

Best practices in availability, management and use of geospatial data to guide reproductive, maternal, child and adolescent health programmes

Yordanos B Molla,¹ Kristine Nilsen,² Kavita Singh,³ Corrine Warren Ruktanonchai,² Michelle M Schmitz,⁴ Jennifer Duong,⁵ Florina Serbanescu,⁴ Allisyn C Moran,⁶ Zoe Matthews,⁷ Andrew J Tatem²

To cite: Molla YB, Nilsen K, Singh K, *et al*. Best practices in availability, management and use of geospatial data to guide reproductive, maternal, child and adolescent health programmes. *BMJ Global Health* 2019;**4**:e001406. doi:10.1136/bmjgh-2019-001406

Handling editor Seye Abimbola

Received 9 January 2019
Revised 28 March 2019
Accepted 13 April 2019

INTRODUCTION

Improving reproductive, maternal, newborn, child and adolescent health (RMNCAH) has been at the centre of international and national development strategies for decades. Closing the inequality gap in mortality and coverage of RMNCAH services continues to be central to global development and is a key objective of the renewed global commitment to 'leave no one behind'.¹ To successfully target poor and other hard-to-reach groups, stakeholders need up-to-date, highly disaggregated subnational estimates of RMNCAH service coverage, quality of care and staffing, which are not readily available in low-income and middle-income countries. The application of geospatial analysis in RMNCAH provides a unique advantage to develop subnational estimates of mortality and service coverage. For example, including geocoded attributes in health facility and household survey datasets can identify subnational inequality hotspots (places of significant inequality) with high-resolution RMNCAH coverage maps.

In a recent review of the application of geospatial methods to maternal and child health, Ebener *et al*² broadly categorised the field into three approaches in increasing order of complexity:

- ▶ thematic mapping (creation of basic maps to convey a topic or theme),
- ▶ spatial analyses (creation or extraction of new information from spatial data), and
- ▶ spatial modelling (spatial analysis with the use of mathematical or statistical models to simulate real-world phenomena).

Makanga *et al*³ noted two major research themes in the use of geographic information

Summary box

- ▶ The commentary provides a set of considerations and some examples for reproductive, maternal, newborn, child and adolescent health (RMNCAH) programmes that wish to use geospatial data.
- ▶ Monitoring RMNCAH data trends or change over time and estimating disease burden remain major challenges due to limited reliable geolocated data sources, inconsistent spatial denominators and technical capacity needs.
- ▶ Increased availability of spatial data, such as satellite imagery and geolocated survey and facility data, coupled with recent methodological refinements, has created new opportunities for use of geographic information systems to achieve spatial disaggregation of RMNCAH data and highlights subnational monitoring among vulnerable populations.
- ▶ More refined geospatial analyses can help to close the gap for countries with high maternity-related deaths and suffering.

system (GIS) in maternal care: modelling access to health services and analysis of risk factors associated with adverse maternal health outcomes. Molla *et al*⁴ discussed potential challenges of, and recommendations for, using GIS in maternal and newborn health including the need to develop standards and parameters critical for mapping.

We review best practices for application of geospatial analysis in RMNCAH while the associated papers in this supplement include examples of how RMNCAH programme has used geospatial analysis. These papers describe diverse geospatial methods and examine diverse topics in RMNCAH, including health service management and targeting of services, standardisation of indicators, and approaches for mapping physical



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Yordanos B Molla;
ymolla@pathfinder.org

access to services, referral system linkages, travel time for care and quality of services. Collectively, these papers raise the following important practical considerations for geospatial applications in RMNCAH.

PRACTICAL NEEDS AND LIMITATIONS OF USING GEOSPATIAL DATA

Technical considerations in geospatial analysis are often overlooked. While we propose that geospatial analysis needs to be used more often, certain criteria first need to be met. Understanding the gaps and limitations of geospatial data will help technical experts, RMNCAH programme implementers and decision-makers to use it more often and effectively at various levels.

