includes 433 surveys from 118 countries providing information on over 375 indicators disaggregated by eight inequality dimensions. The main source of information are two series of surveys known as Demographic and Health Surveys and Multiple Indicator Cluster Survey, but more recently the group has started analyzing country-specific surveys such as the ones carried out in Brazil, Ecuador, and Mexico. From these surveys, we are able to analyze indicators related to reproductive health, antenatal care and birth, vaccines, breastfeeding, nutritional status, mortality, fertility, gender, and many others. All estimates are prepared at national level and disaggregated by place of residence, subnational region, sex of the child, wealth quintiles and deciles, woman's education, and woman's age.

The main activity of the researchers that work at the ICEH is to maintain and update this massive database that is the core of most studies produced by the group and is shared with several international partners including the Health Equity Monitor of the World Health Organization (WHO), the Lives Saved Tool, and the Countdown to 2030 initiative. Despite the evident relevance of the database itself, the most valuable asset may be considered the ability to analyze hundreds of surveys in a short period of time with the flexibility to modify any indicator or inequality dimension.

Capacity building

Starting in 2017, the WHO officialized the partnership with the ICEH by designing it as a collaborating center for health equity monitoring. One of the main activities of this collaboration is to provide training and education on health equity monitoring to strengthen the capacity of the countries to analyze their own data and use them to guide decisions and public health policies. In partnership with the Countdown to 2030 initiative, the ICEH has offered and has been part of several workshops and country case studies to enhance the analytical capacity of researchers and representatives of ministries and institutes of health.

In July of 2018, I was part of the team that organized the workshop entitled "Leaving no women and no child behind: levels and trends in inequalities for RMNCH by wealth, urban-rural residence, and administrative area" in Nairobi, Kenya. The main objectives of the workshop were to 1) develop comprehensive country and regional analyses on inequalities in reproductive, maternal, newborn and child health (RMNCH), and 2) strengthen the analytical skills of the participants in survey and other types of analyses, as well as interpretation and communication of the results. Participants from 15 African countries attended the 5-day event that consisted of lectures, group discussion and hands-on practical sessions.

In November of 2019, the ICEH, through the Countdown to 2030 initiative and affiliated organizations, held a workshop entitled "Female Headed-Households: intersectional analyses of gender and health in low- and middle-income countries". The event took place in Dakar, Senegal with 23 participants from 15 countries and a team of facilitators from Universidade Federal de Pelotas, African Population Health Research Center, WHO, University of Manitoba, University of Pretoria, and the American University of Beirut. The overall goal of the workshop was to enhance capacity on statistical and epidemiological data analyses of national surveys, with particular emphasis on gender equity analyses focused on female-headed households in sub-Saharan Africa.

In August of 2019, I was invited as a representative of the ICEH for a technical meeting on Geospatial Modeling for Immunization Equity in Washington, DC, USA. The meeting was organized by UNICEF and the Bill and Melinda Gates Foundation which brought together 30 modelers and technical partners to review the global efforts on geospatial immunization modeling and begin defining priorities to support the global immunization equity agenda through geospatial modelling.

Additional training

In July of 2019, the Fundação Oswaldo Cruz held a two-week course entitled "Análise Espacial e Geoprocessamento em Saúde" in Rio de Janeiro, Brazil, coordinated by professors Christovam

Barcellos and Monica Magalhães. The course focused on introducing key concepts of spatial analysis, geoprocessing and cartography, as well as enabling students to use geographic information systems to better understand, organize and plan health strategies and draw information from a geographical perspective.

In March of 2019, the University of Bristol held a short course on statistical methods for mediation analysis and repeated measures at the Postgraduate Program in Epidemiology. The course lasted a week and covered topics such as traditional mediation analyses, structural equation models, G-computation and methods for repeated measures.

Scientific collaborations

Beyond the articles that are part of the thesis, I got involved in several collaborations with other researchers that resulted in scientific products. These products are not necessarily related to the main theme of the thesis but were important in my training as a researcher.

Exploring the Potential for a New Measure of Socioeconomic Deprivation Status to Monitor Health Inequality

Submitted to International Journal of Equity in Health.

cixr, siilogit, siilin and equiplot: A set of programs to estimate and visualize inequalities

Submitted to Stata Journal.

Modern contraceptive use among women in need of family planning in India: an analysis of the inequalities related to the mix of methods used

Published at Reproductive Health. 2021. DOI:10.1186/s12978-021-01220-w

Association of the length of time using computers and mobile devices with low back, neck and mid-back pains: findings from a birth cohort

Published at Public Health. 2021. DOI:10.1016/j.puhe.2021.04.003

Are the poorest poor being left behind? Estimating global inequalities in reproductive, maternal, newborn and child health

Published at BMJ Global Health. 2020. DOI:10.1136/bmjgh-2019-002229

Early childhood suspected developmental delay in 63 low- and middle-income countries: Large within- and between-country inequalities documented using national health surveys

Published at Journal of Global Health. 2020. DOI:10.7189/jogh.10.010427

Wealth-related inequalities in the coverage of reproductive, maternal, newborn and child health interventions in 36 countries in the African Region

Published at Bulletin of the World Health Organization. 2020. DOI:10.2471/BLT.19.249078

Large and persistent subnational inequalities in reproductive, maternal, newborn and child health intervention coverage in sub-Saharan Africa

Published at BMJ Global Health. 2020. DOI:10.1136/bmjgh-2019-002232

Analyses of inequalities in RMNCH: rising to the challenge of the SDGs

Published at BMJ Global Health. 2019. DOI:10.1136/bmjgh-2018-001295

Trends in socioeconomic inequalities in stunting prevalence in Latin America and the Caribbean countries: differences between quintiles and deciles

Published at International Journal of Equity in Health. 2019. DOI:10.1186/s12939-019-1046-7

Measurement of social inequalities in health: concepts and methodological approaches in the Brazilian context

Published at Epidemiologia e Serviços de Saude. 2018. DOI:10.5123/s1679-49742018000100017

Article 1

Published at the International Journal of Health Geographics on

October 13, 2020

Ferreira et al. Int J Health Geogr (2020) 19:41
https://doi.org/10.1186/s12942-020-00239-9

International Journal of
Health Geographics

REVIEW

Open Access

Geospatial estimation of reproductive, maternal, newborn and child health indicators: a systematic review of methodological aspects of studies based on household surveys

Leonardo Z. Ferreira[1,2*], Cauane Blumenberg[1], C. Edson Utazi[3], Kristine Nilsen[3], Fernando P. Hartwig[2], Andrew J. Tatem[3] and Aluisio J. D. Barros[1,2]

Abstract

Background: Geospatial approaches are increasingly used to produce fine spatial scale estimates of reproductive, maternal, newborn and child health (RMNCH) indicators in low- and middle-income countries (LMICs). This study aims to describe important methodological aspects and specificities of geospatial approaches applied to RMNCH coverage and impact outcomes and enable non-specialist readers to critically evaluate and interpret these studies.

Methods: Two independent searches were carried out using Medline, Web of Science, Scopus, SCIELO and LILACS electronic databases. Studies based on survey data using geospatial approaches on RMNCH in LMICs were considered eligible. Studies whose outcomes were not measures of occurrence were excluded.

Results: We identified 82 studies focused on over 30 different RMNCH outcomes. Bayesian hierarchical models were the predominant modeling approach found in 62 studies. 5 × 5 km estimates were the most common resolution and the main source of information was Demographic and Health Surveys. Model validation was under reported, with the out-of-sample method being reported in only 56% of the studies and 13% of the studies did not present a single validation metric. Uncertainty assessment and reporting lacked standardization, and more than a quarter of the studies failed to report any uncertainty measure.

Conclusions: The field of geospatial estimation focused on RMNCH outcomes is clearly expanding. However, despite the adoption of a standardized conceptual modeling framework for generating finer spatial scale estimates, methodological aspects such as model validation and uncertainty demand further attention as they are both essential in assisting the reader to evaluate the estimates that are being presented.

Keywords: Geospatial modeling, Small area estimation, Reproductive health, Maternal health, Newborn health, Child health, Low- and middle-income countries, Household surveys

Background

Reproductive, maternal, newborn and child health (RMNCH) is central to the Sustainable Development Goals (SDG) agenda for 2030 given its potential for improving health and quality of life of current and future generations as summarized by the motto "survive, thrive, transform" adopted by the Every Woman Every Child initiative [1]. Despite progress in the area, with the increase

*Correspondence: lferreira@equidade.org
[1] International Center for Equity in Health, Universidade Federal de Pelotas, Pelotas, Brazil
Full list of author information is available at the end of the article

in coverage of several indicators, there is yet much to be achieved [2]. Planning and implementation of essential health interventions, delivered by supporting organizations and governments, is mainly done at small administrative divisions such as districts, states, provinces, regions or counties [3]. This requires geographically disaggregated information, which enables more precise adjustment of policies and targeting of resources [4].

Information on RMNCH indicators is predominantly obtained from national health surveys in low- and middle-income countries (LMIC), which offer standardized and reliable estimates [5]. Still, most surveys are usually designed to provide representative estimates at the largest administrative divisions as further disaggregation would require larger sample sizes [6]. Different estimation methods are required since direct estimation of lower administrative units in these surveys is highly imprecise. Geospatial approaches have been widely used for estimating RMNCH outcomes for small areas using georeferenced survey data. These methods derive indirect estimates from statistical models by 'borrowing strength' across space or from supplementary data, such as geospatial variables, censuses and administrative records [7]. However, censuses are carried out every 10 years or more in LMICs and administrative records are often incomplete, of poor quality or unavailable. Therefore, geospatial variables (information that is continuous across space, often retrieved from satellites or spatial interpolation), have been frequently used as supplementary data given their availability, timeliness, and reliance. The literature often uses the terms model-based geostatistics, small area estimation and (geo)spatial modeling interchangeably as model-based approaches to derive estimates for small geographies assisted by supplementary data.

Despite the rapid increase in the use of geographic information systems in RMNCH over the past decades, only a few studies have attempted to summarize these efforts. Two of them presented a broad review of spatial analyses in RMNCH [8] and health surveys in Sub-Saharan Africa [9], while one study focused on malaria transmission modeling [10]. Lastly, Rahman [11] carried out a review focusing on the methods used for estimation. To our knowledge, no study has comprehensively evaluated the most important methodological aspects for geospatial estimation of RMNCH indicators in LMICs. This assessment is necessary to identify approaches currently being used, their strengths and limitations and to help inform and improve future studies. Also, since these methodologies are relatively complex, non-specialists may struggle to evaluate and correctly interpret such studies. Therefore, this study aims to discuss the core methodological aspects of geospatial estimation, including any specificities employed for each RMNCH outcome, in studies focused on producing fine spatial scale estimates. In addition, we aim to enable non-specialist readers to critically evaluate and interpret these studies.

Methods

Conceptual framework

The structure and methodological aspects discussed in the review are guided by a standard modeling framework, adapted from Mayala et al. [12] and presented in Fig. 1. This conceptual framework is widely adopted in the literature and geospatial estimation studies, as it defines the flow of the modeling process. The use of the conceptual framework is not part of the eligibility criteria and has no effect on the selection of the studies.

Search strategy

Two independent reviewers carried out the same search strategy on August 28th 2020, screened and extracted the characteristics of the studies. Medline, Web of Science, Scopus, SCIELO and LILACS electronic databases were searched for studies based on survey data which applied geospatial approaches to estimate RMNCH outcomes in LMICs.

The search strategy consisted of a combination of health and geospatial keywords. The keywords "*health*" and "*epidemiology*" were used to define a broad health construct, rather than focusing on RMNCH outcomes, to increase the sensitivity of the search. For geospatial approaches, keywords were: "*geostatistical*", "*geo-statistical*", "*spatial modeling*", "*spatial modelling*", "*high-resolution mapping*", "*geospatial*", "*small area estimation*", "*small area estimates*" and "*spatial interpolation*". The complete keywords combination using logical operators is provided in Additional file 1. No restrictions on language or publication date were applied. In addition to the electronic databases, reference lists of the selected articles were searched for additional eligible studies not detected by the initial search strategy.

Articles retrieved from the search strategy were combined using Mendeley and exported to Rayyan, a web application for systematic reviews, for screening [13]. Initial duplicates were automatically removed in Mendeley, and the remainder were manually removed using Rayyan. The protocol for the systematic review was registered on PROSPERO (ID: 206323). This review follows the guidelines from PRISMA, and the checklist is provided in Additional file 2.

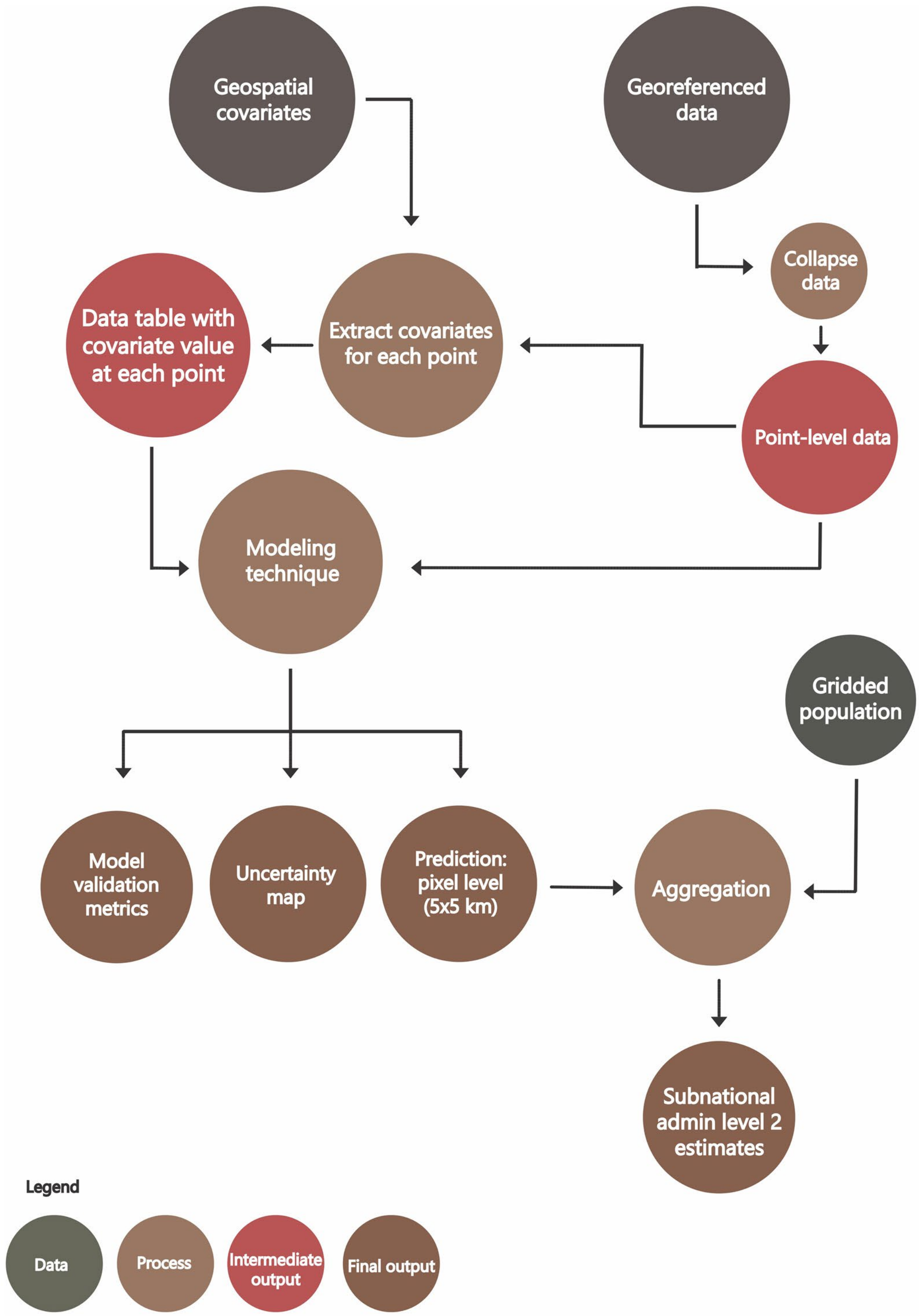


Fig. 1 Overview of a geospatial estimation process (adapted from Mayala et al. [12])

Eligibility criteria

To be eligible, studies must have fulfilled all the following criteria:

1. Carried out model-based geospatial approaches to obtain more geographically precise estimates than allowed by direct estimation due to insufficient sample size or to lack of representativeness;

2. Focused on RMNCH outcomes: coverage or impact indicators relevant to public health policies for women in reproductive age (15 to 49 years) or children aged < 5 years. Studies covering a broader age range for children but including the desired ages were also eligible;
3. Outcomes had to be measured in LMICs as defined by the 2020 World Bank country-income classification [14];
4. The main source of information were survey data and the minimum geographical coverage was an entire country.

The eligibility criteria were applied to all retrieved studies. Records were independently screened by both reviewers, first assessing titles and abstracts, then by reading the full text of the selected studies.

Exclusion criteria

Studies that did not estimate measures of occurrence of the coverage or impact outcomes were excluded.

Data extraction and quality assessment

We developed a Microsoft Excel spreadsheet to extract relevant characteristics of the selected studies based on ten pre-selected studies and on expert opinion. Then, each reviewer manually extracted the information from all the selected studies separately and the spreadsheets were compared later with disagreements dealt by consensus. The extracted characteristics, details, and guidance on how the spreadsheet was filled can be found in Additional file 3. Quality of the studies was assessed using Joanna Briggs Institute checklist for prevalence studies [15] and presented in Additional file 3.

Results of the literature search

After removing duplicates, 5567 records were identified for title and abstract screening, resulting in 126 selected articles. After full-text assessment, another 44 studies were removed yielding a total of 82 studies included in this review (Fig. 2). Several studies using the methods of interest, but estimating outcomes not considered to be RMNCH or covering age ranges outside our focus were not included in the review. The earliest studies identified were carried out in 2000, but the field grew steadily since 2016, comprising over 50% of the included studies (Additional file 1). The following sections discuss methodological aspects and outcomes. Due to the large number of studies reviewed, the following sections do not cite all studies in their respective categories. Details from each study are provided in Additional file 3.

Methodological aspects

Ideally, models built to predict unobserved data aim to minimize prediction errors, bias and overfitting of the data. Certain decisions are taken in each step of the process and presenting them in an organized and clear fashion is essential to allow readers to assess how reliable the estimates are. Based on the conceptual modeling framework presented above, we discuss the most important steps of geospatial estimation and details on how studies are reporting crucial information for their interpretation. These steps include data sources, covariates, modeling techniques, resolution, model validation and uncertainty.

Data sources

RMNCH outcomes in LMICs are often estimated using data from national health surveys. The Demographic and Health Surveys (DHS), a series of nationally representative household surveys conducted in over 85 countries [18], was the leading source of RMNCH information used in 59 of the 82 selected studies (72%). Further data sources include the Multiple Indicator Cluster Surveys [16], Performance Monitoring for Action [17], country-specific health surveys, censuses, and community surveys (main source of information for malaria).

DHS data are available at both administrative (or areal) level (e.g. regions, districts, provinces) and point level, i.e. the centroids of each primary sampling units (or survey clusters). The main difference between areal and point data is the aggregation of the data. While areal data are always summaries of individual level data, points can have both individual and aggregated information. For privacy reasons, DHS adds noise to their GPS coordinates, displacing them in a radius of up to 2 km for urban areas, up to 5 km for rural areas, and up to 10 km in 1% of the rural points. To account for this variation, DHS recommends drawing a buffer around each coordinate and averaging the neighboring values instead of using a precise match [19]. Despite that, only 16 of the 36 studies that used point-level DHS data reported taking steps regarding the displaced coordinates. Gething et al. [20] described the impact of the displacement as modest, overall, but varying between outcomes and locations.

Geospatial covariates

Geospatial variables or covariates are sources of information from determinants or proxies of determinants that are used as predictors in geospatial estimation for any given outcome. Obtaining and processing covariates is the most challenging and time-consuming step of the geospatial estimation process since the availability of this information is often limited to raw satellite indices, previous work, and a few initiatives. Covariates are used in the model for estimation and prediction and must be

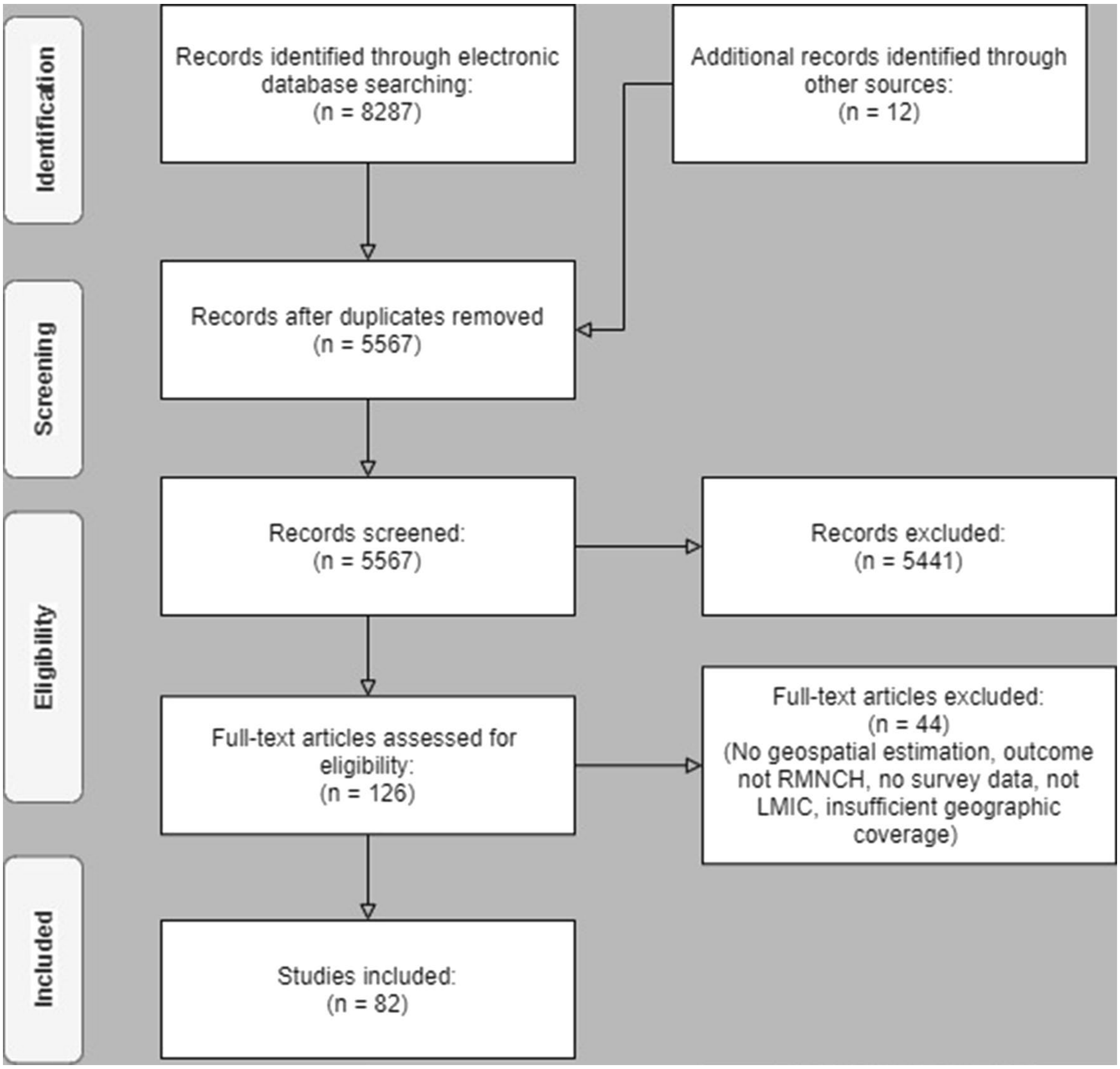


Fig. 2 Flow diagram of study selection

prepared accordingly. For estimation, each covariate information is extracted to the survey cluster location (or the available administrative level for areal models) and provided to the model along with the outcome. After model fitting, for prediction, a surface layer for each covariate is required at the desired resolution. Since these covariates often come from different data sources, aggregation is required when resolution is too high (e.g. satellite information) and interpolation when resolution is too low (e.g. creating surface layers from survey cluster coordinates).

The average number of covariates used across all studies was 9, ranging from 0 to 40. A total of 15 studies did not include any information on covariates into their models. We classified the covariates into seven groups: agriculture and livestock, climate, health-related interventions and outcomes, remoteness, satellite indices, sociodemographic, and topography and land cover. Covariates related to topography and land cover were the most common predictors found in 59 studies, followed by sociodemographic characteristics (53 studies), climate (43 studies) and remoteness (43 studies), as presented in Table 1. Additional file 3 provides the complete list of covariates for each study and their respective classifications.

The optimal number of covariates chosen as predictors, in order to optimize the refined estimation of the outcomes of interest, is a frequent topic of discussion. The principle of parsimony endorses the use of few and strong explanatory covariates to prevent overfitting the

Table 1 Summary of the characteristics for the selected studies

	All studies	Study outcomes[a]				
		Malaria	Child mortality	Malnutrition	Vaccination	Other outcomes
Number of studies	82	34	14	11	8	19
Covariates[a]						
Agriculture and livestock	17	3	3	5	4	5
Climate	43	28	5	5	4	5
Health-related interventions/ outcomes	24	9	5	5	1	7
Remoteness	43	19	7	5	5	11
Satellite indices	19	4	5	3	4	7
Sociodemographic	53	17	6	11	6	17
Topography/land cover	59	30	7	10	4	12
No covariates	15	4	7	1	2	2
Geographic coverage						
Single country	50	26	6	7	2	9
Multi-country	32	8	8	5	6	10
Temporal component						
No	46	19	2	8	7	11
Yes	36	15	12	4	1	8
Spatial resolution[a]						
Less than 5x5km	23	18	0	1	3	2
5x5 to 10x10km	20	6	5	5	3	4
Lower admin. level	30	2	8	6	1	14
Not reported	12	10	1	1	0	0
Uncertainty[a]						
Standard deviation map	14	6	1	1	5	2
Interval map/table	28	12	8	2	0	10
Relative map	7	0	2	3	0	2
Other metrics	13	9	2	0	1	1
Not reported	22	7	3	6	2	4
Modeling technique[a]						
Bayesian–MCMC	35	24	3	3	3	2
Bayesian–INLA	28	4	7	6	3	12
Classical GLM	17	5	2	2	2	6
Spatial interpolation	2	0	1	1	0	0
Ensemble models	12	1	5	4	1	5
Out-of-sample pred.						
Cross-validation	22	3	7	5	4	6
Hold-out	24	18	2	1	0	4
Not reported	36	13	5	6	4	9
Model fit metrics[a]						
Bias	34	12	7	6	4	9
RMSE/MSE	30	3	7	6	6	12
Coverage	24	8	6	4	4	5
DIC/AIC	19	6	3	3	1	6
MAE	16	7	2	3	2	3
Correlation	15	11	0	2	1	2
Other metrics	31	15	4	3	1	9
None reported	11	5	2	3	1	1

[a] These characteristics allow studies to be classified in more than one subgroup

MCMC Markov Chain Monte Carlo, *INLA* Integrated Nested Laplace Approximation, *GLM* Generalized Linear Models, *RMSE* Root Mean Squared Error, *MSE* Mean Squared Error, *DIC* Deviance Information Criterion, *AIC* Akaike Information Criterion, *MAE* Mean Absolute Error data. However, strong predictors are rarely available and insufficient covariate information may lead to model misspecification. This effort in finding the balance reinforces the importance of model validation (discussed later in the paper).

Modeling techniques

Once outcome and covariate information are prepared, they are passed to the chosen modeling technique, including random coefficients to account for spatial correlation and, sometimes, temporal correlation. The Bayesian approach was predominant in the selected studies, as 62 of the 82 studies were based on Bayesian hierarchical models (Table 1). The main conceptual difference, in comparison to the frequentist approach, lies on how the Bayesian framework interpret probabilities. In a frequentist framework, only repeatable events have probabilities, while Bayesian frameworks can assign probabilities to any event [21]. Since Bayesian frameworks also consider the distribution of its parameters, they generate complex posterior distributions, in which exact solutions are often not possible and numerical approximation techniques are required to fit the models. Markov Chain Monte Carlo (MCMC) methods approximate the true posterior distribution by generating dependent samples from it [22]. MCMC can be considered a turning point for Bayesian inference, having been used for model-fitting in 35 of the 62 studies that relied on Bayesian hierarchical models. More recently, Rue and colleagues [23] developed an alternative method called Integrated Nested Laplace Approximation (INLA), which quickly became popular given that it is much faster and yields very similar results compared to MCMC. Despite the first identified studies using INLA being carried out only in 2014, the method has already replaced MCMC in 28 studies. Frequentist estimation was applied in 17 studies through classical generalized linear models. Only two studies used spatial interpolation methods such as kriging [24] and kernel density estimation [25].

Recent studies have started using an ensemble approach, known as stacked generalization, to improve model performance [26]. Briefly, this strategy consists of fitting several models (usually each model uses a different modeling technique), generating intermediate predictions. These predictions are then used as input to a second model. The use of multiple modeling techniques allows any complex non-linear effects of the covariates to be captured, while the final predictions are estimated using a robust, consolidated modeling technique. All 11 studies following this approach [27–37] used a Bayesian hierarchical model fitted using INLA for the final predictions. Only one study relied on ensemble models and did not perform stacked generalization [38].

In addition to borrowing strength from covariance structures through space, spatio-temporal models can also benefit from these structures through time. The inclusion of a temporal component was identified in 36 studies as observed in Table 1. In 34 out of the 36 studies, this approach attempted to evaluate changes over time. These effects were primarily modelled using conditional autoregressive models and stochastic partial differentiation equation as described by Blangiardo et al. [39].

Resolution

Estimates are typically generated at two different levels of aggregation: grid cells or country's administrative divisions. At grid cell-level, the entire country is divided into an equally sized grid and predictions are made for each cell individually. A total of 55 out of the 82 studies opted for gridded-estimates (Table 1). Apart from three studies [40–42], the approximate cell size (grid) for all reported resolutions ranged from 1×1 km to 10×10 km, with 5 x 5 km being the most common one. Smoothed maps were presented by 12 studies without specifying the originally estimated resolution. Estimates for districts, counties, provinces and other low administrative divisions were produced by 30 studies (37%). These administrative level estimates are often produced from the grid level estimates through population-weighted aggregation using gridded population data from, e.g., WorldPop [43]. Only five of the 82 studies presented estimates at both grid cell and administrative levels. There is much discussion regarding the ideal level of aggregation, as it depends on multiple factors including the outcome, the objective of the analysis, how decentralized decision-making is within the country and the trade-off between precision and resolution [41, 44].

Model validation

A good predictive model is a model capable of reproducing the process that generates the outcome. However, depending on the outcome, available covariates and model specification, its performance can vary substantially [45]. Models should be validated against data that was not used in its construction. Otherwise, the model can learn the data instead of their underlying structure, a phenomenon known as overfitting the data. The simplest choice for out-of-sample predictions is known as the hold-out method, which splits the data into two subsets—a training subset used for model fitting and a test subset used for validation. This approach was found in 29% of the studies (Table 1). However, splitting the data and ensuring geographical representativity in both samples is an overlooked challenge. Seven studies [47–53], though, attempted to overcome this limitation using a declustering method, which gives less weight to observations geographically clustered when drawing the samples [54]. An alternative method for out-of-sample prediction is the n-fold cross-validation, found in 22 studies. The algorithm divides the data in "n" parts of equal size (which can be done in a spatially structured or random manner), leaving one for validation and using the remaining to build the model. This process is repeated until each fold is used for validation and the average of the combined measures is taken. Cross-validation is particularly useful when data is limited and holding out data could compromise the model performance, while the drawback is that it must fit one model for each fold, drastically increasing the processing time. Nearly half of the studies (44%) did not report on any validation method.

Within the out-of-sample data, there are several metrics that can be calculated and reported to assess the validity of the model predictions. As shown in Table 1, bias was the most reported validation measure, present in 34 studies (41%). Bias is the average of the difference between the observed and the predicted value. The magnitude of the prediction errors was often reported using the root mean squared error (RMSE) or the mean absolute error (MAE), found in 30 and 15 studies, respectively. RMSE and MAE are both positive values indicating how much, on average, the model predictions differ from the observed results. They differ on how deviations are handled: RMSE takes deviations squared while MAE ignores the signal. This makes RMSE more susceptible to the impact of high magnitude prediction errors (such data points are often referred to as outliers) [46]. A total of 19 studies presented Deviance Information Criterion (DIC) or similar metrics during model selection or validation. While DIC is useful for model selection, it has no direct interpretation and cannot be used to compare different studies. Additionally, several studies reported the achieved coverage within credible intervals [24] and the correlation between predicted and observed values [16].

Presentation of uncertainty

As much as the most precise estimates are desirable, there is always a degree of uncertainty in predictions made. While geographically disaggregated point estimates are easily interpretable when presented in a map, the related uncertainty is much harder to present in an intuitive way. Uncertainty is a complex multi-layer concept and it exists in every step from data collection to the modeled estimates. For the sake of this study, we considered uncertainty as the measures of variability associated to the estimates, since a complete definition includes measurable and unmeasurable components, sampling, modeling strategies, and is out of the scope of this study.

Visualization approaches to present uncertainty in a clear, comprehensive, and interpretable manner are still to be proposed. Bayesian models, for instance, produce full posterior estimates that can be summarized in multiple ways. However, there is no visualization approach that can fully address the challenges of communicating and using uncertainty and, as a result, the literature clearly lacks standardization.

Options for presenting uncertainty are tied to the chosen resolution. At the administrative level, where there is a smaller number of divisions, uncertainty can be described using maps or tables. On the other hand, grids of high resolution can only be represented in maps due to the large quantity of estimates. Uncertainty intervals and standard deviation maps and tables were the most common approaches, found in 28 and 15 studies, respectively. There were also seven studies presenting qualitative measures of uncertainty (e.g., low or high uncertainty). Further approaches include: coefficient of variation [55, 56], exceedance thresholds [57, 58], probability of being correctly classified [48, 50, 52] and Coffey-Feingold Bromberg metric [31]. A total of 22 studies (27%) did not present any measure of uncertainty (Table 1). Aside from the numerous ways of expressing uncertainty, it has been exclusively reported in supplementary files of 23 out of the 60 studies (38%) that presented uncertainty, putting its relevance in check.

Maps with the limits of the uncertainty intervals are often presented in two separate figures, demanding more space, and only covering a best–worst case scenario. Some studies use the width of the interval as an alternative, which is limited when the probabilities are close to zero or 1. Standard deviation maps are harder to interpret, especially for non-specialist readers. Lastly, qualitative measures of uncertainty are likely the easiest to interpret, although defining what is low or high uncertainty is arbitrary.

Key aspects for interpreting geospatial studies

Maps are long used for presenting geographically disaggregated estimates and are often easily interpretable. However, legend scales may be misleading, especially when intervals of different widths are grouped and presented together, or the amplitude is too narrow or too wide. These caveats are particularly important when several maps are presented in sequence and the reader may assume the legend scales are the same.

Every modeled estimate carry assumptions and uncertainties, and several aspects can be observed to assess their reliability. For instance, data sources must provide sufficient information for models to reproduce the occurrence of the outcome. The data must also come from reliable sources and be temporally close to the objective of the study and the covariates used in the process. In the case of multiple data sources and temporal assessment, constant change over time is often an assumption that needs to be taken into consideration.

Models tend to assume the input data are correct, so any estimates from a good predictive model can only be as accurate as the quality of the data sources. As discussed in previous sections, several metrics can be reported and interpreted to evaluate the validity of the predictive model. Bias, the most reported validation measure, indicates whether prediction errors are systematically leaning towards any direction. Therefore, an unbiased estimator should present bias close to zero. However, the scale of the outcome must always be considered when interpreting these measures. For instance, a bias of 1.5 in settings where the average mortality rates are around 3 is huge (50% of the point estimate), but for mortality rates close to 150, the relative importance of the same bias is much smaller (1% of the point estimate). The same applies for interpreting measures of the magnitude of the prediction errors, such as RMSE and MAE. For coverage indicators (bound between 0% and 100%), RMSE or MAE values of 2 indicate the model deviates from the true value, on average, by 2 percentage points.

Incorporating uncertainty in decision-making is often a major challenge. The interpretation of the estimates requires changes in the thought process to consider probabilities rather than an absolute, fixed value. Non-experts tend to depend on heuristics rather than formal statistics when taking decisions [59]. This raises a question on whether considering uncertainty leads to better decisions or simply discredits information in which uncertainty estimates are high [60]. Associated credible intervals can be interpreted as that we are confident (usually 95% confident) that the true estimate is within the interval. Therefore, smaller intervals reduce the probability of our estimate deviating from the true value. The standard error can be roughly interpreted as how precise the sample mean estimate is in relation to the population mean.

Outcomes

Around 30 different outcomes were estimated using geospatial approaches among the selected studies. We classified them into five groups based on their frequency and similarity: malaria, child mortality, malnutrition, vaccination, and other health-related outcomes. Within each family of outcomes, their specificities are highlighted and the summary of characteristics for all studies and by outcome is presented in Table 1.

Malaria

Malaria-related studies could be considered the pioneers in RMNCH geospatial modeling with a large contribution to this field. It took nearly a decade for studies of other RMNCH outcomes to start using geospatial estimation to increase the granularity of their available data. The first identified studies are dated to the early 2000's [61, 62], despite other spatial statistics in the field of malaria having been used for several years before [63]. Malaria is strongly affected by environmental factors. The mosquitoes of the anopheles species require certain climatic conditions to develop themselves and act as transmission vectors for the disease [64]. This geographical dependence along with the burden of the disease led malaria to be the most studied outcome with 34 out of the 82 studies [38, 42, 47–49, 51–53, 56–58, 61, 62, 65–83].

Most of the information for malaria in LMICs comes from combining multiple malariometric surveys conducted at specific locations. Several projects, such as MARA [84] and the Malaria Atlas Project [85], have worked on putting together geo-referenced malaria survey data, allowing researchers to use the pre-processed databases. These surveys were used in 21 of the 34 studies that focused on malaria. Despite concerns over malariometric surveys being carried out only in endemic areas of high prevalence, evidence shows they are well geographically distributed in various settings [47]. A secondary source of information for malaria are nationally representative surveys, either designed for several RMNCH indicators, such as the standard DHS surveys, or focused on malaria as in the Malaria Indicator Surveys, also carried out by the DHS program. A total of 14 studies relied on these surveys.

Although the malaria burden is not limited to children, they are the most affected subset of the population due to the lack of post-infection immunity [86]. Malaria indicators were reported as malaria prevalence, parasitemia risk or number of infected children. Both children under-five and the standardized age range of 2 to 10 years were the most common age subgroups, as observed in 14 and 12 studies, respectively. A few studies presented estimates for other subgroups such as: 6–59 months [79], under-10 years [61], under-16 years [71], 1–10 years [66, 87] and 1–14 years [65].

Different from other outcomes, malaria studies prioritized high-resolution estimates over small administrative units. The only two studies that presented county [58] and regional [72] level estimates also presented estimates at finer resolutions. Single country studies were predominant with 77% of the geographical coverage, while 82% opted for the Bayesian approach for modeling.

Childhood mortality

Child survival is a central goal of maternal and child health interventions and it is considered both a health indicator and a measure of human development [88]. Reducing child mortality rates is a long-term priority defined by international organizations and highlighted in both the Millennium Development Goals and the SDGs [89]. Even in high mortality settings, the death of a child is a rare event, thus requiring larger samples sizes in comparison to other RMNCH indicators. Among the reviewed studies, 14 were focused on child mortality.

Within child mortality studies, we identified ten studies focusing on all-cause mortality and four studies presenting cause-specific deaths. All ten studies reporting on all-cause childhood mortality estimated under-five mortality rate, while a few studies also presented estimates for neonatal [30, 90] and infant [90, 91] mortality rates. For cause-specific mortality, deaths by malaria [38], diarrhea [32, 35] and lower respiratory infection [33] were studied.

Most studies, 12 out of 14, assessed changes over time—a major focus for mortality—most likely relating to monitoring development goals. In terms of resolution, six studies aimed at reaching smaller administrative units such as districts or counties [35, 55, 92–96], six presented gridded estimates [25, 30, 32, 33, 38, 91] and one employed both approaches [90]. As with all outcomes, Bayesian models were predominant, used in 10 out of 14 studies. Four studies attempted to develop or enhance methods to estimate under-five mortality.

Malnutrition

Each year, 3.1 million deaths of under-five children are directly attributable to undernutrition in the form of stunting, wasting and micronutrient deficiencies [97], and overweight in children is an increasing problem. A total of 12 studies focused on malnutrition.

The burden of stunting, wasting, underweight and overweight was estimated for the entire African continent [28] and in all LMICs [27, 37]. There were also several single country studies that account for and focus on local specificities as done in Bangladesh [98], Afghanistan [99], Cambodia [100], India [36], Mexico [101] and Ethiopia [24, 102]. Five studies generated estimates at district or province level [98–100, 102], four studies at 1x1km [45], 5x5km [27, 28] and 10x10km [103], and two studies at both 5x5km and administrative level [36, 37]. Six studies modeled their outcomes using Bayesian models through INLA.

Among all outcomes, uncertainty was least reported on studies focusing on malnutrition, available in only half of the studies.

Immunization

Vaccines save the lives of millions of children every year, and despite being one of the most cost-effective health interventions, many settings have seen coverage levels stall or even decline in recent years [104]. For measles, which is highlighted in six of the seven immunization studies, many outbreaks occurred globally in 2018 and 2019, mainly due to lack of access and anti-vaccination movements [105–107].

Geospatial modeling of immunization started relatively recently, since all identified studies were published from 2015 onwards. Possibly due to being very recent, most of them carried out very comprehensive modeling approaches. Six of the eight studies produced estimates for at least three countries and only two failed to report uncertainty measures. The granularity pursued was also very high, having three studies at 1x1km [108–110], three studies at 5x5km [31, 111, 112] and one at 10x10km [113]. Perhaps due to being the first, Pramanik et al. [114] was the only vaccination study which focused in a single country, aimed at lower administrative units rather than gridded estimates, and one of the two studies that did not report uncertainty measures.

Other RMNCH outcomes

The use of geospatial approaches to produce estimates for small areas has reached a variety of outcomes. Within reproductive health, we identified four studies focusing on contraception [45, 115–117] and two studies on undesired adolescence pregnancies [118, 119]. From pregnancy to child birth, four studies focused on antenatal care, skilled birth attendance, c-section and postnatal care [40, 41, 120, 121]. Diarrhea [32, 122, 123] and respiratory infections [33] were the focus of a total of seven studies, as they are still among the leading causes of death for children in the poorest countries. One study also attempted to map exclusive breastfeeding [34].

Conclusions

The field of geospatial estimation focused on RMNCH outcomes is expanding and the number of published studies has increased more rapidly since 2014. Bayesian hierarchical models have taken place as the preferred modeling technique, but this is a continuously evolving area. More recently, ensemble approaches using several different models that are put together with a Bayesian model have been increasingly used and have the potential to become the approach of choice. The main data sources are likely to remain the same, DHS with a special place among national health surveys, especially that they have been putting a lot of effort in providing geolocated covariates available already harmonized with the surveys [125].

Geospatial models are complex and tend to produce a large number of estimates. Therefore, a validity assessment of how assumptions hold, the estimates precision and the model mean error should always be done and presented. These characteristics should be evaluated out-of-sample using one of the several approaches proposed in the literature, with cross-validation being the most efficient in terms of data use. However, with such complex models, fitting a model repeatedly can demand considerable processing power. Nonetheless, this is a key step to show that the results presented are stable and represent the underlying process in study. Model validation needs to be clearly presented both in terms of how it was done and its results. In our review, a considerable number of studies failed to present clear and convincing model validation—36 out of 82 —what makes the results much harder to interpret.

Other important aspects of geospatial modeling are the resolution of the estimates and how these are presented in terms of both point estimates and their uncertainty. The objective of the work is central to choosing the resolution or the type of aggregation to be used. A study describing the spatial distribution of an outcome or showing associations with geographical aspects can present very high-resolution estimates. On the other hand, if the aim is to support health policy decisions, estimates matching health districts, or geographical units where policies and programs are decided and implemented, are likely to be much more useful. The presentation of estimate uncertainty is also essential. However, we have not identified in the literature a clear and robust approach, as this represents a real challenge. Different measures of uncertainty have been used, as well as a variety of approaches of presentation – from simple to complicated. Given its importance, it seems to us that simpler and more direct visualization approaches could be used in the main body of the paper, while full results could be reported in the supplementary material.

As a final comment, given the often-large number of maps and diagrams presented, special attention has to be devoted to comparability of the scales used, color schemes, and even the map projections. The results need to be presented in an intuitive and understandable fashion so that non-specialists can grasp and make use of such relevant estimates. Authors need to put as much effort in the clarity of their presentations as they invest in the complex process of geospatial estimation.

This study covers the main methodological aspects that are part of a standard conceptual modeling framework adopted by the literature. However, many details that are lightly discussed here could be the focus of further studies such as a thorough evaluation and comparison of modeling techniques, covariates, and uncertainty. In addition, concern should be raised on how far these models can be extended, given the expansion of the field to over 30 different outcomes. Since predictions are based on space and time correlation and explanatory variables, producing fine spatial scale estimates may not be feasible for all outcomes [45].