Availability, quality and management of geospatial data

Data with geographical tags are quickly becoming open access/open source and their use in public health programme implementation is increasing despite uneven quality of shapefile and metadata. Using publicly available data without standardised guidelines means conclusions drawn from analyses of these data are of varying quality. Therefore, publicly available data should follow specific guidelines. Open Street Map,⁵ for example, applies the best practice of including guidance on open source geospatial data (Additional examples of these guideline include Humanitarian Data Exchange FAQs (includes a section on Metadata and a section on Data Quality): <https://data.humdata.org/faq>; Federal Geographic Data Committee (GEOSPATIAL METADATA STANDARDS AND GUIDELINES): <https://www.fgdc.gov/metadata/geospatial-metadata-standards>; Project Open Data Schema: <https://project-open-data.cio.gov/v1.1/schema/>).

Despite the increase in geotagged data availability, there is still a need for complete, higher-resolution geospatial data at different administrative boundaries and for multiple time periods. Alternative solutions for a lack of high-quality data include engaging satellite imagery or technology companies that may be interested in understanding global health needs and stay abreast of health-related geospatial resources development (eg, IBM is working on geospatial big data analytics via the PAIRS system (<https://www.ibm.com/us-en/marketplace/geospatial-big-data-analytics>) and Digital Globe is already providing satellite imagery (with a preference for United States Government clients): <https://www.digitalglobe.com/>).

Satellite imagery can provide proxy measurements for denominators commonly used in estimating key health indicators. Denominators such as population, urban/rural delineations, migration patterns and seasonal changes are critical for geospatial analysis.⁶

Lack of a centralised geospatial data source/repository and data variability between different sources create challenges for robust analysis, which requires access to reliable and readily available datasets. Using geospatial

data from different data sources, formats, aggregations, layers and time points for multiple layers may lead to confusion and misinterpretation of results. Data repositories that consolidate and harmonise information—such as the Integrated Public Use Microdata Series International,⁷ the world's largest collection of publicly available individual-level geotagged census data—are valuable new resources to geospatial analysis.

Definitions of key terms:

- ▶ Shapefile: spatial data format that stores shape attributes (points, lines or areas) and underlying data in tables.
- ▶ Geotagged: have sufficient geographical information, including projection system and coordinates, and time of data capture as necessary.
- ▶ High resolution: having or being capable of producing an image characterised by fine detail.
- ▶ Scalable: ideally can be viewed at both small and large resolutions—for large-scale and small-scale mapping, respectively. If the dataset cannot be used at both scales, provision of information about proper use is helpful.
- ▶ Metadata: data on how the geographical data were collected, or interpolated, and how it should be used.
- ▶ Up-to-date: Data should be no more than 5 years out of date—and should be constantly checked to make sure that it is functional.
- ▶ Publicly available: should be available without a licence.
- ▶ Raster data: Matrix of cells (or pixels) where each cell contains a value representing information.

Another critical consideration for geospatial analysis is that geospatial data management steps are often more labor-intensive than expected. Tabular, raster and vector data from numerous sources must be prepared, cleaned, merged and harmonised. Lack of information about each data source, including time period and metadata documentation, complicates data processing. Therefore, it is a good practice to prepare metadata for any spatial dataset, which can also include guidance on how to apply the data. Examples of good practices in attaching metadata to spatial files include WorldPop and Demographic and Health Survey (DHS) datasets.^{8,9} Similarly, an absence of standardised definitions of variables across sources makes a comparative analysis daunting. Therefore, linking geospatial data with existing standardised health information management systems, such as the commonly used District Health Information System 2, will help to standardise the process and translate the findings to the healthcare system.