The authors encourage future studies focused on modeling RMNCH outcomes using geospatial approaches to make uncertainty presentation and model validation as an integral part of their studies. In light of the issues of handling uncertainty, incorporating it in the discussion of results could assist readers in their interpretations and facilitate the practical application of geospatial approaches for policy making towards improving RMNCH in LMICs.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12942-020-00239-9.

Additional file 1. Search strategy and decisions for each quality criteria.

Additional file 2. PRISMA checklist.

Additional file 3. Methodological aspects of the reviewed studies.

Additional file 4. Decisions behind each screened study.

Abbreviations

RMNCH: Reproductive, maternal, newborn and child health; LMIC: Low- and middle-income countries; DHS: Demographic and Health Surveys; SDG: Sustainable Development Goals; MCMC: Markov Chain Monte Carlo; INLA: Integrated Nested Laplace Approximation; GLM: Generalized Linear Models; RMSE: Root Mean Squared Error; MSE: Mean Squared Error; MAE: Mean Absolute Error; DIC: Deviance Information Criterion; AIC: Akaike Information Criterion.

Acknowledgements

Not applicable.

Authors' contributions

LF and AJDB conceptualized and designed the study. LF and CB carried out the literature search, screening and data extraction. LF wrote the manuscript with inputs from all authors. All authors read and approved the final manuscript.

Funding

This study was funded by Wellcome Trust [Grant Number: 101815/Z/13/Z]; Bill and Melinda Gates Foundation [through the Countdown to 2030 initiative, OPP1148933]; and Associação Brasileira de Saúde Coletiva (ABRASCO).

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] International Center for Equity in Health, Universidade Federal de Pelotas, Pelotas, Brazil. [2] Post-Graduate Program in Epidemiology, Universidade Federal

de Pelotas, Pelotas, Brazil. [3] WorldPop, Geography and Environmental Science, University of Southampton, Southampton, UK.

Received: 30 June 2020 Accepted: 5 October 2020
Published online: 13 October 2020

References

1. UN/WHO. Survive, Thrive, Transform The Global Strategy for Women's, Children's and Adolescents' Health (2016-2030). 2016.
2. Boerma T, Requejo J, Victora CG, Amouzou A, George A, Agyepong I, et al. Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health. Lancet. 2018;391(10129):1538–48. https://doi.org/10.1016/S0140-6736(18)30104-1.
3. Faye CM, Wehrmeister FC, Melesse DY, Mutua MKK, Maïga A, Taylor CM, et al. Large and persistent subnational inequalities in reproductive, maternal, newborn and child health intervention coverage in sub-Saharan Africa. BMJ Glob Heal. 2020;5(1):e002232. https://doi.org/10.1136/bmjgh-2019-002232.
4. Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nature. 2016;540(7632):189–91. https://doi.org/10.1038/540189a.
5. Hancioglu A, Arnold F. Measuring Coverage in MNCH: Tracking Progress in Health for Women and Children Using DHS and MICS Household Surveys. Madise N, editor. PLoS Med. 2013;10(5):e1001391. https://doi.org/10.1371/journal.pmed.1001391.
6. Rahman A. Estimating small area health-related characteristics of populations: a methodological review. Geospat Health. 2017;12(1). https://doi.org/10.4081/gh.2017.495.
7. Rao JNK. Small-Area Estimation. Wiley StatsRef Stat Ref Online. Chichester, UK: John Wiley & Sons, Ltd; 2017. p. 1–8. https://doi.org/10.1002/9781118445112.stat03310.pub2.
8. Ebener S, Guerra-Arias M, Campbell J, Tatem AJ, Moran AC, Amoako Johnson F, et al. The geography of maternal and newborn health: The state of the art. Int J Health Geogr. 2015;14(1). https://doi.org/10.1186/s12942-015-0012-x.
9. Manda S, Haushona N, Bergquist R. A Scoping Review of Spatial Analysis Approaches Using Health Survey Data in Sub-Saharan Africa. Int J Environ Res Public Health. 2020. 17(9). https://doi.org/10.3390/ijerph17093070.
10. Gebreslasie MT. A review of spatial technologies with applications for malaria transmission modelling and control in Africa. Geospat Health. 2015;10(2):328.
11. Rahman A. Estimating small area health-related characteristics of populations: a methodological review. Geospat Health. 2017;12(1):495. https://doi.org/10.4081/gh.2017.495.
12. Mayala BK, Dontamsetti T, Fish TD, Croft TN. Interpolation of DHS survey data at subnational administrative Level 2 [Internet]. DHS Spat. Anal. Reports No. 17. Rockville, Maryland, USA; 2019.
13. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4.
14. Fantom NJ, Serajuddin U. The World Bank's classification of countries by income (English) [Internet]. Washington. 2016. Report No.: WPS 7528.
15. Porritt K, Gomersall J, Lockwood C. JBI's Systematic Reviews. AJN, Am J Nurs. 2014;114(6):47–52. https://doi.org/10.1097/01.NAJ.0000450430.97383.64.
16. Murray C, Newby H. Data resource profile: United Nations children's fund (UNICEF). Int J Epidemiol. 2012;41(6):1595–601. https://doi.org/10.1093/ije/dys185.
17. Johns Hopkins University. Performance Monitoring for Action [Internet]. 2020https://www.pmadata.org/about. Accessed 2020 Sep 18.
18. Corsi DJ, Neuman M, Finlay JE, Subramanian SV. Demographic and health surveys: A profile. Int J Epidemiol. 2012;41(6):1602–13. https://doi.org/10.1093/ije/dys184.
19. Burgert CR, Colston J, Roy T, Zachary B. Geographic displacement procedure and GEOREFERENCED data release policy for the demographic and health surveys. Maryland: Calverton; 2013.
20. Gething P, Tatem A, Bird T, Burgert-Brucker CR. Creating spatial interpolation surfaces with DHS data [Internet]. DHS Spat. Anal. Reports No. 11. USAID. 2015. https://doi.org/10.13140/rg.2.1.1938.5362.
21. Congdon PD. Bayesian Hierachical Models with applications using R. 2nd ed. Chapman and Hall/CRC; 2019.
22. Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in practice. Chapman and Hall/CRC; 1995.
23. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Statistical Methodol. 2009;71(2):319–92. https://doi.org/10.1111/j.1467-9868.2008.00700.x.
24. Liyew AM, Kebede SA, Agegnehu CD, Teshale AB, Alem AZ, Yeshaw Y, et al. Spatiotemporal patterns of anemia among lactating mothers in Ethiopia using data from Ethiopian Demographic and Health Surveys (2005, 2011 and 2016). PLoS ONE. 2020;15(8):e0237147. https://doi.org/10.1371/journal.pone.0237147.
25. Burke M, Heft-Neal S, Bendavid E. Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. Lancet Glob Heal. 2016;4(12):e936–45. https://doi.org/10.1016/S2214-109X(16)30212-1.
26. Bhatt S, Cameron E, Flaxman SR, Weiss DJ, Smith DL, Gething PW. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. J R Soc Interface. 2017. 14(134).
27. Kinyoki DK, Osgood-Zimmerman AE, Pickering BV, Schaeffer LE, Marczak LB, Lazzar-Atwood A, et al. Mapping child growth failure across low- and middle-income countries. Nature. 2020;577(7789):231–4. https://doi.org/10.1038/s41586-019-1878-8.
28. Osgood-Zimmerman A, Millear AI, Stubbs RW, Shields C, Pickering BV, Earl L, et al. Mapping child growth failure in Africa between 2000 and 2015. Nature. 2018;555(7694):41–7.
29. Wiens KE, Lindstedt PA, Blacker BF, Johnson KB, Baumann MM, Schaeffer LE, et al. Mapping geographical inequalities in oral rehydration therapy coverage in low-income and middle-income countries, 2000-2017. Lancet Glob Heal. 2020;8(8):e1038–60. https://doi.org/10.1016/S2214-109X(20)30230-8.
30. Golding N, Burstein R, Longbottom J, Browne AJ, Fullman N, Osgood-Zimmerman A, et al. Mapping under-5 and neonatal mortality in Africa, 2000–2015: a baseline analysis for the Sustainable Development Goals. Lancet. 2017;390(10108):2171–82.
31. Mosser JF, Gagne-Maynard W, Rao PC, Osgood-Zimmerman A, Fullman N, Graetz N, et al. Mapping diphtheria-pertussis-tetanus vaccine coverage in Africa, 2000–2016: a spatial and temporal modelling study. Lancet. 2019;393(10183):1843–55. https://doi.org/10.1016/S0140-6736(19)30226-0.
32. Reiner RC, Graetz N, Casey DC, Troeger C, Garcia GM, Mosser JF, et al. Variation in childhood diarrheal morbidity and mortality in Africa, 2000-2015. N Engl J Med. 2018;379(12):1128–38.
33. Reiner RC, Welgan CA, Casey DC, Troeger CE, Baumann MM, Nguyen QP, et al. Identifying residual hotspots and mapping lower respiratory infection morbidity and mortality in African children from 2000 to 2017. Nat Microbiol. 2019;4(12):2310–8.
34. Bhattacharjee NV, Schaeffer LE, Marczak LB, Ross JM, Swartz SJ, Albright J, et al. Mapping exclusive breastfeeding in Africa between 2000 and 2017. Nat Med. 2019;25(8):1205–12.
35. Reiner RC, Wiens KE, Deshpande A, Baumann MM, Lindstedt PA, Blacker BF, et al. Mapping geographical inequalities in childhood diarrhoeal morbidity and mortality in low-income and middle-income countries, 2000–2017: analysis for the Global Burden of Disease Study 2017. Lancet. 2020;395(10239):1779–801. https://doi.org/10.1016/S0140-6736(20)30114-8.
36. Hemalatha R, Pandey A, Kinyoki D, Ramji S, Lodha R, Kumar GA, et al. Mapping of variations in child stunting, wasting and underweight within the states of India: the Global Burden of Disease Study 2000–2017. E Clin Med. 2020. 22. https://doi.org/10.1016/j.eclinm.2020.100317.
37. Kinyoki DK, Ross JM, A. L-A, Munro SB, Schaeffer LE, Abbasalizad-Farhangi M, et al. Mapping local patterns of childhood overweight and wasting in low- and middle-income countries between 2000 and 2017. Nat Med. 2020. 26(5):750–9. https://doi.org/10.1038/s41591-020-0807-6.
38. Gething PW, Casey DC, Weiss DJ, Bisanzio D, Bhatt S, Cameron E, et al. Mapping Plasmodium falciparum Mortality in Africa between 1990 and

2015. N Engl J Med. 2016;375(25):2435–45. https://doi.org/10.1056/NEJMoa1606701.
39. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. Spat Spatiotemporal Epidemiol. 2013;4:33–49. https://doi.org/10.1016/j.sste.2012.12.001.
40. Ruktanonchai CW, Ruktanonchai NW, Nove A, Lopes S, Pezzulo C, Bosco C, et al. Equality in maternal and newborn health: Modelling geographic disparities in utilisation of care in five East African countries. PLoS ONE. 2016;11(8):e0162006. https://doi.org/10.1371/journal.pone.0162006.
41. Ruktanonchai CW, Nieves JJ, Ruktanonchai NW, Nilsen K, Steele JE, Matthews Z, et al. Estimating uncertainty in geospatial modelling at multiple spatial resolutions: the pattern of delivery via caesarean section in Tanzania. BMJ Glob Heal. 2020;4(Suppl 5):e002092. https://doi.org/10.1136/bmjgh-2019-002092.
42. Giardina F, Franke J, Vounatsou P. Geostatistical modelling of the malaria risk in Mozambique: effect of the spatial resolution when using remotely-sensed imagery. Geospat Health. 2015;10(2):232–8. https://doi.org/10.4081/gh.2015.333.
43. Tatem AJ. WorldPop, open data for spatial demography. Sci Data. 2017;4(1):170004. https://doi.org/10.1038/sdata.2017.4.
44. Gatrell AC, Loytonen M. GIS and Health. 1st ed. Taylor & Francis; 1998.
45. Bosco C, Alegana V, Bird T, Pezzulo C, Bengtsson L, Sorichetta A, et al. Exploring the high-resolution mapping of gender-disaggregated development indicators. J R Soc Interface. 2017. 14(129).
46. Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 2005;30:79–82. https://doi.org/10.3354/cr030079.
47. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, et al. A world malaria map: plasmodium falciparum endemicity in 2007. PLoS Med. 2009;6(3):286–302. https://doi.org/10.1371/journal.pmed.1000048.
48. Noor AM, Gething PW, Alegana VA, Patil AP, Hay SI, Muchiri E, et al. The risks of malaria infection in Kenya in 2009. BMC Infect Dis. 2009;9:180.
49. Bennett A, Kazembe L, Mathanga DP, Kinyoki D, Ali D, Snow RW, et al. Mapping malaria transmission intensity in Malawi, 2000-2010. Am J Trop Med Hyg. 2013;89(5):840–9.
50. Noor AM, Clements ACA, Gething PW, Moloney G, Borle M, Shewchuk T, et al. Spatial prediction of *Plasmodium falciparum* prevalence in Somalia. Malar J. 2008. 7. https://doi.org/10.1186/1475-2875-7-159.
51. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-2010: a spatial and temporal analysis of transmission intensity. Lancet. 2014;383(9930):1739–47.
52. Noor AM, ElMardi KA, Abdelgader TM, Patil AP, Amine AAA, Bakhiet S, et al. Malaria risk mapping for control in the republic of Sudan. Am J Trop Med Hyg. 2012;87(6):1012–21.
53. Elyazar IRF, Gething PW, Patil AP, Rogayah H, Kusriastuti R, Wismarini DM, et al. *Plasmodium falciparum* malaria endemicity in Indonesia in 2010. PLoS ONE. 2011;6(6):e21315. https://doi.org/10.1371/journal.pone.0021315.
54. Pyrcz MJ, Deutsch C V. Declustering and Debiasing Nonrepresentative Sampling Declustering Polygonal Declustering. 2002. 1–25.
55. Wakefield J, Fuglstad G-A, Riebler A, Godwin J, Wilson K, Clark SJ. Estimating under-five mortality in space and time in a developing world context. Stat Methods Med Res. 2019;28(9):2614–34. https://doi.org/10.1177/0962280218767988.
56. Noor AM, Alegana VA, Patil AP, Moloney G, Borle M, Yusuf F, et al. Mapping the receptivity of malaria risk to plan the future of control in Somalia. BMJ Open. 2012. 2(4). https://doi.org/10.1136/bmjopen-2012-001160.
57. Yankson R, Anto EA, Chipeta MG. Geostatistical analysis and mapping of malaria risk in children under 5 using point-referenced prevalence data in Ghana. Malar J. 2019. 18(1). https://doi.org/10.1186/s12936-019-2709-y.
58. Macharia PM, Giorgi E, Noor AM, Waqo E, Kiptui R, Okiro EA, et al. Spatio-temporal analysis of Plasmodium falciparum prevalence to understand the past and chart the future of malaria control in Kenya. Malar J. 2018. 17(1). https://doi.org/10.1186/s12936-018-2489-9.
59. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science (80-). 1974. 185(4157):1124–31. https://doi.org/10.1126/science.185.4157.1124.
60. MacEachren AM, Robinson A, Hopper S, Gardner S, Murray R, Gahegan M, et al. Visualizing geospatial information uncertainty: what we know and what we need to know. Cartogr Geogr Inf Sci. 2005;32(3):139–60. https://doi.org/10.1559/1523040054738936.
61. Kleinschmidt I, Bagayoko M, Clarke GP, Craig M, Le Sueur D. A spatial statistical approach to malaria mapping. Int J Epidemiol. 2000;29(2):355–61.
62. Kleinschmidt I, Omumbo J, Briet O, van de Giesen N, Sogoba N, Mensah NK, et al. An empirical malaria distribution map for West Africa. Trop Med Int Heal. 2001;6(10):779–86. https://doi.org/10.1046/j.1365-3156.2001.00790.x.
63. Robinson TP. Spatial statistics and geographical information systems in epidemiology and public health. 2000. p. 81–128. https://doi.org/10.1016/s0065-308x(00)47007-7.
64. Afrane YA, Githeko AK, Yan G. The ecology of Anopheles mosquitoes under climate change: case studies from the effects of deforestation in East African highlands. Ann N Y Acad Sci. 2012;1249(1):204–10. https://doi.org/10.1111/j.1749-6632.2011.06432.x.
65. Craig MH, Sharp BL, Mabaso MLH, Kleinschmidt I. Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. Int J Health Geogr. 2007;6:44.
66. Gosoniu L, Vounatsou P, Sogoba N, Maire N, Smith T. Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. Comput Stat Data Anal. 2009;53(9):3358–71. https://doi.org/10.1016/j.csda.2009.02.022.
67. Gosoniu L, Veta AM, Vounatsou P. Bayesian geostatistical modeling of Malaria Indicator Survey data in Angola. PLoS ONE. 2010;5(3):e9322.
68. Riedel N, Vounatsou P, Miller JM, Gosoniu L, Chizema-Kawesha E, Mukonka V, et al. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). Malar J. 2010. 9(1). https://doi.org/10.1186/1475-2875-9-37.
69. Giardina F, Gosoniu L, Konate L, Diouf MB, Perry R, Gaye O, et al. Estimating the burden of malaria in Senegal: Bayesian zero-inflated binomial geostatistical modeling of the MIS 2008 data. PLoS ONE. 2012;7(3):e32625.
70. Giardina F, Kasasa S, Sie A, Utzinger J, Tanner M, Vounatsou P. Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: a spatial and temporal analysis. Lancet Glob Heal. England; 2014. 2(10):e601-15. https://doi.org/10.1016/s2214-109x(14)70300-6.
71. Raso G, Schur N, Utzinger J, Koudou BG, Tchicaya ES, Rohner F, et al. Mapping malaria risk among children in Cote d'Ivoire using Bayesian geo-statistical models. Malar J. 2012. 11. https://doi.org/10.1186/1475-2875-11-160.
72. Gosoniu L, Msengwa A, Lengeler C, Vounatsou P. Spatially explicit burden estimates of malaria in Tanzania: bayesian geostatistical modeling of the malaria indicator survey data. PLoS ONE. 2012;7(5):e23966.
73. Samadoulougou S, Maheu-Giroux M, Kirakoya-Samadoulougou F, De Keukeleire M, Castro MC, Robert A. Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. Parasites and Vectors. 2014. 7(1). https://doi.org/10.1186/1756-3305-7-350.
74. Onyiri N. Estimating malaria burden in Nigeria: a geostatistical modelling approach. Geospat Health. 2015;10(2):306.
75. Adigun AB, Gajere EN, Oresanya O, Vounatsou P. Malaria risk in Nigeria: Bayesian geostatistical modelling of 2010 malaria indicator survey data. Malar J. 2015;14:156.
76. Diboulo E, Sié A, Vounatsou P. Assessing the effects of malaria interventions on the geographical distribution of parasitaemia risk in Burkina Faso. Malar J. 2016. 15(1). https://doi.org/10.1186/s12936-016-1282-x.
77. Ssempiira J, Nambuusi B, Kissa J, Agaba B, Makumbi F, Kasasa S, et al. The contribution of malaria control interventions on spatio-temporal changes of parasitaemia risk in Uganda during 2009–2014. Parasit Vectors. 2017;10(1):450.
78. Ssempiira J, Nambuusi B, Kissa J, Agaba B, Makumbi F, Kasasa S, et al. Geostatistical modelling of malaria indicator survey data to assess the effects of interventions on the geographical distribution of malaria

prevalence in children less than 5 years in Uganda. PLoS One. 2017. 12(4). https://doi.org/10.1371/journal.pone.0174948.
79. Kang SY, Battle KE, Gibson HS, Ratsimbasoa A, Randrianarivelojosia M, Ramboarina S, et al. Spatio-temporal mapping of Madagascar's Malaria Indicator Survey results to assess *Plasmodium falciparum* endemicity trends between 2011 and 2016. BMC Med. 2018. 16(1). https://doi.org/10.1186/s12916-018-1060-4.
80. Massoda Tonye SG, Kouambeng C, Wounang R, Vounatsou P. Challenges of DHS and MIS to capture the entire pattern of malaria parasite risk and intervention effects in countries with different ecological zones: The case of Cameroon. Malar J. 2018. 17(1). https://doi.org/10.1186/s12936-018-2284-7.
81. Snow RW, Sartorius B, Kyalo D, Maina J, Amratia P, Mundia CW, et al. The prevalence of *Plasmodium falciparum* in sub-Saharan Africa since 1900. Nature. 2017;550(7677):515–8. https://doi.org/10.1038/nature24059.
82. Chipeta MG, Giorgi E, Mategula D, Macharia PM, Ligomba C, Munyenyembe A, et al. Geostatistical analysis of Malawi's changing malaria transmission from 2010 to 2017. Wellcome open Res. 2019;4:57.
83. Weiss DJ, Lucas TCD, Nguyen M, Nandi AK, Bisanzio D, Battle KE, et al. Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–2017: a spatial and temporal modelling study. Lancet. 2019;394(10195):322–31. https://doi.org/10.1016/S0140-6736(19)31097-9.
84. MARA–Mapping Malaria Risk in Africa [Internet]. 2020. https://www.mara-database.org//login.html. Accessed 2020 Jun 24.
85. Hay SI, Snow RW. The Malaria Atlas Project: Developing Global Maps of Malaria Risk. PLoS Med. 2006;3(12):e473. https://doi.org/10.1371/journal.pmed.0030473.
86. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the Plasmodium falciparum parasite rate. Malar J. 2007;6(1):131. https://doi.org/10.1186/1475-2875-6-131.
87. Kazembe LN, Kleinschmidt I, Holtz TH, Sharp BL. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int J Health Geogr. 2006;5:41.
88. Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. Hill K, editor. PLoS Med. 2010. 7(4):e1000253. https://doi.org/10.1371/journal.pmed.1000253.
89. SACHS JD. From millennium development goals to sustainable development goals. Lancet. 2012. 379(9832):2206–11.
90. Burstein R, Henry NJ, Collison ML, Marczak LB, Sligar A, Watson S, et al. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. Nature. 2019;574(7778):353–8. https://doi.org/10.1038/s41586-019-1545-0.
91. Chin B, Montana L, Basagaña X. Spatial modeling of geographic inequalities in infant and child mortality across Nepal. Health Place. 2011;17(4):929–36.
92. Dwyer-Lindgren L, Squires ER, Teeple S, Ikilezi G, Allen Roberts D, Colombara DV, et al. Small area estimation of under-5 mortality in Bangladesh, Cameroon, Chad, Mozambique, Uganda, and Zambia using spatially misaligned data. Popul Health Metr. 2018;16(1):13.
93. Dwyer-Lindgren L, Kakungu F, Hangoma P, Ng M, Wang H, Flaxman AD, et al. Estimation of district-level under-5 mortality in Zambia using birth history data, 1980-2010. Spat Spatiotemporal Epidemiol Netherlands. 2014;11:89–107. https://doi.org/10.1016/j.sste.2014.09.002.
94. Asiimwe JB, Jehopio P, Atuhaire LK, Mbonye AK. Examining small area estimation techniques for public health intervention: lessons from application to under-5 mortality data in Uganda. J Public Health Policy. 2011. 32(1):1–14; discussion 15.
95. Mercer LD, Wakefield J, Pantazis A, Lutambi AM, Masanja H, Clark S. Space-time smoothing of complex survey data: small area estimation for child mortality. Ann Appl Stat. 2015;9(4):1889–905.
96. Macharia PM, Giorgi E, Thuranira PN, Joseph NK, Sartorius B, Snow RW, et al. Sub national variation and inequalities in under-five mortality in Kenya since 1965. BMC Public Health. 2019;19(1):146. https://doi.org/10.1186/s12889-019-6474-1.
97. Bhutta ZA, Das JK, Rizvi A, Gaffey MF, Walker N, Horton S, et al. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet. 2013;382(9890):452–77. https://doi.org/10.1016/S0140-6736(13)60996-4.
98. Mia MN, Rahman MS, Roy PK. Sociodemographic and geographical inequalities in under- and overnutrition among children and mothers in Bangladesh: A spatial modelling approach to a nationally representative survey. Public Health Nutr. 2018;21(13):2471–81. https://doi.org/10.1017/S1368980018000988.
99. Akseer N, Bhatti Z, Mashal T, Soofi S, Moineddin R, Black RE, et al. Geospatial inequalities and determinants of nutritional status among women and children in Afghanistan: an observational study. Lancet Glob Heal. 2018;6(4):e447–59. https://doi.org/10.1016/S2214-109X(18)30025-1.
100. Fujii T. Micro-Level Estimation of Child Undernutrition Indicators in Cambodia. World Bank Econ Rev. 2010;24(3):520–53. https://doi.org/10.1093/wber/lhq016.
101. Catalán Nájera HE. Small-area estimates of stunting. Mexico 2010: Based on a hierarchical Bayesian estimator. Spat Spatiotemporal Epidemiol. 2019. 29:1–11. https://doi.org/10.1016/j.sste.2019.01.001.
102. Sohnesen TP, Ambel AA, Fisker P, Andrews C, Khan Q. Small area estimation of child undernutrition in Ethiopian woredas. PLoS ONE. 2017;12(4):e0175445.
103. Magalhaes RJS, Clements ACA. Mapping the Risk of Anaemia in Preschool-Age Children: The Contribution of Malnutrition, Malaria, and Helminth Infections in West Africa. PLOS Med. 2011.8(6). https://doi.org/10.1371/journal.pmed.1000438.
104. Peck M, Gacic-Dobo M, Diallo MS, Nedelec Y, Sodha SS, Wallace AS. Global Routine Vaccination Coverage, 2018. MMWR Morb Mortal Wkly Rep. 2019;68(42):937–42. https://doi.org/10.15585/mmwr.mm6842a1.
105. CDC. Global Measles Outbreaks [Internet]. 2020https://www.cdc.gov/globalhealth/measles/globalmeaslesoutbreaks.htm. Accessed 2020 May 24.
106. World Health Organization. Measles – Global situation [Internet]. Dis. outbreak news. 2019https://www.who.int/csr/don/26-november-2019-measles-global_situation/en/. Accessed 2020 May 24.
107. Zucker JR, Rosen JB, Iwamoto M, Arciuolo RJ, Langdon-Embry M, Vora NM, et al. Consequences of Undervaccination—Measles Outbreak, New York City, 2018–2019. N Engl J Med. 2020;382(11):1009–17. https://doi.org/10.1056/NEJMoa1912514.
108. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. Vaccine. 2018. 36(12):1583–91. https://doi.org/10.1016/j.vaccine.2018.02.020.
109. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. Mapping vaccination coverage to explore the effects of delivery mechanisms and inform vaccination strategies. Nat Commun. 2019;10(1):1633. https://doi.org/10.1038/s41467-019-09611-1.
110. Utazi CE, Wagai J, Pannell O, Cutts FT, Rhoda DA, Ferrari MJ, et al. Geospatial variation in measles vaccine coverage through routine and campaign strategies in Nigeria: Analysis of recent household surveys. Vaccine. 2020;38(14):3062–71. https://doi.org/10.1016/j.vaccine.2020.02.070.
111. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Nilsen K, Takahashi S, et al. A spatial regression model for the disaggregation of areal unit based data to high-resolution grids with application to vaccination coverage mapping. Stat Methods Med Res. 2019;28(10):3226–41.
112. Takahashi S, Metcalf CJE, Ferrari MJ, Moss WJ, Truelove SA, Tatem AJ, et al. Reduced vaccination and the risk of measles and other childhood infections post-Ebola. Science (80-). 2015. 347(6227):1240–2. https://doi.org/10.1126/science.aaa3438.
113. Takahashi S, Metcalf CJE, Ferrari MJ, Tatem AJ, Lessler J. The geography of measles vaccination in the African Great Lakes region. Nat Commun. 2017;8(1):15585. https://doi.org/10.1038/ncomms15585.
114. Pramanik S, Muthusamy N, Gera R, Laxminarayan R. Vaccination coverage in India: a small area estimation approach. Vaccine. 2015;33(14):1731–8.
115. Mercer LD, Lu F, Proctor JL. Sub-national levels and trends in contraceptive prevalence, unmet need, and demand for family planning in Nigeria with survey uncertainty. BMC Public Health. 2019;19(1):1752.
116. Li Q, Louis TA, Liu L, Wang C, Tsui AO. Subnational estimation of modern contraceptive prevalence in five sub-Saharan African countries: a Bayesian hierarchical approach. BMC Public Health. 2019. 19. https://doi.org/10.1186/s12889-019-6545-3.
117. Johnson FA, Padmadas SS, Chandra H, Matthews Z, Madise NJ. Estimating unmet need for contraception by district within Ghana: An

application of small-area estimation techniques. Popul Stud J Demogr. 2012;66(2):105–22. https://doi.org/10.1080/00324728.2012.678585.
118. Neal S, Ruktanonchai C, Chandra-Mouli V, Matthews Z, Tatem AJ. Mapping adolescent first births within three east African countries using data from Demographic and Health Surveys: Exploring geospatial methods to inform policy. Reprod Health. Reproductive Health. 2016. 13(1):1–29. https://doi.org/10.1186/s12978-016-0205-1.
119. Neal S, Ruktanonchai CW, Chandra-Mouli V, Harvey C, Matthews Z, Raina N, et al. Using geospatial modelling to estimate the prevalence of adolescent first births in Nepal. BMJ Glob Heal. 2019;4:e000763.
120. Amoako Johnson F, Padmadas SS, Chandra H, Matthews Z, Madise NJ. Estimating unmet need for contraception by district within Ghana: an application of small-area estimation techniques. Popul Stud. 2012;66(2):105–22.
121. Ruktanonchai CW, Nilsen K, Alegana VA, Bosco C, Ayiko R, Seven Kajeguka AC, et al. Temporal trends in spatial inequalities of maternal and newborn health services among four east African countries, 1999-2015. BMC Public Health. 2018. 18(1):1–13. https://doi.org/10.1186/s12889-018-6241-8.
122. Das S, Chandra H, Saha UR. District level estimates and mapping of prevalence of diarrhoea among under-five children in Bangladesh by combining survey and census data. PLoS ONE. 2019;14(2):e0211062.
123. Kazembe LN, Muula AS, Simoonga C. Joint spatial modelling of common morbidities of childhood fever and diarrhoea in Malawi. Heal Place. 2009;15(1):165–72. https://doi.org/10.1016/j.healthplace.2008.03.009.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Additional file 1 - Search strategy and decisions for each quality criteria

Search strategy

Search strategy for MEDLINE, LILACS, Web of Science and Scielo

(health OR epidemiology) AND (geostatistical OR geo-statistical OR “spatial modeling” OR “spatial modelling” OR “high-resolution mapping” OR geospatial OR “small area estimation” OR “small area estimates” OR “spatial interpolation”)

Search strategy for Scopus

TITLE-ABS-KEY ((health OR epidemiology) AND (geostatistic OR "geo statistic" OR "spatial modeling" OR "high-resolution mapping" OR geospatial OR "small area estimation" OR "spatial interpolation")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (EXCLUDE (SUBJAREA , "AGRI") OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA , "MATH") OR EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA , "BUSI") OR EXCLUDE (SUBJAREA , "VETE") OR EXCLUDE (SUBJAREA , "CHEM") OR EXCLUDE (SUBJAREA , "ARTS") OR EXCLUDE (SUBJAREA , "PHYS") OR EXCLUDE (SUBJAREA , "PSYC") OR EXCLUDE (SUBJAREA , "NEUR") OR EXCLUDE (SUBJAREA , "ENER") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA , "DENT"))

Decisions on quality assessment criteria

Were study participants sampled in an appropriate way? Studies using several data sources in which most of the data come from structured national health surveys were classified as ‘yes’.

Was the sample size adequate? All studies were classified as ‘Not applicable’ because the rationale behind small area estimation assumes insufficient sample sizes for direct estimation.

Were the study subjects and the setting described in detail? Studies covering multiple countries were classified as ‘Not applicable’.

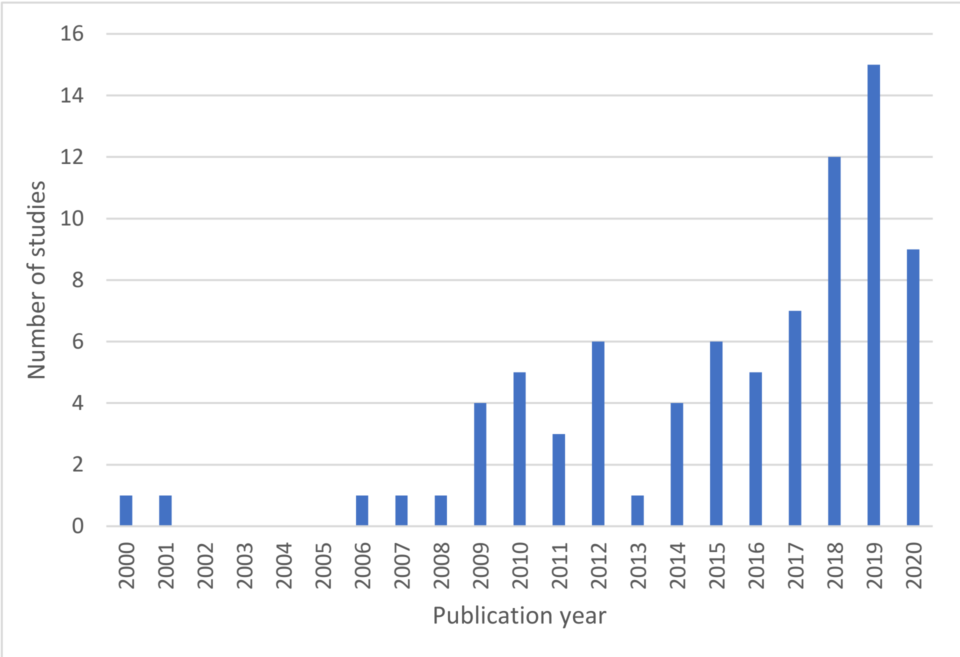
Was data analysis conducted with sufficient coverage of the identified sample? All studies were classified as ‘Not applicable’ because most (if not all) of them are based on secondary data and this information is usually presented by the responsible for the data collection process.

Was the condition measured in a standard, reliable way for all participants? All studies were classified as ‘Not applicable’ because most (if not all) of them are based on secondary data and it is uncommon for studies to discuss the validity of the data collection methods of the data sources.

Was there appropriate statistical analysis? All studies not reporting uncertainty measures were classified as “no” based on the criteria definition.

Was the response rate adequate, and if not, was the low response rate managed appropriately? All studies were classified as 'Not applicable' because most (if not all) of them are based on secondary data and it is uncommon for studies report the response rate of their data sources, especially when multiple data sources are utilized.

Distribution of reviewed studies over time



Additional file 2 – PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	1-2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3-4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6

Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	6
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	6-18
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	AF1
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	18-20
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	20
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	18-20

FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	21

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Additional file 3 - Methodological aspects of the reviewed studies

Available at: https://static-content.springer.com/esm/art%3A10.1186%2Fs12942-020-00239-9/MediaObjects/12942_2020_239_MOESM3_ESM.xlsx

Additional file 4 - Decisions behind each screened study

Available at: https://static-content.springer.com/esm/art%3A10.1186%2Fs12942-020-00239-9/MediaObjects/12942_2020_239_MOESM4_ESM.xlsx

Article 2

Mapping inequalities in a composite reproductive, maternal, neonatal and child health indicator in Peru using geospatial modeling

Leonardo Z. Ferreira[1 2*], C. Edson Utazi[3], Luis Huicho[4 5 6], Kristine Nilsen[3], Fernando P. Hartwig[2], Andrew J. Tatem[3], Aluisio JD Barros[1 2]

1. International Center for Equity in Health, Universidade Federal de Pelotas, Pelotas, Brazil

2. Post-Graduate Program in Epidemiology, Universidade Federal de Pelotas, Pelotas, Brazil

3. WorldPop, Geography and Environmental Science, University of Southampton, Southampton, UK

4. Centro de Investigación en Salud Materna e Infantil, Universidad Peruana Cayetano Heredia, Lima, Peru

5. Centro de Investigación para el Desarrollo Integral y Sostenible, Universidad Peruana Cayetano Heredia, Lima, Peru

5. School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

***Corresponding author**: International Center for Equity in Health, Universidade Federal de Pelotas, R. Marechal Deodoro 1160, Centro, Pelotas, Brazil. E-mail: lferreira@equidade.org

Word count: 2901 words.

Abstract

Background: The composite coverage index (CCI) provides an integrated perspective towards universal health coverage in the context of reproductive, maternal, newborn and child health. However, the sample design of the surveys is not sufficient for an in-depth analysis from a geographical perspective. This study aims to describe and compare the CCI coverage at multiple resolutions in Peru to support decision-makers with actionable information at local scale.

Methods: Using a model-based geostatistical approach, we generated estimates for all eight indicators of the CCI (which were further combined using the index formula) for the departments, provinces, and areas of 5 x 5 km of Peru using data from two national household surveys carried out in 2018 and 2019 and geospatial covariates. Models were fit following a Bayesian framework using INLA-SPDE and assessed using validation metrics and comparisons at the department-level.

Results: Coverage in the provinces throughout the coast were consistently higher than the remainder of the country. Areas in the north and east of the country, especially within the Amazon jungle, were found to have the largest gaps between and within provinces. These gaps are otherwise masked when looking at department-level only.

Conclusions: Our study highlights areas of low CCI coverage within departments and provinces of Peru, showcasing the importance of estimates at high-resolution to unveil inequalities within highly heterogeneous areas. Our results constitute a valuable guide for local policy makers and managers to focus efforts in disadvantaged areas.

Keywords: geospatial modeling, child health, woman's health, composite coverage index, Peru

Key messages

Geospatial modeling techniques allow the use of estimates for small areas in which direct estimation from national household surveys is not possible or yield imprecise results.

Local managers and decision-makers benefit of information for smaller areas, such as provinces, since the planning and allocation of resources is often done at local scale.

Increasing the resolution reveals striking inequalities from a geographical perspective masked by aggregation, highlighting the most vulnerable areas and subgroups of the population.

Introduction

Peru has shown tremendous progress in improving the health and survival of women and children in the past few decades (1). Under-five mortality rates and undernutrition dropped by over 50% since 2000, mainly due to the equitable increase in the coverage of reproductive, maternal, newborn and child health (RMNCH) indicators, better water and sanitation conditions, along with improvements in the social determinants of health (2). These efforts elevated the country to a prime position in the pursuit of universal health coverage (UHC) in terms of access to the full range of quality health services without undue financial burden.

Since the concept of UHC relies on a broad set of services and interventions, monitoring its progress requires data on multiple RMNCH indicators. And even overlooking the difficulty of reporting substantial amounts of data, visualizing and advocating for dozens of indicators hampers the prioritization of areas and subgroups that are farther from receiving a comprehensive assistance. In order to account for several indicators and present an integrated summary measure, the composite coverage index (CCI) was created and has been widely used as a proxy for tracking UHC in low- and middle-income countries in the context of RMNCH (3–5). The CCI is a weighted average of eight essential preventive and curative interventions along the continuum of care, covering four stages including reproductive, pregnancy, newborn and child health. Its composition has proven to be robust as the inclusion of other important interventions have shown to have little impact in the estimates (6). Also, its strong associations with under-five mortality rate and stunting further support that the CCI can capture adequately the combined effect of health interventions (7).

Like the majority of low- and middle-income countries, Peru relies heavily on information from national health surveys to monitor the progress of many RMNCH interventions, allowing data-driven actions to increase coverage and reduce inequalities (8–10). These actions are more effective when stakeholders and policy planners have available data disaggregated at local levels, where policy is ultimately implemented (11,12). However, the sampling design of the national surveys only provides reliable estimates for large subnational divisions, as further geographical disaggregation would require much larger (often prohibitively so) sample sizes. Alternatively, indirect estimates can be derived for smaller areas using geospatial modeling approaches, as previous studies have done for RMNCH outcomes in recent years (13–15). These strategies combine the georeferenced data from the surveys with relevant geospatial covariates, while also taking advantage of spatial correlation, to predict small area estimates.

While a few studies have generated estimates for individual interventions or health outcomes at global scale in which Peru was included, such as malnutrition (16), mortality (17) and diarrhea management (18), no studies on RMNCH interventions were exclusively focused on Peru. Based on fine-scale estimates generated using geospatial modeling techniques, this study aims to describe the CCI coverage at province and grid-level in Peru, enabling local managers to identify and act on areas in need of prioritization.

Methods

The following sections describe each stage of the modelling process. Further details can be found in the supplementary materials.

Study area

Peru is an upper-middle income country located in the South American continent. Its lands cover around 1.28 million km^2 making it the 19th largest country in the world with a total population of nearly 31 million (19). It shares borders with Ecuador, Colombia, Brazil, Bolivia, Chile, and the

Pacific Ocean. The country is divided into 25 first administrative units (24 departments plus the Callao province) which are subdivided into 196 provinces, and further into districts. The geography of Peru is often divided into three main ecological zones known as a) coast, a semiarid margin bordered by the Pacific Ocean, b) highlands or the Andean mountains, a climatic diverse area separating the other two ecological zones from north to south, and c) jungle which is the most extensive zone mostly covered by the Amazon rainforest (20).

Composite coverage index data

Carried out annually since 2004, the Encuesta Demográfica y de Salud Familiar (ENDES) is a household survey designed to provide estimates at the national and departmental levels for several health and nutritional indicators for women and children in Peru. The 2018 ENDES survey carried out a multi-stage sampling process by selecting 3,254 enumeration areas (EAs), or primary sampling units (also known as clusters – the unit of analysis in this study), proportionally distributed in all departments, followed by 36,760 households in the second stage. Similarly, the 2019 survey sampled 36,745 households in 3,254 EAs, totalizing 73,505 households within 6,508 EAs (Figure 1). More details on the sampling methodology can be found in the surveys' reports (21,22).

To increase the sample sizes, we combined data from the 2018 and 2019 ENDES surveys and used them to calculate each of the indicators that are part of the CCI. The CCI is a weighted average of eight essential maternal and child health interventions and comprises the four stages of the continuum of care. Its formula is given by

$$CCI = \frac{1}{4}\left(DFPSm + \frac{SBA + ANC4}{2} + \frac{2(DPT3) + BCG + MSL}{4} + \frac{ORS + CAREP}{2}\right)$$

where the interventions are: demand for family planning satisfied with modern methods (DFPSm), skilled birth attendant (SBA), at least four antenatal care visits (ANC4), one dose of bacille Calmette-Guérin vaccine (BCG), three doses of diphtheria-pertussis-tetanus vaccine

(DPT3), one dose of measles vaccine (MSL), oral rehydration salts for diarrhoea (ORS), and care-seeking for suspected childhood pneumonia (CAREP) (6). The complete definition on the calculation of each indicator can be found in the supplementary table 1.

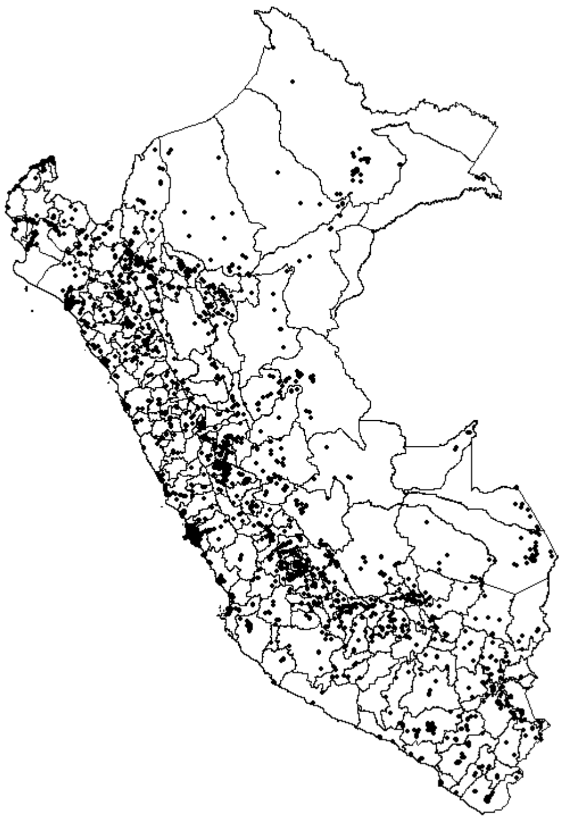


Figure 1 - Geographical distribution of cluster locations for ENDES 2018 and 2019 surveys

The coordinates for each cluster are displaced by up to 2km in urban areas and 5km in rural areas for protect the anonymity of the respondents. This displacement was considered in the covariate extraction process by drawing a buffer and taking its mean according to the place of residence (23).

Geospatial covariates and covariate selection

A suite of 14 covariate layers known to correlate directly or indirectly with RMNCH indicators were considered as predictors for each of the eight modelled indicators. These covariates include measures of accessibility, remoteness, urbanicity, and sociodemographic characteristics that were found to be associated with health interventions in previous studies (14,24–26). Some of these covariates were surfaces obtained from satellite imagery and publicly available repositories while others were interpolated using survey and health facility data points. Further information for each of the covariates is found in the supplementary tables 2 and 3.

We carried out a covariate selection strategy divided in two stages to achieve the best fit without overparameterizing the models: 1) testing the predictors seeking the best outcome-covariate relationship, followed by 2) backward elimination process within a stepwise logistic regression, where variables were dropped starting with the ones with the highest p values until none had a p value greater than 5%. Fractional polynomials of up to the second order were tested for all predictors, as well as the logarithmic transformation, to allow for model flexibility. We also assessed the association between all covariates using Pearson's correlation and we checked any combination of predictors with a high correlation coefficient (> 0.8) to address the problem of (multi)collinearity.