Advanced technical considerations

Predictive modelling of indicators in places where measurements were not taken is one advanced solution to address lack of data availability across multiple geographical areas. This is becoming an increasingly common practice in spatial modelling, although the methodologies applied vary.^{9–11} Similar to any statistical modelling,

Table 1 Geospatial resources freely available to support reproductive, maternal, newborn, child and adolescent health analyses

Resource type	Content	Link for access	Main use
Software package	QGIS (a free and open-source geographic information system)	https://qgis.org/en/	To create maps for use in fieldwork planning, scientific analysis or M&E dissemination
Analysis tool	AccessMod (a tool for modelling physical accessibility to healthcare)	https://www.accessmod.org/	To create travel time surfaces that can estimate time needed to reach a facility given land cover, speed, road coverage, weather, etc
Data management tool	MEASURE Evaluation's Spatial Quality and Anomalies Diagnosis tool	https://www.measureevaluation.org/resources/tools/geographic-information-systems/squad-tool/spatial-quality-and-anomalies-diagnosis-squad-tool	To clean point-level facility and village data provided from fieldwork surveys
Data	Spatial distribution of births, pregnancies and women of childbearing age raster data	http://www.worldpop.org.uk/data/data_sources/	To estimate the spatial distribution of births, pregnancies and women of childbearing age, for use in spatial joins and survey models
	Spatially modelled surfaces for DHS indicators including family planning and maternal and child health indicators	https://spatialdata.dhsprogram.com/modeled-surfaces	To estimate the spatial distribution of various DHS indicators, for use in spatial joins and survey models
	DHS geospatial covariate contain data on topics including population, climate and environmental factors packaged as a zip and include a.csv data file	https://spatialdata.dhsprogram.com/covariates/	To conduct geospatial statistical analysis without having to manually source and link these covariates to cluster locations
	Integrated Public Use Microdata Series Terra (TerraPop) integrates population and environmental data	https://terra.ipums.org/	To conduct comparative analysis using environmental data and census data from over 160 countries around the world

DHS, Demographic and Health Survey.

there are underlying assumptions that make some models better than others. In addition to the model's strength in measuring the variation, the robustness of the prediction can be a result of the level of granularity of the sampling location distribution and the frequency of the event being measured.⁹ Therefore, one consideration in such techniques is a modelled surface map output where spatial uncertainties are measured, quantified and mapped to facilitate informed interpretation. Sharing the assumptions used and underlying statistical measures in parallel with the modelled output should be a standard practice in spatial modelling. Such practices have been documented in publicly available spatially modelled maps of DHS indicators,⁹ for WorldPop's global population distribution maps,⁸ and in the Malaria Atlas project.¹¹ These programmes are producing maps for modelled indicators in parallel with prediction uncertainty maps.

Maintaining confidentiality is the other critical issue to take into consideration when using geospatial products, since geographical location can enable identification of individuals in confined areas. Geomasking approaches,

which can preserve confidentiality of records by obscuring the underlying location, must be properly implemented to prevent identification while preserving sufficient detail for spatial analysis. The most common technique to preserve confidentiality of health records is record aggregation within the boundaries of geographical areas.¹² In maternal and child health, this technique allows estimation of such factors as maternal mortality ratio, antenatal care coverage, antiretroviral prophylaxis among pregnant women and skilled attendants at birth, within national and subnational boundaries.¹³

Recent advances in geographical technologies and data collection methods have increased the availability of georeferenced RMNCAH data and the opportunity for geomasking. Georeferenced verbal autopsies, for example, are aggregated to the community level to allow for geospatial analysis of causes of death.¹⁴ Population-based surveys use geographical displacement procedures to obscure locations by a few kilometres. This technique still allows identification of geographical patterns in demographic and health outcomes and

analysis of service availability and distance to care; it also permits aggregation to other geographical boundaries without considering specific boundary locations.¹⁵ Geomasked data, while potentially de-identified, will not be suitable for direct survey mapping or for small-scale mapping, as the data will no longer directly correspond to the point in space it aligns with.

GLOBAL INITIATIVES/RESOURCES

Geospatial data and tools can reveal greater insights into RMNCAH outcomes and assist with programme planning and/or resource allocation through spatial analysis and pattern mapping.⁴ Multiple recent endeavours have offered solutions to the aforementioned limitations and challenges of spatial data and analysis. Table 1 presents examples of open source spatial data, tools and software related to RMNCAH.