Geospatial model

We followed a model-based geostatistical approach (27), similar to what was done previously (28,29), to predict estimates at areas of 5 x 5 km in Peru based on geospatial covariate information and spatial correlation. We fitted eight different models, one for each indicator, and combined their posterior distributions to obtain estimates for the CCI. Let $Y(s_i)$ be the number of individuals with a given outcome at cluster location s_i (i=1, ..., n), out of a total of $N(s_i)$ individuals sampled at the location. The model can be defined as:

$$Y(s_i) \sim Binomial\ (N(s_i), p(s_i))$$

$$logit\big(p(s_i)\big) = x(s_i)^T\,\beta + \ \omega(s_i) + \epsilon(s_i)$$

where $x(s_i)$ is a set of covariates values associated with cluster s_i, β are the corresponding regression parameters, $\omega(s_i)$ is a Gaussian spatial random effect used to capture residual spatial correlation in the model, and $\epsilon(s_i)$ is a Gaussian random effect used to model non-spatial residual variation. The geostatistical model described above followed a Bayesian framework using the integrated nested Laplace approximation with the stochastic partial differential equations (30).

We drew 1000 samples from the posterior distribution generated from each model and combined them using the CCI formula. Then, we used the combined posterior to predict a grid of 5 x 5 km covering the entire study area. The estimates and uncertainty measures were further aggregated into the first and the second administrative divisions using population layers for weighting.

Uncertainty estimates were drawn from the posterior distributions and are presented as the width of the credible intervals (difference between the 97.5 and the 2.5 percentiles).

Model validation

The validity of the estimated models was assessed using an out-of-sample cross validation strategy. Data from all indicators were divided into five folds to ensure a minimum sample size of 50 clusters within each fold. We calculated and presented the following metrics: bias (mean error), the magnitude of the error (mean absolute error - MAE) and the correlation between predicted and observed values. We also compared predicted estimates aggregated at the first administrative division to the observed estimates directly derived from the surveys.

We used Stata 16 (31) for survey data analysis and the covariate selection process and R 4.0.2 (32) for the processing of geospatial covariates, model fitting and validation.

Results

Overall, the national coverage of the CCI in the country is 71.6% (Table 1). Interventions like SBA and BCG vaccination are nearly universal in Peru with coverage above 95% (Table 1). On the other hand, treatment of diarrhea using ORS is surprisingly low with only 33.6% (Table 1). Sample sizes are large for pregnancy and reproductive health indicators, moderate for vaccines and low for treatment of childhood illnesses.

Table 1. Description of the CCI and its indicators in the sample.

Indicator	Number of clusters	Number of individuals	National coverage
Skilled birth attendant	6,383	24,513	94.6%
Antenatal care 4+ visits	6,383	24,141	96.3%
Demand for family planning satisfied with modern methods	5,296	12,059	65.5%
DPT3 vaccine coverage	4,861	8,706	85.3%
BCG vaccine coverage	4,861	8,706	95.3%
Measles vaccine coverage	4,861	8,706	80.0%
Oral rehydration salts for diarrhea	3,301	5,014	33.6%
Care-seeking for suspected pneumonia	1,639	1,956	70.1%
Composite coverage index	-	-	71.6%

Out of the 14 geospatial covariates, improved sanitation coverage was the most stable predictor as it was selected in 7 out of the 8 indicators, followed by the mean number of years of education for women, used in six indicators. Conversely, improved water coverage failed to remain as a predictor in all models and was left out of analysis. Both travel time to health facilities and urbanicity were only eligible for one model each, putting them among the least relevant covariates. The median number of covariates used to fit the models was seven, with SBA using 10 covariates and CAREP fitting the final model with a single predictor (Supplementary table 4).

The geospatial estimates for CCI coverage at the first and second administrative divisions of Peru (departments and provinces, respectively) are presented in Figure 2. Coverage ranged from

59.6% in Puno, in the south-east of the country, to 79.1% in Tumbes, in the north-west (Figure 2A). Figure 2B shows a clear pattern of higher coverage for the provinces along the coast while most of the provinces with the lowest coverages are in the jungle area. Also, substantial disparities in coverage are observed between the provinces within each department. The maximum difference between the CCI coverage in provinces of Madre de Dios, Tumbes and Ica range from 1 percentage point (p.p.) to 2.3 p.p. while these differences between the provinces of Amazonas and Ucayali go up to 20 p.p. The widest gaps between provinces are found in jungle departments but important gaps can also be seen in several departments along the coast, such as Piura, La Libertad, Ancash, and Lima.

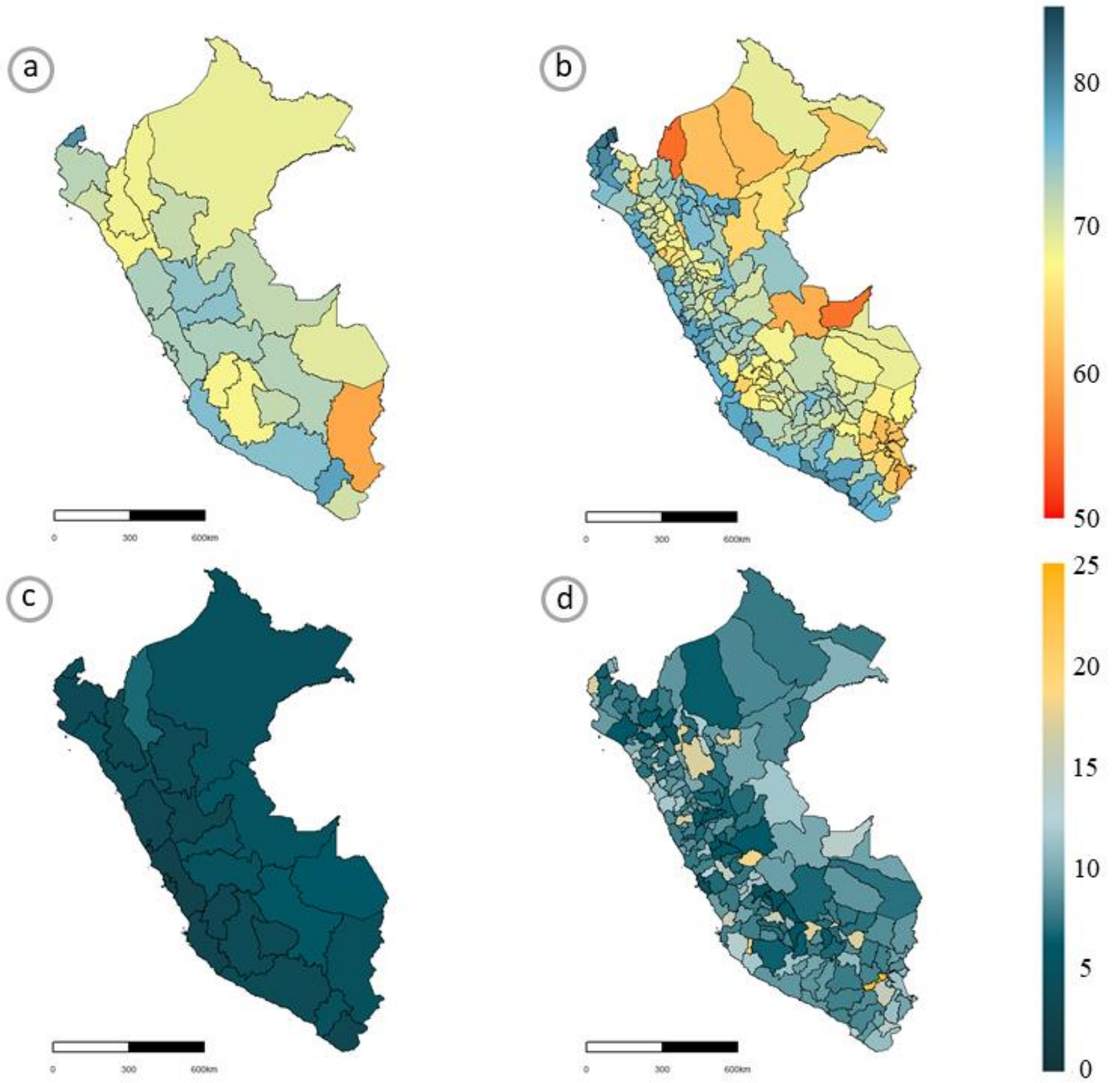


Figure 2 – Geographical distribution of the CCI in Peru for a) the 26 departments (observed data); b) the 195 provinces (predicted data), and associated uncertainty measured as c) the width of the 95% confidence intervals for departmental estimates and, d) the width of the 95% credible intervals for provincial estimates.

The same way estimates from sampled data present variability, the estimates generated by the geospatial models are sensitive to the amount of data available in each area, and the further we disaggregate, the more uncertainty we may observe. Figure 2C presents the width of the confidence intervals as a measure of uncertainty for the estimates at the first administrative division that were directly derived from the surveys. The width of the credible intervals for each of the provincial estimates generated by the geospatial models are shown in Figure 2D. The median width of the credible intervals for the provinces is 8.6 p.p., meaning that at least half of the estimates should vary no more than 4.3 p.p. around the point estimate (median coverage of 71.7%). However, estimates for the provinces colored in orange should be interpreted with more caution since those estimates could lay within a 15 to 20 p.p. interval, with a maximum interval width of 23 p.p.

Geospatial models can provide estimates for much smaller areas than the political administrative divisions of a country. Using grids of high resolution can shed light on pockets of low or high coverage otherwise masked by aggregation that may hide the most vulnerable subgroups of the population. Figure 3 presents high-resolution estimates in grids of 5 x 5 km along with uncertainty maps measured using the width of the credible intervals. Pockets of low coverage can be seen in several provinces across the jungle, while some smaller pockets exist along the highlands throughout the country. Although these estimates show a more detailed scenario, increasing the resolution also increases the uncertainty of these estimates, as presented in Figure 3B. Of note, Figures 2B and 2D are not directly comparable to Figure 3 because scales differ.

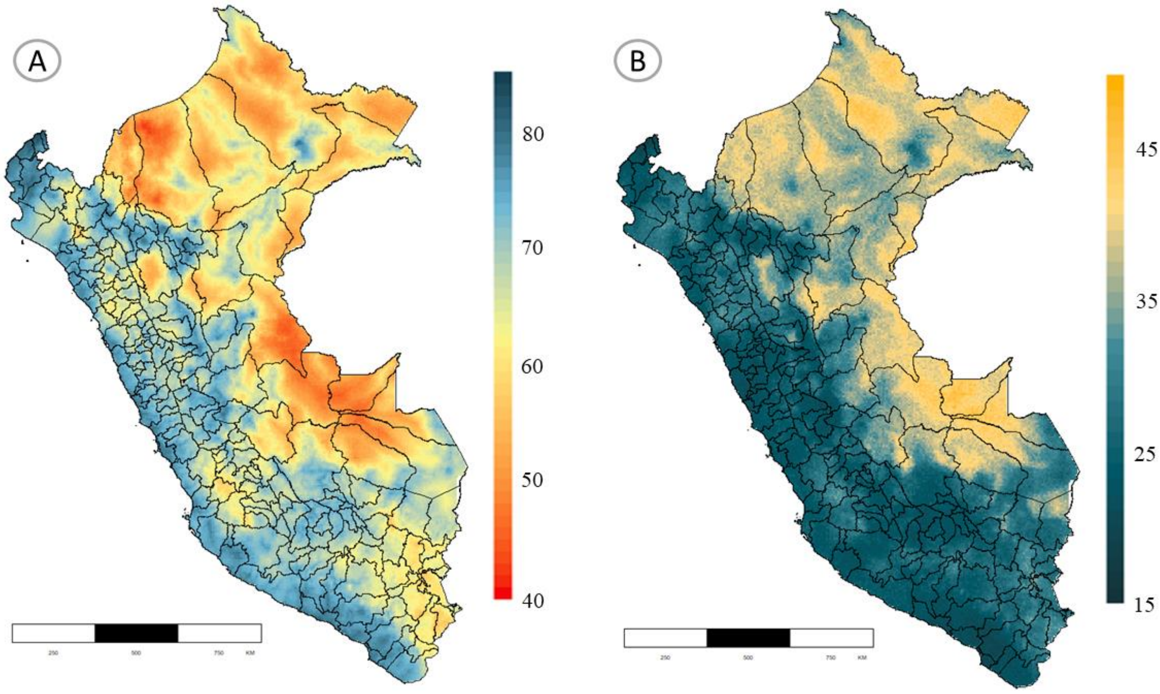


Figure 3 – a) Geographical distribution of the CCI in Peru by 5 x 5 km, and b) associated uncertainty map measured as the width of the 95% credible intervals.

The correlation of the predicted coverage against the observed cluster estimates was strong for SBA, weak for CAREP and moderate for the remainder. Bias was close to zero for all indicators and the magnitude of the errors was between 4 and 7 percentage points for BCG, SBA and ANC4, around 20 for DFPSm, MSL and DPT, and close to 30 for CAREP and ORS. We also generated estimates at the departmental-level and compared them to the published survey results, where the mean difference at the departmental-level for the CCI was 1.7% and the largest difference was 5.2% in Madre de Dios. All validation metrics and further details on the validations process are available in the supplementary tables 5 and 6.

Discussion

Geospatial models have become an important resource for assisting researchers and stakeholders in unveiling hidden areas and populations in need of prioritization in the context of RMNCH (13). Through them, it is possible to obtain information that is not available through conventional methods and can be easily used in health policy planning and decision-making. When comparing department-level estimates to province-level estimates, it is evident that very different coverage levels exist throughout the country, and even more detailed patterns are observable when moving down to the grid-level.

In general, the size of the provinces in Peru seems sufficient to accurately inform the local context in most of the country. However, this may not be true when looking at larger provinces predominantly located in the Amazon rainforest. Some of these provinces, especially in the north and east of Peru, do not stand out as low coverage in the province-level coverage map, mainly because the low coverage areas are less populated and, on average, province coverage is not particularly low. With the high-resolution map, it is possible to identify areas with very low coverage within such provinces, supporting the use of both maps as complementary resources since the provinces account for the size of the population affected and the grids focus on anyone that lives on a specific zone. By this means, local managers will have their attention drawn to these places with enough information to verify what type of action is most needed.

The rationale behind choosing the CCI rather than one or several of the many essential RMNCH interventions is simple – one composite indicator gives a broad perspective of the status of health intervention coverage (6). Looking at antenatal care or treatment of childhood diseases instead, we found distinct levels of coverage without marked geographical variation. These patterns are completely different from indicators such as demand for family planning satisfied or skilled birth attendant – both presented huge spatial heterogeneity with pockets of low and high coverage. As a composite measure, the CCI is able to highlight areas and subgroups that are struggling in multiple fronts and are far away from UHC. Additionally, being a weighted average, it is less susceptible to imprecision of some specific indicator.

Peru was able to improve substantially its RMNCH indicators and the availability and quality of the departmental level data through annual ENDES to monitor appropriately the progress of coverage and impact indicators throughout the latest few decades (33,34). Now, it faces the challenge of sustaining such a progress by tracking reliably the evolution of interventions coverage at the provincial and more local levels. The use of granular information like the CCI based on geospatial methods and high-resolution mapping may allow an increased efficiency of policy makers in the design, implementation and impact evaluation of programs and interventions for further improvement of maternal and child health, with a special focus on the most disadvantaged communities, which are located mainly within the Amazon and the Andean provinces. In particular, it may facilitate the identification of success and limiting factors at those levels, thus contributing to an effective decentralization process.

Some limitations should be considered when interpreting the estimates described in our study. All modeled estimates carry uncertainty, and it should be observed when interpreting the coverage estimates. Due to low sample sizes, indicators for remote and less populated areas as well as those related to treatment of childhood diseases can be unstable, as they are based on information from few clusters or depend on children presenting with pneumonia or diarrhea at the time of the survey. The most critical areas were concentrated in the Amazon jungle where population density is low, and many preservation areas exist. Also, increased granularity implies greater uncertainty. This phenomenon is evident when comparing the uncertainty produced in the different levels of aggregation.

Conclusions

In summary, our study presents CCI coverage at three disaggregation levels in Peru, pinpointing where are the population segments with the lowest coverage levels. It also showcases the importance of geospatial methods and high-resolution mapping in comparison to coverage estimates at administrative division level, especially where the divisions cover a large area and are highly heterogeneous. Our results constitute a valuable guide for local policy makers and managers to focus efforts in disadvantaged areas.

Ethics approval

The study used anonymized publicly available data. Ethical clearance was done by the institutions responsible for carrying out the original surveys.

Funding

This work was supported by Wellcome Trust [Grant Number: 101815/Z/13/Z]; Bill and Melinda Gates Foundation [through the Countdown to 2030 initiative, OPP1148933]; and Associação Brasileira de Saúde Coletiva (ABRASCO).

Data availability

Data on the national household surveys can be obtained from the Instituto Nacional de Estadística e Informática (https://www.inei.gob.pe/). Data sources from the covariates are listed in the supplementary material. Codes and further data can be obtained directly from the authors.

Author contributions

LF and AJDB conceptualized and designed the study. LF extracted the data, carried out the analysis, and wrote the first draft of the manuscript. All authors critically reviewed and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Conflict of interest

None declared.

References

1. PAHO/WHO, Partnership for Maternal Newborn & Child Health, World Bank and Alliance for Health Policy and Research Systems. Success Factors for Women's and Children's Health: Peru. Geneva; 2015.

2. Huicho L, Segura ER, Huayanay-Espinoza CA, de Guzman JN, Restrepo-Méndez MC, Tam Y, et al. Child health and nutrition in Peru within an antipoverty political agenda: a Countdown to 2015 country case study. Lancet Glob Heal [Internet]. 2016 Jun;4(6):e414–26. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2214109X16000851

3. Ewerling F, Wehrmeister FC, Victora CG, Raj A, McDougal L, Barros AJ. Is women's empowerment associated with coverage of RMNCH interventions in low- and middle-income countries? An analysis using a survey-based empowerment indicator, the SWPER. J Glob Health [Internet]. 2021 Mar 1;11:04015. Available from: http://jogh.org/documents/2021/jogh-11-04015.pdf

4. Amouzou A, Jiwani SS, da Silva ICM, Carvajal-Aguirre L, Maïga A, Vaz LME. Closing the inequality gaps in reproductive, maternal, newborn and child health coverage: slow and fast progressors. BMJ Glob Heal [Internet]. 2020 Jan 26;5(1):e002230. Available from: https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2019-002230

5. Barros AJD, Wehrmeister FC, Ferreira LZ, Vidaletti LP, Hosseinpoor AR, Victora CG. Are the poorest poor being left behind? Estimating global inequalities in reproductive, maternal, newborn and child health. BMJ Glob Heal [Internet]. 2020 Jan 26;5(1):e002229. Available from: https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2019-002229

6. Wehrmeister FC, Barros AJD, Hosseinpoor AR, Boerma T, Victora CG. Measuring universal health coverage in reproductive, maternal, newborn and child health: An update of the composite coverage index. Zhang S, editor. PLoS One [Internet]. 2020 Apr 29;15(4):e0232350. Available from: https://dx.plos.org/10.1371/journal.pone.0232350

7. Wehrmeister FC, Restrepo-Mendez MC, Franca GVA, Victora CG, Barros AJD. Summary indices for monitoring universal coverage in maternal and child health care. Bull World Health Organ. 2016;94(12):903–12.

8. Huicho L, Vidal-Cárdenas E, Akseer N, Brar S, Conway K, Islam M, et al. Drivers of stunting reduction in Peru: a country case study. Am J Clin Nutr [Internet]. 2020 Sep 14;112(Supplement_2):816S-829S. Available from: https://academic.oup.com/ajcn/article/112/Supplement_2/816S/5898918

9. Huicho L, Huayanay-Espinoza CA, Herrera-Perez E, Niño de Guzman J, Rivera-Ch M, Restrepo-Méndez MC, et al. Examining national and district-level trends in neonatal health in Peru through an equity lens: a success story driven by political will and societal advocacy. BMC Public Health [Internet]. 2016 Sep 12;16(S2):796. Available from: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-016-3405-2

10. Huicho L, Huayanay-Espinoza CA, Herrera-Perez E, Segura ER, Niño de Guzman J, Rivera-Ch M, et al. Factors behind the success story of under-five stunting in Peru: a district ecological multilevel analysis. BMC Pediatr [Internet]. 2017 Dec 19;17(1):29. Available from: https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-017-0790-3

11. Hosseinpoor AR, Bergen N. Area-based units of analysis for strengthening health inequality monitoring. Bull World Health Organ [Internet]. 2016 Nov 1;94(11):856–8. Available from: http://www.who.int/entity/bulletin/volumes/94/11/15-165266.pdf

12. Agyepong IA, Sewankambo N, Binagwaho A, Coll-Seck AM, Corrah T, Ezeh A, et al. The path to longer and healthier lives for all Africans by 2030: the Lancet Commission on the future of health in sub-Saharan Africa. Lancet [Internet]. 2017 Dec;390(10114):2803–59. Available from: https://linkinghub.elsevier.com/retrieve/pii/S014067361731509X

13. Ferreira LZ, Blumenberg C, Utazi CE, Nilsen K, Hartwig FP, Tatem AJ, et al. Geospatial estimation of reproductive, maternal, newborn and child health indicators: a systematic review of methodological aspects of studies based on household surveys. Int J Health Geogr [Internet]. 2020 Dec 13;19(1):41. Available from: https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-020-00239-9

14. Ruktanonchai CW, Ruktanonchai NW, Nove A, Lopes S, Pezzulo C, Bosco C, et al. Equality in maternal and newborn health: Modelling geographic disparities in utilisation of care in five East African countries. PLoS One. 2016;11(8):e0162006.

15. Ruktanonchai CW, Nilsen K, Alegana VA, Bosco C, Ayiko R, Seven Kajeguka AC, et al. Temporal trends in spatial inequalities of maternal and newborn health services among four east African countries, 1999-2015. BMC Public Health. 2018 Dec;18(1):1339.

16. Kinyoki DK, Osgood-Zimmerman AE, Pickering B V, Schaeffer LE, Marczak LB, Lazzar-Atwood A, et al. Mapping child growth failure across low- and middle-income countries. Nature [Internet]. 2020;577(7789):231–4. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077638090&doi=10.1038%2Fs41586-019-1878-8&partnerID=40&md5=2a89a102657b0e1cbc9a1ad5da85e48b

17. Burstein R, Henry NJ, Collison ML, Marczak LB, Sligar A, Watson S, et al. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. Nature [Internet]. 2019;574(7778):353–8. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073458638&doi=10.1038%2Fs41586-019-1545-0&partnerID=40&md5=17a1970c5f10decfd08811677a9ef3e7

18. Wiens KE, Lindstedt PA, Blacker BF, Johnson KB, Baumann MM, Schaeffer LE, et al. Mapping geographical inequalities in oral rehydration therapy coverage in low-income and middle-income countries, 2000-17. Lancet Glob Heal. 2020 Aug;8(8):e1038–60.

19. Instituto Nacional de Estadística e Informática. Perú: Perfil sociodemográfico [Internet]. 2017. Available from: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1539/index.html

20. Torres-Roman JS, Ronceros-Cardenas L, Valcarcel B, Arce-Huamani MA, Bazalar-Palacios J, Ybaseta-Medina J, et al. Cervical cancer mortality in Peru: regional trend analysis from 2008–2017. BMC Public Health [Internet]. 2021 Dec 26;21(1):219. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-021-10274-1

21. Instituto Nacional de Estadística e Informatica. Perú: Encuesta Demográfica y de Salud Familiar 2018. 2019.

22. Instituto Nacional de Estadística e Informatica. Perú: Encuesta Demográfica y de Salud

Familiar 2019. 2020.

23. Burgert CR, Colston J, Roy T, Zachary B. Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys. Calverton, Maryland, USA; 2013.

24. Gething PW, Matthews Z, Nyarko P, Falkingham J, Atkinson PM, Johnson FA, et al. Geographical access to care at birth in Ghana: a barrier to safe motherhood. BMC Public Health [Internet]. 2012;12(1):1. Available from: BMC Public Health

25. Gabrysch S, Campbell OM. Still too far to walk: Literature review of the determinants of delivery service use. BMC Pregnancy Childbirth [Internet]. 2009 Dec 11;9(1):34. Available from: http://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/1471-2393-9-34

26. METCALF CJE, TATEM A, BJORNSTAD ON, LESSLER J, O'REILLY K, TAKAHASHI S, et al. Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. Epidemiol Infect [Internet]. 2015 May 14;143(7):1457–66. Available from: https://www.cambridge.org/core/product/identifier/S0950268814001988/type/journal_article

27. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. J R Stat Soc Ser C (Applied Stat [Internet]. 1998;47:299–350. Available from: http://doi.wiley.com/10.1111/1467-9876.00113

28. Utazi CE, Wagai J, Pannell O, Cutts FT, Rhoda DA, Ferrari MJ, et al. Geospatial variation in measles vaccine coverage through routine and campaign strategies in Nigeria: Analysis of recent household surveys. Vaccine. 2020 Mar;38(14):3062–71.

29. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. Vaccine [Internet]. 2018;36(12):1583–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29454519

30. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B (Statistical Methodol [Internet]. 2009 Apr;71(2):319–92. Available from: http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x

31. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LP; 2019.

32. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.r-project.org/

33. Huicho L, Huayanay-Espinoza CA, Hernandez P, Niño de Guzman J, Rivera-Ch M. Enabling reproductive, maternal, neonatal and child health interventions: Time trends and driving factors of health expenditure in the successful story of Peru. Horton S, editor. PLoS One [Internet]. 2018 Oct 31;13(10):e0206455. Available from: https://dx.plos.org/10.1371/journal.pone.0206455

34. Huicho L, Segura ER, Huayanay-Espinoza CA, de Guzman JN, Restrepo-Méndez MC, Tam Y, et al. Child health and nutrition in Peru within an antipoverty political agenda: A Countdown to 2015 country case study. Lancet Glob Heal. 2016;4(6):e414–26.

Supplementary table 1 - Complete definitions for the composite coverage index (CCI) and its indicators

Indicator	Acronym	Numerator	Denominator
Demand for family planning satisfied by modern methods	FPSm	Who is using (or whose partner is using) a modern contraceptive method	Women aged 15-49 years either married or in union in need of contraception
Antenatal care 4 or more visits	ANC4	Attended at least four antenatal care (ANC) visits with any provider	Women aged 15-49 years who had a birth in the last 2/3 years
Skilled attendant at delivery	SBA	Delivered by a skilled attendant (based on each country's definition of skilled attendant)	Women aged 15-49 years who had a birth in the last 2/3 years
BCG vaccination	BCG	Received Bacillus Calmette-Guérin (BCG) vaccine	All live-children, 12-23/18-29/15-26 months (according to country's calendar)
DPT3 vaccination	DPT3	Received 3 doses of Diphteria, Pertussis, Tetanus (DPT) vaccine	All live-children, 12-23/18-29/15-26 months (according to country's calendar)
Measles vaccination	MSL	Received measles vaccine	All live-children, 12-23/18-29/15-26 months (according to country's calendar)
Treatment for diarrhea	ORS	Received oral rehydration salts (ORS)	All live children aged 0-59 months with diarrhea in the last 2 weeks
Care-seeking for pneumonia	CAREP	Sought treatment from an appropriate health facility or provider.	Live children, 0-59 months, suspected pneumonia in the last 2 weeks
Composite coverage index	CCI	$CCI = \frac{1}{4}\left(FPSm + \frac{SBA + ANC4}{2} + \frac{2(DPT3) + BCG + MSL}{4} + \frac{ORS + CAREP}{2} \right)$	

Supplementary table 2 – Description of covariates included in the analysis

Covariate	Year	Resolution	Unit	Source
Altitude	NA	1km	meters	SRTM (raster package)
Travel time to cities >50,000	2000	1km	minutes	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/
Distance to health facilities	2017	1km	meters	https://www.datosabiertos.gob.pe/dataset/minsa-ipress
Enhanced vegetation index	2017	1km	0 to 10000 (least to most vegetation)	https://ladsweb.modaps.eosdis.nasa.gov/search/history
Urbanicity	2014	1km	0.00 to 1.00 (extremely rural to urban)	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/GHSL/GHS_BUILT_LDSMT_GLOBE_R2018A/
Nighttime lights	2016	100m	nW/cm2/sr	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/
Improved water coverage	2018	1km	proportion	Interpolated using kriging
Improved sanitation coverage	2018	1km	proportion	Interpolated using kriging
Mean number of household members	2018	1km	number	Interpolated using kriging
% of households in Q1 or Q2	2018	1km	proportion	Interpolated using kriging
Mean women's years of education	2018	1km	proportion	Interpolated using kriging
% of indigenous population	2018	1km	proportion	Interpolated using kriging
Distance to protected areas	2017	100m	km	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/
Distance to build settlements (BGSM)	2017	100m	km	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/

Supplementary Table 3 - Summary statistics for covariates included in the analysis

Covariate	Mean	Median	Min	P1%	P99%	Max
Altitude	1373.00	487.00	0.00	6.60	4350.00	4876.00
Distance to health facilities	0.02	0.01	0.00	0.00	0.09	0.31
Travel time to health facilities	14.67	2.21	0.00	0.00	219.00	888.00
Travel time to cities 50k	172.50	44.49	0.51	0.51	1438.00	3269.00
Enhanced vegetation index	2025.00	1771.00	-3000.00	-1666.00	5581.00	6378.00
Urbanicity	15.65	0.36	0.00	0.00	92.99	96.95
Nighttime lights	13.67	1.85	-0.07	-0.03	75.00	85.00
Improved water	0.96	0.98	0.01	0.37	1.00	1.00
Improved sanitation	0.63	0.75	0.00	0.00	0.98	0.99
% of households in Q1 or Q2	0.49	0.45	0.00	0.00	0.99	0.99
Mean number of household members	3.84	3.82	2.35	2.72	4.84	5.45
Distance to built settlements	1.14	0.33	-2.17	-1.78	13.90	64.00
Distance to protected areas	705.00	702.00	56.00	195.00	1240.00	1278.00
Years of education (women)	10.00	10.40	4.10	5.30	13.60	14.00
% of indigenous population	0.07	0.00	0.00	0.00	0.99	0.99

Note: Zeros and negative values were replaced by a positive value close to 1 for some covariates

Supplementary Table 4 - Covariates selected for each modeled indicator

Covariate	BCG	DPT	MSL	CAREP	ORS	SBA3	ANC4	FPSmo
Altitude		X			X	X		
Distance to health facilities			X		X		X	X
Travel time to health facilities					X			
Travel time to cities 50k		X	X			X		
Enhanced vegetation index	X		X		X	X		
Urbanicity						X		
Nighttime lights		X			X	X		X
Improved water								
Improved sanitation	X	X	X		X	X	X	X
% of households in Q1 or Q2		X		X			X	X
Mean number of household members	X		X			X	X	
Distance to built settlements			X			X		
Distance to protected areas			X		X	X	X	
Years of education (women)	X	X			X	X	X	X
% of indigenous population		X					X	X

BCG Bacillus Calmette-Guérin vaccine; *DPT3* 3 doses of Diphteria, pertussis, tetanus vaccine; *CAREP* Care-seeking for pneumonia; *SBA* Skilled attendant at delivery; *ANC4* Antenatal care 4 or more visits; *ORS* Oral rehydration salts; *FPSmo* Demand for family planning satisfied by modern methods

Supplementary Table 5 – Model validation metrics

	Cross-validation			In-sample		
Indicator	**Correlation**	**Bias**	**MAE**	**Correlation**	**Bias**	**MAE**
BCG	0.31	0.00	0.07	0.61	0.00	0.06
DPT	0.20	0.00	0.18	0.55	0.00	0.16
Measles	0.20	0.00	0.22	0.59	0.00	0.20
CAREP	0.03	0.00	0.30	0.80	0.00	0.25
SBA	0.73	0.00	0.04	0.88	0.00	0.03
ANC4	0.25	0.00	0.06	0.45	0.00	0.06
ORS	0.26	0.00	0.27	0.60	0.00	0.24
FPSmo	0.34	-0.01	0.24	0.54	-0.01	0.22

MAE Mean absolute error; *BCG* Bacillus Calmette-Guérin vaccine; *DPT3* 3 doses of Diphteria, pertussis, tetanus vaccine; *CAREP* Care-seeking for pneumonia; *SBA* Skilled attendant at delivery; *ANC4* Antenatal care 4 or more visits; *ORS* Oral rehydration salts; *FPSmo* Demand for family planning satisfied by modern methods

Supplementary Table 6 – Comparison of predicted vs observed estimates at department-level (adm1) for the composite coverage index (CCI) and its 8 indicators

Department	BCG			DPT3			MEASLES			CARE-SEEKING			ORS			ANC4			SBA			FPSmo			CCI		
	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff	Pred	Obs	Diff
Amazonas	94.6	93.6	1.0	84.5	83.4	1.1	85.9	81.1	4.7	66.9	65.9	1.0	27.7	24.2	3.5	94.7	92.4	2.3	85.7	81.3	4.4	66.0	61.2	4.7	72.7	69.6	3.1
Ancash	97.6	98.0	-0.4	87.6	90.0	-2.4	83.3	89.0	-5.7	69.6	74.5	-4.8	26.6	19.1	7.5	97.7	98.2	-0.5	97.9	98.0	-0.2	64.7	63.0	1.8	74.9	74.9	0.0
Apurímac	97.2	99.2	-2.0	89.0	90.3	-1.3	82.9	82.6	0.3	68.4	75.6	-7.2	23.7	22.9	0.9	96.8	97.7	-0.9	99.1	99.9	-0.7	59.8	61.5	-1.7	73.3	75.0	-1.7
Arequipa	97.8	98.6	-0.8	91.9	93.9	-2.1	80.5	78.4	2.1	69.7	80.5	-10.8	30.3	30.8	-0.5	96.7	96.3	0.4	98.8	99.0	-0.1	71.0	67.3	3.7	77.3	77.9	-0.6
Ayacucho	95.7	93.8	1.8	88.2	87.6	0.6	81.0	84.9	-3.9	68.6	59.4	9.2	22.9	22.0	1.0	95.6	95.2	0.4	97.5	98.3	-0.8	53.2	56.8	-3.6	71.0	70.7	0.3
Cajamarca	94.8	95.3	-0.5	87.2	90.2	-3.0	85.4	91.3	-5.9	67.3	62.4	4.9	21.8	21.4	0.5	95.5	95.4	0.0	87.1	87.4	-0.3	56.2	57.4	-1.2	70.2	70.6	-0.4
Callao	96.4	96.1	0.3	81.4	82.5	-1.1	81.5	78.0	3.5	73.7	74.6	-0.9	40.8	40.5	0.3	97.2	96.6	0.5	99.6	99.8	-0.2	73.4	72.4	1.0	78.5	78.2	0.3
Cusco	96.5	96.1	0.4	88.4	85.9	2.5	83.2	74.7	8.5	66.5	74.8	-8.3	28.1	33.3	-5.2	97.1	97.9	-0.8	97.5	98.2	-0.7	55.1	58.2	-3.1	72.2	74.0	-1.8
Huancavelica	95.6	97.8	-2.2	86.9	88.7	-1.8	84.7	89.0	-4.4	66.1	71.6	-5.5	18.3	19.5	-1.2	95.1	95.9	-0.7	92.9	93.1	-0.3	42.9	46.6	-3.7	66.9	69.4	-2.5
Huánuco	95.6	96.4	-0.8	89.6	92.2	-2.6	84.2	86.5	-2.4	65.9	64.9	1.0	22.7	27.0	-4.3	96.4	96.4	0.0	94.2	96.1	-1.9	60.8	66.9	-6.1	72.5	75.2	-2.7
Ica	97.1	97.3	-0.1	83.8	80.0	3.9	87.0	88.2	-1.2	73.0	85.8	-12.8	39.1	38.2	1.0	97.0	97.3	-0.3	99.2	99.4	-0.2	67.2	65.1	2.1	77.3	78.0	-0.6
Junín	96.5	95.8	0.7	89.6	91.8	-2.2	80.6	81.0	-0.4	63.2	60.5	2.7	29.0	23.6	5.4	95.6	96.3	-0.7	89.7	91.7	-1.9	62.1	64.2	-2.1	72.5	72.6	-0.1
La Libertad	97.1	96.7	0.4	82.6	80.6	2.0	82.9	80.7	2.2	68.0	64.3	3.7	32.1	34.9	-2.8	96.8	96.3	0.4	92.1	89.8	2.3	66.5	62.6	3.9	74.3	72.5	1.8
Lambayeque	95.9	95.4	0.4	86.1	84.5	1.7	74.5	71.5	3.0	68.9	72.6	-3.6	36.6	37.4	-0.8	95.1	94.0	1.0	97.1	95.2	1.8	68.6	60.9	7.7	75.8	73.6	2.2
Lima Province	95.2	94.9	0.3	83.6	82.8	0.8	80.3	79.0	1.3	73.1	75.5	-2.4	38.4	37.4	1.0	97.3	97.6	-0.2	99.6	99.5	0.0	72.6	70.2	2.4	78.1	77.5	0.6
Loreto	83.8	85.6	-1.8	79.3	81.7	-2.5	70.4	75.8	-5.4	66.2	73.1	-6.9	32.5	40.0	-7.5	90.2	90.1	0.1	66.1	72.3	-6.2	59.2	61.3	-2.1	66.2	70.1	-3.8
Madre de Dios	95.5	99.5	-4.0	81.0	82.2	-1.2	71.8	73.8	-1.9	61.7	52.2	9.5	31.2	45.7	-14.6	93.4	94.7	-1.2	88.6	97.9	-9.3	54.5	65.4	-10.9	68.6	73.8	-5.2
Moquegua	98.0	98.3	-0.3	90.1	88.8	1.4	85.9	88.8	-2.9	70.1	77.0	-6.9	29.4	36.3	-6.8	97.4	97.9	-0.5	97.7	99.5	-1.8	69.2	72.2	-3.0	76.9	79.7	-2.8
Pasco	96.3	95.3	1.0	85.6	83.9	1.7	82.3	81.0	1.3	65.6	71.9	-6.3	22.5	31.2	-8.7	96.2	96.7	-0.5	93.6	96.8	-3.3	65.4	69.9	-4.5	72.9	76.1	-3.1
Piura	97.2	98.0	-0.8	85.7	87.5	-1.8	81.1	81.8	-0.7	67.4	69.3	-1.9	36.8	37.5	-0.7	96.6	96.8	-0.2	91.9	91.9	0.0	71.9	70.8	1.1	76.4	76.8	-0.4
Puno	94.5	91.4	3.1	76.3	75.0	1.4	72.7	70.9	1.9	64.4	40.8	23.6	21.3	18.6	2.7	92.5	91.9	0.6	92.5	95.0	-2.5	41.2	43.4	-2.1	64.1	61.1	3.0
San Martín	94.6	93.2	1.4	90.2	90.9	-0.7	82.7	78.5	4.1	64.0	50.0	14.0	39.4	42.1	-2.7	97.5	98.1	-0.6	93.5	92.0	1.5	66.2	66.5	-0.3	75.7	74.0	1.7
Tacna	98.5	99.3	-0.8	90.2	90.2	0.0	81.7	78.9	2.9	70.9	64.8	6.0	32.1	21.2	10.9	97.5	97.7	-0.2	98.2	98.6	-0.4	63.6	60.0	3.6	75.8	72.7	3.1
Tumbes	98.0	98.4	-0.4	92.9	93.9	-1.1	87.2	89.7	-2.4	63.9	67.9	-4.1	43.8	51.1	-7.3	96.5	96.7	-0.2	98.4	99.1	-0.6	85.2	83.2	2.0	82.3	83.6	-1.3
Ucayali	92.7	94.5	-1.8	84.7	87.1	-2.4	70.2	69.1	1.0	61.4	60.8	0.6	41.5	42.0	-0.5	92.0	94.0	-2.1	82.5	88.6	-6.2	66.3	65.9	0.4	72.0	73.3	-1.3

BCG Bacillus Calmette-Guérin vaccine; *DPT3* 3 doses of Diphteria, pertussis, tetanus vaccine; *CAREP* Care-seeking for pneumonia; *SBA* Skilled attendant at delivery; *ANC4* Antenatal care 4 or more visits; *ORS* Oral rehydration salts; *FPSmo* Demand for family planning satisfied by modern methods; *CCI* Composite coverage index

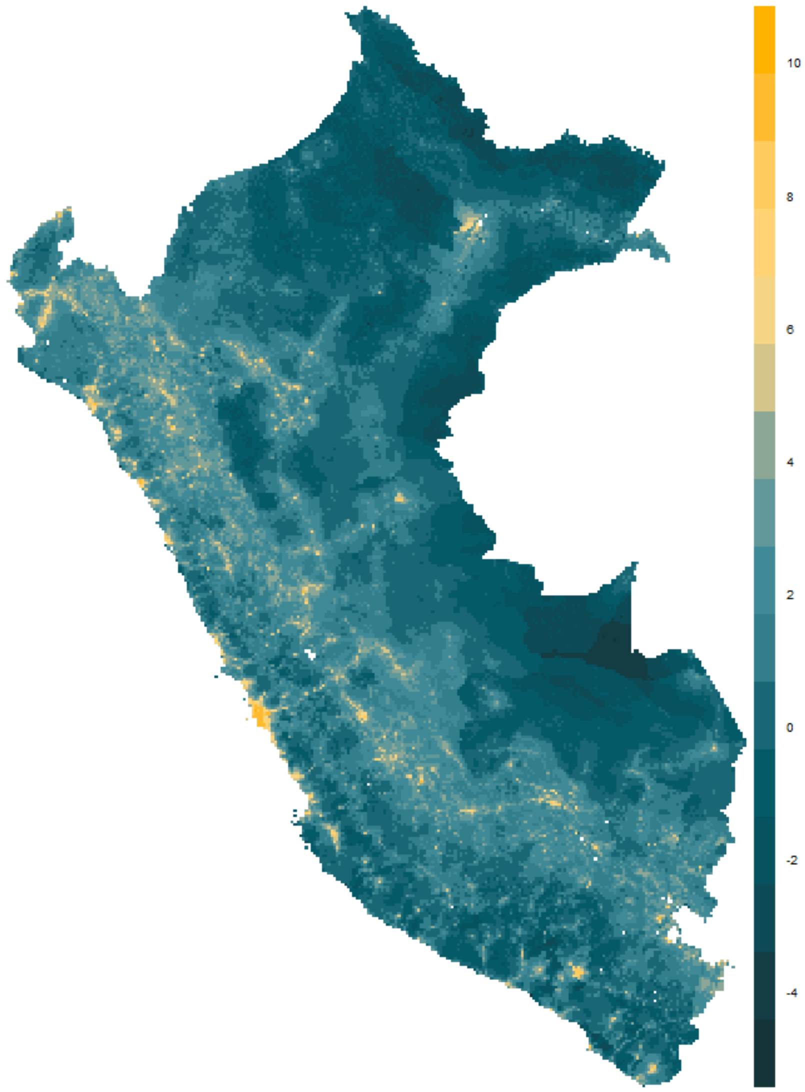
Supplementary table 7 – Predicted estimates for the composite coverage index (CCI) for the provinces of Peru

Provinces	Point estimate	Standard error	2.5th percentile	97.5th percentile
Abancay	75.9	4.3	66.7	83.7
Acobamba	66.9	2.2	62.4	71.0
Acomayo	72.3	2.8	66.5	77.6
Aija	70.2	2.5	65.0	74.8
Alto Amazonas	61.9	1.7	58.6	65.1
Ambo	72.4	1.7	69.0	75.6
Andahuaylas	71.8	1.4	69.0	74.4
Angaraes	67.1	2.0	62.9	70.7
Anta	73.2	1.9	69.4	76.6
Antabamba	71.7	2.0	67.8	75.6
Antonio Raymondi	69.9	2.2	65.5	74.2
Arequipa	77.5	2.1	73.3	81.3
Ascope	76.8	1.9	73.0	80.3
Asunción	71.9	2.4	66.9	76.7
Atalaya	60.2	2.5	55.0	64.9
Ayabaca	69.8	2.0	65.9	73.6
Aymaraes	73.1	1.7	69.8	76.2
Azángaro	61.6	2.0	57.5	65.4
Bagua	73.6	2.0	69.5	77.2
Barranca	79.4	2.5	74.2	84.0
Bellavista	72.8	1.9	69.0	76.3
Bolognesi	71.5	1.8	68.1	74.9
Bolívar	63.3	2.4	58.5	68.0
Bongará	73.5	2.7	67.8	78.5
Cajabamba	66.9	2.0	63.1	70.8
Cajamarca	69.8	2.0	65.8	73.4
Cajatambo	71.4	2.2	67.1	75.4
Calca	71.6	2.0	67.5	75.3
Callao	78.5	2.8	72.1	83.8
Camaná	79.5	2.3	74.9	83.9
Canas	69.6	2.2	65.1	73.7
Canchis	70.8	4.3	61.9	78.9
Candarave	70.8	2.4	65.8	75.6
Cangallo	67.5	1.8	63.9	70.8
Canta	74.3	3.2	67.3	79.8
Carabaya	66.5	2.2	62.3	70.9
Caravelí	76.4	2.3	71.6	80.6
Carhuaz	72.4	2.3	67.7	76.6
Carlos Fermin Fitzcarrald	70.8	2.2	66.5	75.1
Casma	76.0	2.2	71.6	80.1
Castilla	76.1	2.4	71.1	80.3
Castrovirreyna	63.6	2.5	58.7	68.4
Caylloma	70.6	2.0	66.5	74.3
Cañete	76.5	2.4	71.6	80.9
Celendín	67.5	2.0	63.5	71.2
Chachapoyas	77.3	4.6	68.2	85.3