In summary, as new, higher-quality datasets are made available to the public, GIS professionals can work alongside global health practitioners in RMNCAH to conduct more complex analysis beyond thematic maps. More refined geospatial analyses can help to close the gap for countries with high maternity-related deaths and suffering. Initial investments in resources, personnel training, and time to prepare higher-quality datasets and conduct more refined analyses can be substantial. Investments in applying better quality data and more refined analytic techniques are critical to increasing the potential breadth and depth of geospatial analysis to understand and improve RMNCAH health outcomes.

Author affiliations

¹Pathfinder International, Washington, District of Columbia, USA

²WorldPop Research Group, University of Southampton, Southampton, UK

³MEASURE Evaluation/UNC, Chapel hills, North Carolina, USA

⁴US Centers for Disease Control and Prevention, Atlanta, Georgia, USA

⁵American Red Cross, Washington, District of Columbia, USA

⁶Global Health Fellows Program II, United States Agency for International Development (USAID), Washington, DC, USA

⁷Department of Social Statistics and Demography, University of Southampton, Southampton, UK

Contributors All authors contributed to the development and review of the commentary. All authors read and approved the final commentary.

Funding This manuscript is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the terms of the Cooperative Agreement AID-OAA-A-14-00028.

Disclaimer The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Samman ESaE. *Defining 'leave no one behind'*. London, UK: Overseas Development Institute, 2017.
2. Ebener S, Guerra-Arias M, Campbell J, et al. The geography of maternal and newborn health: the state of the art. *Int J Health Geogr* 2015;14.
3. Makanga PT, Schuurman N, von Dadelszen P, et al. A scoping review of geographic information systems in maternal health. *Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet* 2016;134:13–17.
4. Molla YB, Rawlins B, Makanga PT, et al. Geographic information system for improving maternal and newborn health: recommendations for policy and programs. *BMC Pregnancy Childbirth* 2017;17.
5. OpenStreetMap. Open Street Map Wiki: Beginners' Guide, 2017. Available: https://wiki.openstreetmap.org/wiki/Main_Page [Accessed 30 Sep 2018].
6. Tatem AJ. Mapping the denominator: spatial demography in the measurement of progress. *Int Health* 2014;6:153–5.
7. IPUMS International. Harmonized international census data for social science and health research, 2018. Available: <https://international.ipums.org/international/> [Accessed 30 Sep 2018].
8. WorldPop. WorldPop open access demographic data. What WorldPop, 2018. Available: <http://www.worldpop.org.uk/> [Accessed 30 Sep 2018].
9. Burgert-Brucker CR, Dontamsetti T, Mashall A, et al. Guidance for use of the DHS program modeled MAP surfaces. 2016; DHS spatial analysis reports No. 14. Available: <https://dhsprogram.com/publications/publication-SAR14-Spatial-Analysis-Reports.cfm> [Accessed 30 Sep 2018].
10. Diggle PJ, Giorgi E. Model-based Geostatistics for prevalence mapping in low-resource settings. *Journal of the American Statistical Association* 2016;111:1096–120.
11. Tatem AJ. The malaria atlas project: cartographic approaches to estimating populations at risk burden and elimination feasibility. Available: <http://www.healthdata.org/events/seminar/malaria-atlas-project-cartographic-approaches-estimating-populations-risk-burden-and> [Accessed 3 Oct 2018].
12. Brownstein JS, Cassa CA, Mandl KD. No place to hide — reverse identification of patients from published maps. *N Engl J Med* 2006;355:1741–2.
13. WHO. Accountability for Women's and Children's Health. Recommendation 2: Health indicators, 2018. Available: http://www.who.int/woman_child_accountability/progress_information/recommendation2/en/ [Accessed 30 Sep 2018].
14. Hanson C, Cox J, Mbaruku G, et al. Maternal mortality and distance to facility-based obstetric care in rural southern Tanzania: a secondary analysis of cross-sectional census data in 226 000 households. *Lancet Glob Health* 2015;3:e387–95.
15. Burgert CR, Colston J, Roy T, et al. *Geographic displacement procedure and georeferenced data release policy for the demographic and health surveys*. Calverton, Maryland, USA: ICF International, 2013.