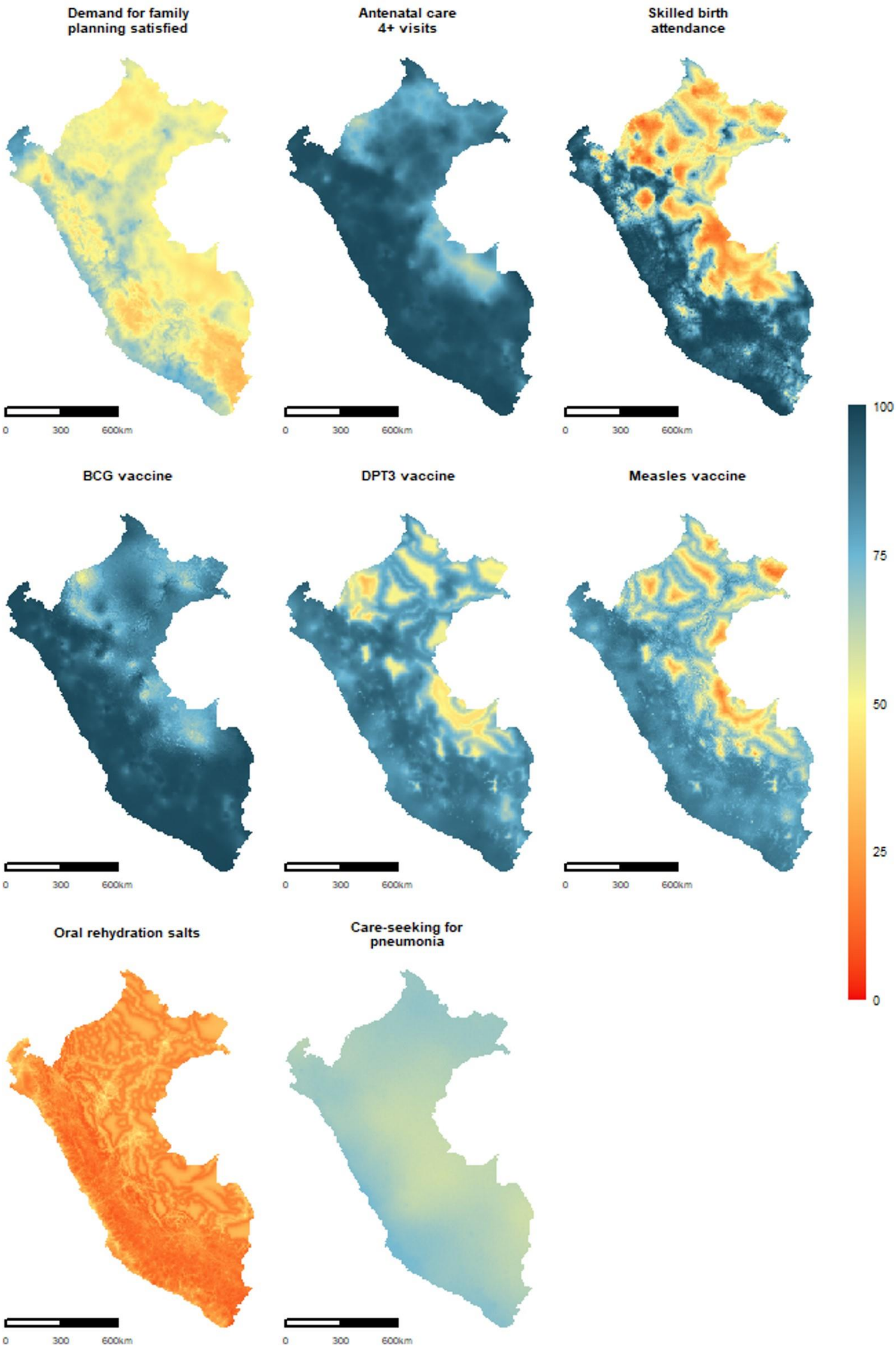
Chanchamayo	74.3	4.7	64.0	82.3
Chepén	77.6	4.0	69.4	84.4
Chiclayo	77.0	2.7	71.4	81.8
Chincha	77.3	4.0	68.9	84.2
Chincheros	71.2	2.7	65.7	76.2
Chota	69.1	1.5	65.9	72.0
Chucuíto	59.9	2.5	55.0	64.8
Chumbivilcas	69.9	1.8	66.4	73.3
Chupaca	71.5	2.9	65.8	76.9
Churcampa	66.1	1.7	62.6	69.3
Concepción	71.0	3.0	65.0	76.5
Condesuyos	74.3	2.3	69.7	78.6
Condorcanqui	54.8	2.4	50.3	59.3
Contralmirante Villar	80.7	1.7	77.1	83.8
Contumazá	71.1	2.1	66.8	74.9
Coronel Portillo	74.2	2.9	68.1	79.7
Corongo	67.0	2.6	61.8	71.6
Cotabambas	72.7	1.7	69.5	75.9
Cusco	73.9	4.4	65.1	82.0
Cutervo	71.2	2.3	66.4	75.8
Daniel Alcides Carrión	71.9	1.6	68.7	74.9
Dos de Mayo	69.4	1.9	65.8	73.2
El Collao	62.9	2.8	57.0	68.3
El Dorado	73.6	2.3	68.9	77.7
Espinar	67.2	2.7	61.5	72.3
Ferreñafe	71.2	2.1	67.1	75.3
General Sánchez Cerro	72.3	2.2	67.9	76.4
Gran Chimú	70.3	3.6	62.8	76.6
Grau	72.7	2.3	68.1	76.9
Huacaybamba	67.5	2.2	63.2	71.8
Hualgayoc	69.8	4.1	61.4	77.1
Huallaga	74.4	3.4	67.2	80.8
Huamalíes	71.7	1.5	68.8	74.3
Huamanga	72.0	4.1	64.1	79.6
Huanca Sancos	68.2	2.2	64.0	72.6
Huancabamba	64.4	2.0	60.2	68.2
Huancane	61.8	2.2	57.3	66.1
Huancavelica	67.1	2.0	63.0	70.9
Huancayo	73.9	3.0	68.0	79.4
Huanta	68.4	1.4	65.6	71.0
Huaral	78.1	2.0	73.8	81.9
Huaraz	74.8	4.3	65.7	82.2
Huari	71.0	1.6	67.8	74.2
Huarmey	75.7	2.0	71.5	79.3
Huarochiri	74.1	1.8	70.4	77.5
Huaura	77.5	2.0	73.5	81.2
Huaylas	70.1	3.0	64.1	75.8
Huaytara	66.3	2.1	62.3	70.2
Huenuco	75.0	2.2	70.2	79.0
Ica	77.3	3.5	70.0	83.7
Ilo	80.4	3.4	73.3	86.8
Islay	78.6	2.2	74.3	82.4

Jauja	72.3	1.9	68.6	76.0
Jaén	72.5	1.6	69.1	75.5
Jorge Basadre	77.0	2.3	72.2	81.2
Julcan	62.8	2.8	56.9	68.2
Junín	72.9	1.8	69.1	76.3
La Convención	71.7	1.7	68.0	74.9
La Mar	70.0	1.4	67.0	72.7
La Unión	72.3	2.6	67.3	77.1
Lago Titicaca	63.3	2.8	57.4	68.6
Lamas	76.1	2.3	71.0	80.4
Lambayeque	73.8	1.5	70.5	76.8
Lampa	63.3	2.1	59.1	67.3
Lauricocha	69.4	1.9	65.3	73.2
Leoncio Prado	75.8	1.7	72.4	78.9
Lima	78.1	1.3	75.6	80.6
Loreto	61.7	2.1	57.8	66.1
Lucanas	72.3	1.7	69.1	75.7
Luya	72.8	1.5	69.8	75.6
Manu	68.0	2.4	63.6	72.6
Marañón	69.2	1.7	65.7	72.4
Mariscal Cáceres	75.4	4.3	65.9	82.9
Mariscal Luzuriaga	71.6	2.2	67.4	75.8
Mariscal Nieto	76.9	2.1	72.5	80.8
Mariscal Ramón Castilla	62.9	2.6	57.7	68.1
Maynas	69.1	2.0	65.0	72.7
Melgar	65.4	2.0	61.4	69.3
Moho	63.9	2.8	58.2	69.0
Morropón	73.4	1.9	69.7	77.0
Moyobamba	76.3	3.2	69.8	81.9
Nazca	78.7	3.0	72.2	83.9
Ocros	73.7	2.2	69.5	77.9
Otuzco	67.0	1.9	63.2	70.7
Oxapampa	71.0	1.6	67.8	74.0
Oyon	72.6	3.0	66.5	77.8
Pacasmayo	76.8	2.3	71.8	81.0
Pachitea	70.5	1.9	66.7	74.1
Padre Abad	72.8	1.8	69.4	76.4
Paita	77.8	2.8	71.9	82.8
Pallasca	65.2	2.5	60.0	69.8
Palpa	76.4	4.6	66.2	84.6
Parinacochas	71.7	2.9	65.7	77.1
Paruro	72.4	1.8	68.8	75.9
Pasco	74.7	1.5	71.6	77.4
Pataz	68.3	2.5	63.2	73.1
Paucar del Sara Sara	73.1	2.4	68.3	77.7
Paucartambo	69.0	1.8	65.1	72.7
Picota	73.6	2.5	68.7	78.3
Pisco	76.4	2.1	72.3	80.4
Piura	78.3	2.2	73.7	82.3
Pomabamba	70.6	2.3	66.0	75.0
Puerto Inca	70.2	1.9	66.5	73.8
Puno	64.2	3.9	56.2	71.5

Purús	55.2	3.5	48.3	62.0
Quispicanchi	70.5	1.9	66.7	74.0
Recuay	72.8	2.0	68.4	76.4
Requena	65.1	2.2	60.7	69.3
Rioja	73.7	1.9	69.8	77.1
Rodríguez de Mendoza	75.9	1.9	72.0	79.6
San Antonio de Putina	62.9	2.5	58.4	67.7
San Ignacio	71.2	1.7	67.8	74.6
San Marcos	69.0	3.2	62.4	74.7
San Martín	78.7	4.4	69.5	86.2
San Miguel	71.2	1.9	67.2	74.9
San Pablo	71.0	3.6	63.6	77.5
San Román	65.6	6.1	53.6	76.8
Sandia	67.4	2.3	62.7	71.8
Santa	78.2	3.0	71.9	83.7
Santa Cruz	71.7	2.1	67.6	75.6
Santiago de Chuco	65.3	2.3	60.8	69.8
Satipo	68.2	2.4	63.3	72.7
Sechura	75.4	2.4	70.6	79.9
Sihuas	69.8	2.0	65.6	73.6
Sucre	70.9	1.9	66.9	74.8
Sullana	79.5	1.8	75.8	82.8
Sánchez Carrión	68.6	2.1	64.3	72.5
Tacna	76.2	2.8	70.2	81.2
Tahuamanu	69.0	2.4	64.0	73.6
Talara	78.8	4.1	69.7	86.5
Tambopata	68.8	1.8	64.9	72.2
Tarata	70.2	3.5	62.5	76.8
Tarma	74.8	1.9	71.1	78.2
Tayacaja	67.6	1.2	65.2	70.1
Tocache	73.7	1.8	69.8	77.0
Trujillo	77.2	3.7	69.4	83.6
Tumbes	82.5	2.6	77.0	87.3
Ucayali	64.1	2.5	59.6	69.4
Urubamba	73.3	2.2	69.0	77.4
Utcubamba	74.7	1.5	71.6	77.7
Victor Fajardo	69.2	1.8	65.5	72.7
Vilcas Huamán	68.2	1.9	64.6	71.8
Viru	71.8	2.9	66.2	77.4
Yarowilca	69.2	2.4	64.4	73.9
Yauli	73.9	3.7	66.1	80.7
Yauyos	67.0	2.0	62.9	70.6
Yungay	70.9	2.1	66.6	75.0
Yunguyo	67.9	4.2	59.2	75.5
Zarumilla	82.2	2.9	75.7	87.2



Supplementary Figure 1 – Map of log of population density in Peru



Supplementary Figure 2 – Predicted coverage for all eight composite coverage index (CCI) indicators in Peru

Article 3

Measuring time trends in geographic health inequalities at different resolutions: the scale effect

Leonardo Z. Ferreira[1 2], Cauane Blumenberg[2], Fernando P. Hartwig[2], C. Edson Utazi[3], Kristine Nilsen[3], Andrew J. Tatem[3], Aluisio JD Barros[1 2]

1. International Center for Equity in Health, Universidade Federal de Pelotas, Pelotas, Brazil

2. Post-Graduate Program in Epidemiology, Universidade Federal de Pelotas, Pelotas, Brazil

3. WorldPop, Geography and Environmental Science, University of Southampton, Southampton, UK

Abstract

Background: Monitoring health inequalities is an essential and continuous process to ensure equitable progress in the countries. However, studying the geographic inequality dimension presents many challenges including the modifiable areal unit problem (MAUP). This study aims to quantify the magnitude of the scale effect, one of MAUP's issues, when assessing geographic inequalities over time in household surveys using complex measures of inequality.

Methods: Using data from two national health surveys carried out in Peru in 2009 and 2019, we applied a model-based geostatistical approach to generate estimates for stunting in children under-5 at different resolutions. Then, we calculated four complex measures of inequality (weighted and unweighted mean absolute difference to the mean and index of disparity) and used them to compare inequalities between resolutions and over time.

Results: The magnitude of the inequality measures increases as the number of geographical units increase, but it reached a plateau around 2000 geographical units for all measures. The absolute difference over the years remained stable for matching resolutions. Population-

weighted measures of inequality varied substantially less between resolutions than their unweighted counterparts.

Conclusions: Overall, weighted inequality measures presented more stability, especially for absolute measures, and the magnitude of the changes over time was not affected by the MAUP when comparing years at the same resolution.

Introduction

Monitoring and reducing inequalities is as crucial as ever as the world marches in the pursuit of the Sustainable Development Goals set for 2030 (1). Dimensions such as wealth, education and place of residence are long studied and have been monitored at global, regional, and local levels in most low- and middle-income countries (2). However, within-country subnational geographical units (districts, provinces, regions, etc.) are often overlooked especially in multi-country comparisons and in the assessment of progress over time. Although central to the process of policy design and decision-making, monitoring inequalities for subnational units require accounting for the modifiable areal unit problem (MAUP) and facing the challenges of measuring inequalities among unordered groups (3,4), which could partly explain the reason other inequality dimensions are preferred. Estimates for subnational geographical units are obtained by aggregating data points (e.g., sample clusters in household surveys) based on a modifiable number of areas and shapes for the boundaries. Thus, the MAUP implies that changing these parameters may affect the interpretation of geographical analyses due to differences in scale (number of areas) or zoning (boundaries) alone, despite being originated from the same data points. This potential issue has been acknowledged by many studies in the field of geography and health inequalities (4–6), but a clearer picture of the amount of uncertainty inherent from the variation in the size and aggregation of these areas is desired.

Due to fragile and inefficient health information systems, most low- and middle-income countries rely on national household surveys as their primary source of information on reproductive, maternal, newborn and child health (RMNCH). These surveys are often carried out every 3 to 5 years collecting information on multiple health indicators whose estimates are representative nationally and for the first administrative division of the country - generally large areas such as macroregions (states, regions) or the aggregation of smaller subdivisions. As of late, studies have used geospatial models to obtain reliable estimates for smaller geographical units from the surveys based on spatial correlation and geospatial covariates (7). This strategy allows the use of subnational divisions whose estimates are more suitable for policy planners and expand the possibilities for monitoring geographical inequalities over time and between countries. Yet, empirical evidence on the impact of the MAUP is necessary to guide further studies that aim to use geospatial modeling to quantify inequalities from a geographical perspective.

This study aims to test the effects of the MAUP in assessing geographic inequalities over time by using geospatial models and complex measures of inequality at different resolutions in two Peruvian surveys as a case study. We expect to show the variation in absolute and relative measures at different resolutions (the scale effect) and compare whether differences over time vary as the scale changes.

Methods

Data

We used data from the Encuesta Demográfica y de Salud Familiar (ENDES), a population-based survey carried out in 2009 and 2019 in Peru (8,9). These surveys are designed to provide representative estimates at national and departmental level for several RMNCH indicators. They use a multi-stage sampling design accounting for age, sex and other characteristics of the population where the clusters are selected in the first stage, followed by the selection of households in the second stage. Peru is administratively divided into 25 departments, subdivided into 195 provinces which are further disaggregated into 1761 districts.

The outcome selected for the analysis was the prevalence of stunting defined as the percentage of under-five children with height-for-age below -2 standard deviations in comparison to the World Health Organization Child Growth Standard (10), based on anthropometric measures collected by trained interviewers.

Geospatial modelling

A total of 11 geospatial covariates were prepared and tested to be used as predictors in the geospatial model. The covariates relate to sociodemographic characteristics, urbanization, accessibility, vegetation, and topography which were previously used as potential predictors in other studies (11,12). The final set of covariates for each model were selected following a backwards stepwise logistic regression at 5% significance level. The list of candidate covariates is presented in the Supplementary Table 1.

We followed a Bayesian framework to generate estimates at 5 x 5 km, 10 x 10 km and 25 x 25 km resolutions using a model-based geostatistical approach (13) based on spatial correlation and geospatial variables (14,15). Given $Y(s_i)$ is the number of households at cluster location s_i (i=1, ..., n), out of a total of $N(s_i)$ households sampled at the location, the model can be defined as follows:

$$Y(s_i) \sim Binomial\ (N(s_i), p(s_i))$$

$$logit\big(p(s_i)\big) = x(s_i)^T\,\beta + \ \omega(s_i) + \epsilon(s_i)$$

where $x(s_i)$ is a set of covariates for each cluster s_i, β are the corresponding regression parameters, $\omega(s_i)$ is a Gaussian spatial random effect used to capture residual spatial correlation in the model, and $\epsilon(s_i)$ is a Gaussian random effect used to model non-spatial residual variation. We used the integrated nested Laplace approximation with the stochastic partial differential equations to fit the geostatistical model (16). Estimates for the grid-level resolutions were obtained by drawing 1000 samples from the posterior distribution generated by the model and were further aggregated into the first, second and third administrative divisions using population estimates obtained from Worldpop (17) for weighting.

We carried out an out-of-sample cross validation strategy with 5 folds to ensure model stability. Model performance was assessed using mean error (bias), the mean absolute error and the correlation between predicted and observed cluster values.

Inequality measures

We used four complex measures of inequality for unordered groups to quantify the magnitude of the inequalities in the prevalence of stunting estimated at different resolutions. The mean absolute difference to the mean (MADM) is an absolute measure that uses the national average as the reference and takes the mean of all the absolute differences of each subnational unit to the reference. As a relative measure, we opted for the index of disparity (IDISP), which can be interpreted as the relative counterpart of the MADM. It takes the mean of the absolute differences to the reference value (national average) and further divides it by the reference value and multiplies the quotient by 100. We calculated for both measures unweighted and population-weighted versions in each resolution. The inequality measures were calculated according to the following equations

$$MADM = \frac{\sum_j \left| r_j - r_{ref} \right|}{J}$$

$$IDISP = \frac{\sum_j \left| r_j - r_{ref} \right| / J}{r_{ref}} * 100$$

$$WMADM = \sum\nolimits_j \left| r_j - r_{ref} \right| * \; w_j$$

$$WIDISP = \frac{\sum_j \left| r_j - r_{ref} \right| * w_j}{r_{ref}} * 100$$

where r_j is the estimate for each subnational unit, r_{ref} represents the national average, J is the total number of subnational units and w_j is the proportion of the total population in each subnational unit.

MAUP effects

As mentioned above, MAUP has two main components: scale and zoning effects. The statistical analyses focused on scale effects, which were assessed by comparing results for the inequality measures calculated at the following resolutions: departments (25 units), provinces (195 units), districts (1761 units), 25 x 25km areas (2,657 units), 10 x 10 km areas (15,869 units) and 5 x 5 km areas (62,388 units). We evaluated the potential impact of the scale effect in two scenarios: a) the behavior of each inequality measures at different resolutions in each time point and b) the gap between time points in each inequality measures at different resolutions.

By comparing estimates between districts and areas of 25 x 25 km, where the number of geographical units is similar, the zoning effect can also be observed. However, this effect was not formally assessed in this study, and it is only briefly mentioned in the discussion section.

Survey data analysis, the covariate selection process and the calculation of the inequality measures was done in Stata 16 (18) while R 4.0.2 (19) was used for the processing of geospatial covariates, model fitting and validation.

Results

Geospatial models for both years were fit using altitude, improved sanitation, and the mean number of household members as covariates. For 2009, they also included travel time to cities, improved water, percentage of households in the first two quintiles of wealth, distance to protected areas and percentage of indigenous population. For 2019, the additional covariates were enhanced vegetation index and the mean years of education for women (Supplementary

Table 2). Based on the cross-validated results, the correlation of predicted and observed estimates at cluster-level was 0.65 and 0.56 for 2009 and 2019, respectively. Bias was very close to zero for both years and the mean absolute error was 13 percentage points (p.p.) in 2009 and 10p.p. in 2019 (Supplementary Table 3). We also compared published and predicted estimates at the first administrative level (ADM1) and found mean differences of 0.5p.p and -0.9p.p in 2009 and 2019, respectively, while the largest differences were 7.1p.p. in 2009 and -4.2p.p. in 2019 (Supplementary Table 4).

The national prevalence of stunting decreased from 23.9% in 2009 to 12.2% in 2019. At the first administrative level (ADM1), the gap between the best and worst performing subnational units was 49.7p.p. in 2009, decreasing to 28.2p.p. in 2019 (Table 1). Absolute inequalities decreased from around 11p.p. to a little over 6p.p. from 2009 to 2019, but this pattern was not observed for relative measures in the same period (Table 1).

When comparing subnational units disaggregated at the second administrative level (ADM2) to ADM1, the gaps are larger in both years but the reduction over time in percentage points is very similar (21p.p. for ADM1 vs 24p.p. for ADM2) for the absolute measures. The same pattern is observed in higher resolutions where the gap between regions continues to increase but the absolute difference over the years is stable. However, in relative terms, the inequality direction seems to reverse as the number of subnational units increase when comparing 2009 to 2019 (Table 1).

Both absolute measures of inequality increased as the resolution got higher, although the unweighted MADM varied substantially more than its weighted version (Figure 1). In the relative measures, a much steeper increase can be observed with the increase in the size of the subnational units, and a more stable pattern is seen in the weighted version of the IDISP (Figure 2). For all measures, the increase tends to reach a plateau from the 25 x 25 km to higher resolutions.

Table 1 – Inequality measures calculated for stunting in children under-five at each of the predicted resolutions for Peru in 2009 and 2019

Resolution	Number of units	Worst-best		MADM		WMADM		IDISP		WIDISP	
		2009	2019	2009	2019	2009	2019	2009	2019	2009	2019
ADM1	25	49.7	28.2	11.4	6.5	10.6	6.3	54.7	52.0	51.0	50.7
ADM2	195	60.8	36.8	15.1	9.2	12.2	7.6	72.1	73.4	58.6	61.1
ADM3	1,761	66.1	47.5	15.8	9.6	12.5	8.0	75.7	77.3	60.0	64.1
25 x 25 km	2,657	70.9	48.8	18.6	13.4	12.7	7.9	93.0	105.9	63.6	62.6
10 x 10 km	15,869	74.2	55.3	18.6	13.8	12.5	8.0	92.2	113.3	62.1	66.0
5 x 5 km	62,388	76.8	56.7	18.7	13.8	12.5	8.1	91.5	113.8	61.5	66.5

MADM: Mean absolute difference to the mean; WMADM: Weighted mean absolute difference to the mean; IDISP: Index of disparity; WIDISP: Weighted index of disparity

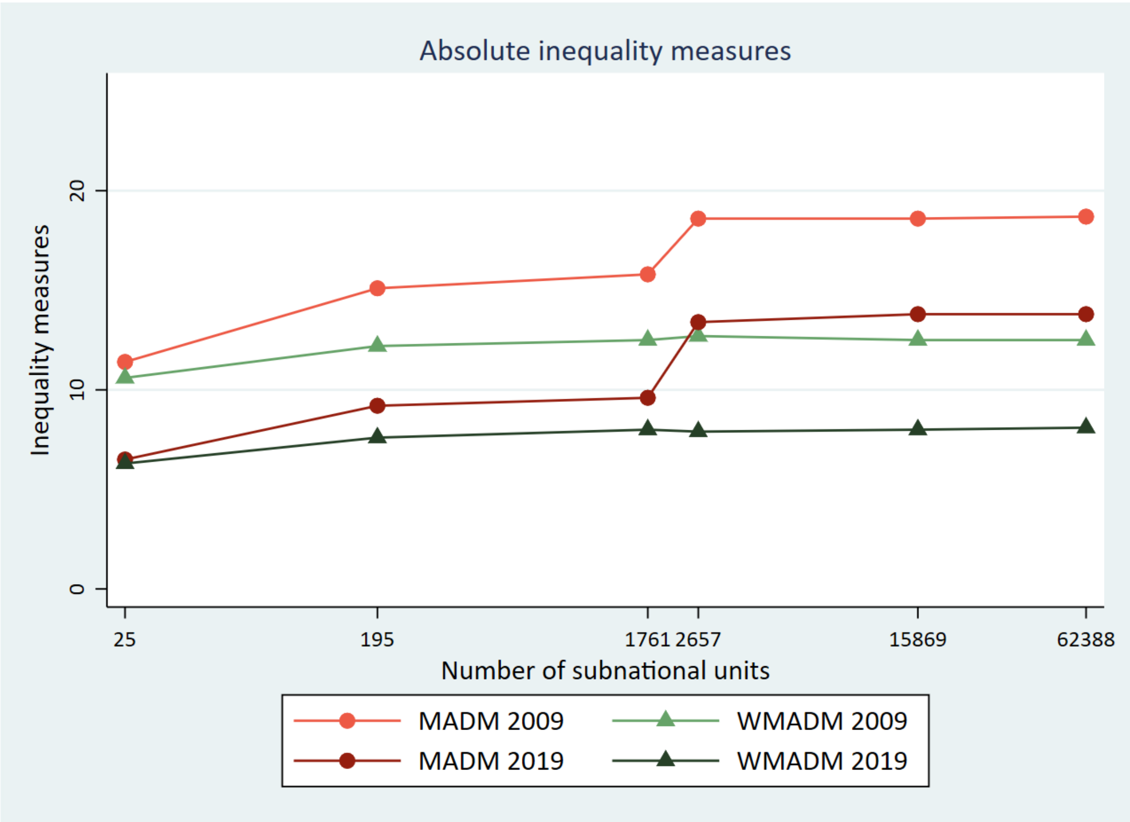


Figure 1 – Absolute inequalities measures for each of the estimated resolutions in 2009 and 2019.

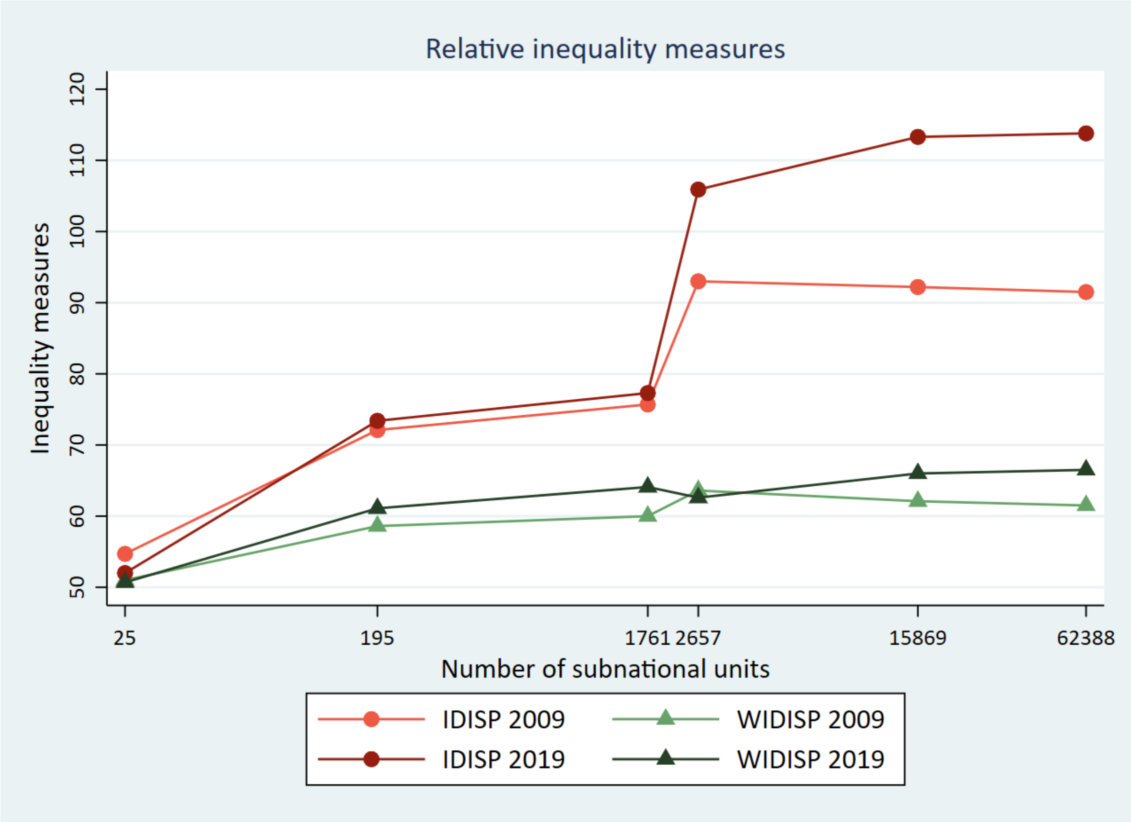


Figure 2 – Relative inequalities measures for each of the estimated resolutions in 2009 and 2019.

Discussion

Here, we assessed the behavior of four complex measures of inequality between six levels of aggregation and compared them in two points in time. The magnitude of the inequalities increased along the number of subnational units in all scenarios, but such increase stabilized in different resolutions for each inequality measure. This plateau is better observed when estimates are disaggregated after the grid-level, where very little variation is seen regardless of the grid size. From the third administrative level to the most aggregated grid-level estimates, a different effect of the MAUP is noticed. Also known as the zoning effect (20), estimates seem to change based on the shape of the geographical units rather than its number, since we move from administratively defined borders to equal-sized grid-cells. Weighted measures accounting for the population in each subnational unit have consistently shown to be less affected by the scale effect, suggesting they are more stable for monitoring inequalities in the dimension of subnational geographic units. Lastly, the magnitude of changes over time seems to be consistent regardless of the resolution for absolute inequality measures, but such interpretation should be taken cautiously for relative inequality measures, especially in higher resolutions.

The influence of the MAUP over geographically disaggregated estimates presented in this study seems to be consistent with the literature. The scale effect tends to be minimized when working with geographical units in higher resolutions (6), although this may not be available in many cases without indirect estimation such as geospatial modeling due to the sample design of the surveys. Also, comparing inequalities in one country over time using matching resolutions is unlikely to be impacted by the MAUP effects (5) for absolute inequality measures. However, the same may not be true when assessing inequalities through relative measures.

The inequality measures evaluated in our study are a part of a large set of complex measures, all of which have strengths and limitations. Both the MADM and the IDISP use the national average as a benchmark in their calculations and changing the benchmark may lead to different interpretations of the results. We have also compared the use of weighted and unweighted measures of inequality, and they both tell different stories. While the weighted measures tend to be less affected by the scale effect, they are driven by estimates of areas with larger populations, capturing the density of the spatial outcome more than its spatial heterogeneity, in which the latter could be better assessed using unweighted measures. Opting to go with unweighted or weighted measures mostly depends on the nature of the research question and evaluating them as complementary results should always be considered. The same applies for absolute versus relative ones. Nonetheless, these measures were chosen to illustrate how inequalities may shift in the proposed scenarios and conclusions may not hold for different inequality measures.

All in all, some limitations and caveats should be considered when interpreting our findings. First, the estimates produced at the studied resolutions are derived from geospatial models, which also carry their share of uncertainty. Also, despite the large increase of the use of geospatial models in the past years and the likeliness that they will become an integral part of the data used for monitoring and accountability, there is still very scarce information of global health indicators at resolutions beyond what is offered by the household surveys. Moreover, if the quality of fit differs between surveys, it could also impact on the inequality measures. Lastly, the assessment of the scale effect using the inequality measures did not account for the uncertainty of each estimate.

Conclusions

This study attempted to quantify the magnitude of the scale effect of geographical units on empirical data and understand the impact on inequality measures and its variation over the years. We found that weighted inequality measures presented more stability in all scenarios especially for absolute measures, and the magnitude of the changes over time was not affected by the MAUP when comparing years at the same resolution. Further empirical studies are desired to validate the magnitude of the scale effect in other countries and to confirm that measuring geographic inequalities through time is viable and reliable.

Declarations

Ethics approval and consent to participate

The study used anonymized publicly available data. Ethical clearance was done by the institutions responsible for carrying out the original surveys.

Consent for publication

Not applicable

Availability of data and materials

Data on the national household surveys can be obtained from the Instituto Nacional de Estadística e Informática (https://www.inei.gob.pe/). Data sources from the covariates are listed in the supplementary material. Codes and further data can be obtained directly from the authors.

Competing interests

We have no competing interest to declare.

Funding

This work was supported by Wellcome Trust [Grant Number: 101815/Z/13/Z]; Bill and Melinda Gates Foundation [through the Countdown to 2030 initiative, OPP1148933]; and Associação Brasileira de Saúde Coletiva (ABRASCO).

Author's contributions

LF and AJDB conceptualized and designed the study. LF extracted the data, carried out the analysis, and wrote the first draft of the manuscript. All authors critically reviewed and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

References

1. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development [Internet]. Vol. 16301, General Assembley 70 session. 2015. p. 1–35. Available from: http://undocs.org/A/RES/70/1

2. Hosseinpoor AR, Bergen N, Schlotheuber A, Grove J. Measuring health inequalities in the context of sustainable development goals. Bull World Health Organ [Internet]. 2018 Sep 1;96(9):654–9. Available from: http://www.who.int/entity/bulletin/volumes/96/9/18-210401.pdf

3. Wong DWS. The Modifiable Areal Unit Problem (MAUP). In: WorldMinds: Geographical

Perspectives on 100 Problems [Internet]. Dordrecht: Springer Netherlands; 2004. p. 571–5. Available from: http://link.springer.com/10.1007/978-1-4020-2352-1_93

4. Hosseinpoor AR, Bergen N, Barros AJD, Wong KLM, Boerma T, Victora CG. Monitoring subnational regional inequalities in health: measurement approaches and challenges. Int J Equity Health [Internet]. 2016 Dec 28;15(1):18. Available from: http://www.equityhealthj.com/content/15/1/18

5. Speybroeck N, Paraje G, Prasad A, Goovaerts P, Ebener S, Evans DB. Inequality in Human Resources for Health: Measurement Issues. Geogr Anal [Internet]. 2012 Apr;44(2):151–61. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.2012.00842.x

6. Schuurman N, Bell N, Dunn JR, Oliver L. Deprivation Indices, Population Health and Geography: An Evaluation of the Spatial Effectiveness of Indices at Multiple Scales. J Urban Heal [Internet]. 2007 Jul 13;84(4):591–603. Available from: http://link.springer.com/10.1007/s11524-007-9193-3

7. Ferreira LZ, Blumenberg C, Utazi CE, Nilsen K, Hartwig FP, Tatem AJ, et al. Geospatial estimation of reproductive, maternal, newborn and child health indicators: a systematic review of methodological aspects of studies based on household surveys. Int J Health Geogr [Internet]. 2020 Dec 13;19(1):41. Available from: https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-020-00239-9

8. Instituto Nacional de Estadística e Informatica. Perú: Encuesta Demográfica y de Salud Familiar 2009. 2010.

9. Instituto Nacional de Estadística e Informatica. Perú: Encuesta Demográfica y de Salud Familiar 2019. 2020.

10. de Onis M. WHO Child Growth Standards based on length / height , weight and age. Acta Paediatr. 2006;Suppl 450(Suppl 450):76–85.

11. Osgood-Zimmerman A, Millear AI, Stubbs RW, Shields C, Pickering B V, Earl L, et al. Mapping child growth failure in Africa between 2000 and 2015. Nature [Internet]. 2018;555(7694):41–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29493591

12. Akseer N, Bhatti Z, Mashal T, Soofi S, Moineddin R, Black RE, et al. Geospatial inequalities and determinants of nutritional status among women and children in Afghanistan: an observational study. Lancet Glob Heal [Internet]. 2018;6(4):e447–59. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29454554

13. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. J R Stat Soc Ser C (Applied Stat [Internet]. 1998;47:299–350. Available from: http://doi.wiley.com/10.1111/1467-9876.00113

14. Utazi CE, Wagai J, Pannell O, Cutts FT, Rhoda DA, Ferrari MJ, et al. Geospatial variation in measles vaccine coverage through routine and campaign strategies in Nigeria: Analysis of recent household surveys. Vaccine. 2020 Mar;38(14):3062–71.

15. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. Vaccine [Internet]. 2018;36(12):1583–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29454519

16. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models

by using integrated nested Laplace approximations. J R Stat Soc Ser B (Statistical Methodol [Internet]. 2009 Apr;71(2):319–92. Available from: http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x

17. Tatem AJ. WorldPop, open data for spatial demography. Sci Data [Internet]. 2017 Dec 31;4(1):170004. Available from: http://www.nature.com/articles/sdata20174

18. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LP; 2019.

19. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.r-project.org/

20. Bell N, Schuurman N. GIS and Injury Prevention and Control: History, Challenges, and Opportunities. Int J Environ Res Public Health [Internet]. 2010 Mar 11;7(3):1002–17. Available from: http://www.mdpi.com/1660-4601/7/3/1002

Supplementary table 1 – Description of covariates considered for the analysis

Covariate	Year	Resolution	Unit	Source
Altitude	NA	5km	meters	SRTM (raster package)
Travel time to cities >50,000	2000	5km	minutes	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/
Distance to health facilities	2017	5km	meters	https://www.datosabiertos.gob.pe/dataset/minsa-ipress
Enhanced vegetation index	2009, 2019	5km	0 to 10000 (least to most vegetation)	https://ladsweb.modaps.eosdis.nasa.gov/search/history
Nighttime lights	2009, 2016	5km	nW/cm2/sr	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/
Improved water coverage	2009, 2019	5km	proportion	Interpolated using kriging
Improved sanitation coverage	2009, 2019	5km	proportion	Interpolated using kriging
Mean number of household members	2009, 2019	5km	number	Interpolated using kriging
% of households in Q1 or Q2	2009, 2019	5km	proportion	Interpolated using kriging
Mean women's years of education	2009, 2019	5km	proportion	Interpolated using kriging
% of indigenous population	2009, 2019	5km	proportion	Interpolated using kriging
Distance to protected areas	2009, 2017	5km	km	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/
Distance to build settlements (BGSM)	2009, 2017	5km	km	ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/PER/

Supplementary Table 2 - Covariates selected for the models in each survey year

Covariate	2009	2019
Altitude	X	X
Distance to health facilities		
Travel time to cities 50k	X	
Enhanced vegetation index		X
Nighttime lights		
Improved water	X	
Improved sanitation	X	X
% of households in Q1 or Q2	X	
Number of household members	X	X
Distance to built settlements		
Distance to protected areas	X	
Years of education (women)		X
% of indigenous population	X	

Supplementary Table 3 – Model validation metrics

	Cross-validation			In-sample		
Survey	**Correlation**	**Bias**	**MAE**	**Correlation**	**Bias**	**MAE**
Peru 2009	0.65	0.01	0.13	0.87	0.01	0.09
Peru 2019	0.56	0.00	0.10	0.75	0.00	0.08

Supplementary Table 4 – Comparison of predicted vs observed estimates at department-level (adm1) for stunting in 2009 and 2019

Department	2009			2019		
	Pred	**Obs**	**Diff**	**Pred**	**Obs**	**Diff**
Amazonas	31.1	27.1	-4.0	17.4	17.9	0.5
Ancash	32.6	28.2	-4.3	17.9	16.2	-1.7
Apurímac	37.7	34.6	-3.1	19.1	16.2	-2.8
Arequipa	9.9	12.2	2.3	5.6	6.1	0.4
Ayacucho	41.0	41.4	0.4	21.4	17.3	-4.2
Cajamarca	36.2	39.8	3.7	26.3	25.4	-0.9
Callao	7.4	6.0	-1.5	4.5	3.8	-0.7
Cusco	33.5	38.5	5.1	15.9	13.9	-2.1
Huancavelica	52.8	53.7	0.9	31.2	30.4	-0.8
Huánuco	37.9	39.9	2.0	22.5	19.3	-3.2
Ica	10.3	10.2	-0.1	5.6	5.5	-0.1
Junín	32.7	33.5	0.9	17.5	18.7	1.2
La Libertad	20.1	27.2	7.1	13.5	12.7	-0.8
Lambayeque	15.0	18.2	3.1	8.5	11.1	2.7
Lima Province	8.3	9.0	0.7	5.0	5.5	0.5
Loreto	32.6	28.9	-3.6	26.4	23.7	-2.8
Madre de Dios	15.3	12.6	-2.7	8.6	8.3	-0.3
Moquegua	8.4	5.3	-3.0	4.2	2.5	-1.8
Pasco	37.0	38.5	1.4	17.2	15.8	-1.4
Piura	21.5	22.9	1.3	13.1	13.0	-0.1
Puno	26.5	27.3	0.9	15.6	12.6	-3.0
San Martín	24.4	28.2	3.8	12.4	11.5	-0.9
Tacna	3.1	2.1	-1.0	3.0	2.4	-0.6
Tumbes	13.4	13.5	0.1	7.4	7.5	0.1
Ucayali	28.5	29.9	1.4	17.4	17.4	0.0

Press release [English]

Press release

Geospatial modeling is a powerful tool to reveal geographic inequalities and target health interventions

A study conducted at the International Center for Equity in Health from the Federal University of Pelotas evaluated the use of geospatial modeling methodologies as a tool to increase the quality and availability of data used to monitor inequalities in maternal and child health in low- and middle-income countries. These results were produced in the PhD research of the student Leonardo Zanini Ferreira under the supervision of professor Aluísio JD Barros. “Analyzing the estimates produced by these models, in comparison to those directly observed through the surveys, is almost like adding a magnifying glass to a huge map. Despite adding some degree of uncertainty, this allows us to assess specific areas in much more detail.”, comments Leonardo.

The researchers also assessed how these methodologies are being applied in the area of maternal and child health and the challenges of comparing geographic inequalities at different levels of aggregation. “We are talking about methodologies that have grown exponentially in the literature in recent years and that have much to contribute to the study of inequalities”. According to the authors, high-resolution maps are extremely powerful resources due to their ease of interpretation, in addition to their ability to present large amounts of data simultaneously. In Peru, a clear pattern of high coverage in the coastal areas and low coverage in the north and east of country emerges when estimates for the provinces are observed. And, especially in the jungle areas, the high-resolution maps reveal huge differences even within the provinces. “Managers often ask where the most vulnerable populations are and maps allow us to point directly at them. With these technologies, we continue to increase the chances of these marginalized groups to be identified and receive the care and attention that they are entitled to”, concludes the author.

Press release [Portuguese]

Nota de imprensa

Modelagem geoespacial é uma ferramenta poderosa para expor desigualdades geográficas e direcionar intervenções em saúde

Um estudo conduzido no Centro Internacional de Equidade em Saúde da Universidade Federal de Pelotas analisou o uso de metodologias de modelagem geoespacial como uma ferramenta para aumentar a qualidade e disponibilidade dos dados utilizados para o monitoramento de desigualdades em saúde materno-infantil em países de baixa e média renda. Estes resultados foram obtidos através da pesquisa de doutorado do estudante Leonardo Zanini Ferreira sob orientação do professor Aluísio JD Barros. “Analisar as estimativas produzidas por esses modelos, em comparação com as observadas diretamente através dos inquéritos, é quase como adicionar uma lupa a um enorme mapa. Mesmo com algum ganho de incerteza, isso nos permite avaliar áreas específicas com muito mais detalhes.”, comenta Leonardo.

Os pesquisadores também avaliaram como estas metodologias estão sendo aplicadas na área de saúde materno-infantil e os desafios de comparar desigualdades geográficas em diferentes níveis de agregação. “Estamos falando de metodologias que cresceram exponencialmente na literatura nos últimos anos e que tem muito a contribuir para o estudo das desigualdades”. Segundo os autores, mapas de alta resolução são recursos extremamente poderosos pela sua facilidade de interpretação, além de sua capacidade de apresentar grandes quantidades de dados simultaneamente. No Peru, um claro padrão de alta cobertura nas áreas de costa e baixa cobertura no norte e leste do país emerge quando estimativas para as províncias são observadas. E, especialmente para as áreas de selva, os mapas de alta resolução revelam enormes diferenças mesmo dentro das províncias. “Gestores costumam perguntar onde estão as populações mais vulneráveis e os mapas nos permitem apontar diretamente para eles. Com essas tecnologias, cada vez mais aumentamos as chances desses grupos marginalizados serem identificados e receberem o cuidado e atenção que lhes é de direito”, conclui o autor.