How much is 'enough'? Exploring the impact of volunteered dataset completeness upon healthcare accessibility modelling

DOI:
10.5281/zenodo.6410144

How much is 'enough'? Exploring the impact of volunteered dataset completeness upon healthcare accessibility modelling

Kirsty Watkinson[*] [1], Jonathan Huck[†][1] and Angela Harris[‡][1]

[1]MCGIS, Department of Geography, University of Manchester

January 13, 2022

Summary

An inequality in access to healthcare exists globally, preventing many people from living long, healthy lives. Availability is an important component of access and is often used to inform healthcare policy. However, our ability to measure availability is impacted by the completeness of geographical datasets, in particular road and path datasets. This paper presents a novel sensitivity analysis whereby the completeness of a volunteered road and path dataset is incrementally downgraded and the change in distance to healthcare measured. In doing so we hope to answer "*how much data is 'enough'*?" to enable reliable estimations of healthcare availability.

KEYWORDS: Volunteered geographic information, healthcare, access, mapping inequalities

1. Introduction

1.1. Inequalities: Access to healthcare and geographic data

Access to healthcare is important to enable people to live long, healthy lives (Luqman and Khan, 2021). However, access to healthcare is highly unequal, with populations in rural regions of lower income countries often experiencing significantly poorer access (Dowhaniuk 2021). Access is a complex concept, comprising multiple inter-related dimensions often summarised as availability, affordability and acceptability (Iyer et al. 2020). Although affordability and acceptability are important, due to the importance of distance in determining access, especially in rural locations, measuring access tends to focus on the spatial dimension availability. Availability describes the distribution of available healthcare facilities and the route used to travel to them, and is often used to inform healthcare policy interventions (Iyer et al. 2020). Despite its importance to policy, the measurement of availability is often hampered by a lack of geographic data. Inequalities in the coverage of geographic data determine how availability is measured, and consequently access to healthcare may be over or underestimated (Verma and Dash 2020; Figure 1). It is therefore critical to understand the impact of data inequalities on our ability to measure spatial healthcare availability, to enable informed decision making and service allocation.

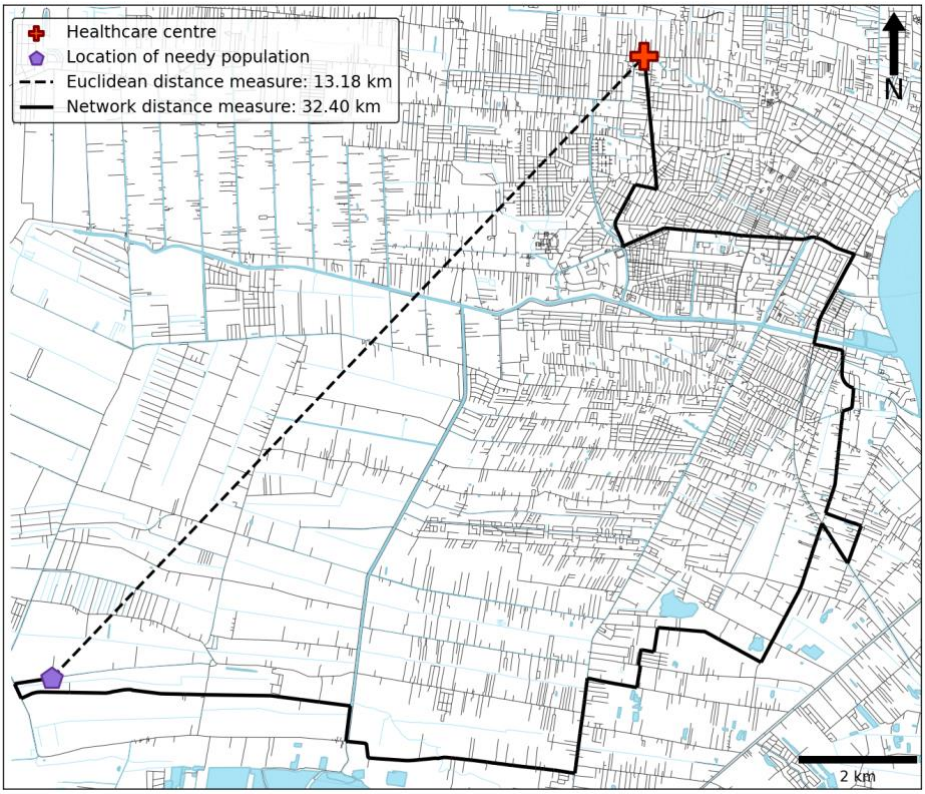

Figure 1 The difference in distance to healthcare measured using Euclidean and network-based measures. Example analysis run in Suriname.

1.2. Volunteered geographic information (VGI)

VGI has been suggested as a solution to deficiencies in geographic data (Huck et al., 2021). VGI enables the collection of up-to-date, local data relevant to health and provides health researchers with a "v*irtual audit instrument*" (Mooney, Corcoran and Ciepluch, 2013 p 741). However, using VGI for health research is limited due to concerns over its quality, particularly data completeness e.g. the presence or absence of features (Fonte et al. 2017). Difficulties engaging volunteers in the widespread production of VGI leads to an inequality in its production, often reproducing global inequalities in geographic data. Although efforts have been made to improve VGI production, through automation (Watkinson, Huck, and Harris 2021; Herfort et al. 2019) and gamification (Herfort et al. 2017), many features remain unmapped.

When considering the need for 'complete' datasets, it is important to consider the meaning of 'completeness'. Labelling a map as 'complete' is subjective and reflects the assertion of one person's understanding of 'complete' (Kitchen, Dodge, and Perkins, 2011). Completeness is scale-dependent and highly contextual, reflecting decisions on what can and should be included (Quinn and Leon, 2019). Acknowledging the challenges in defining 'completeness', this research aims to explore the notion of 'completeness' through a novel sensitivity analysis applied to the context of healthcare availability. Adopting a quantitative definition of 'completeness' (Fonte et al. 2017) we explore "*how much data is 'enough'*?" to provide reliable measures of healthcare availability. In doing so we hope to quantify how much VGI should be produced in a location before redirecting volunteer efforts.

2. Methods

To answer the question "*how much data is 'enough'*?", road and path networks from selected countries will be incrementally downgraded (road and path edges from the network randomly removed each time) and the mean distance to healthcare measured at each step. Changes in distance measured will be

compared between countries and a level of 'completeness' where changes in distance to healthcare become negligible identified.

2.1. Study areas

Countries were selected from a quantitative assessment of OpenStreetMap road network 'completeness' conducted by Barrington-Leigh and Millard-Ball (2019). To be selected countries had to meet two conditions: 1) Have an estimated road network completeness over 70% to ensure the dataset could be incrementally downgraded through sufficient iterations before reaching 0%, 2) Be classed as a 'developing' country by the United Nations Development Report. Countries classified as 'developing' were chosen as they are more likely to face distance-based barriers to healthcare access. In total, 36 countries were selected for analysis (Figure 2). Locations of healthcare centres and the road and path network were downloaded from the OpenStreetMap Geofabrik service and loaded into a PG Routing (PostgreSQL)[1] database using Osmosis[2] and OSM2PO[3].

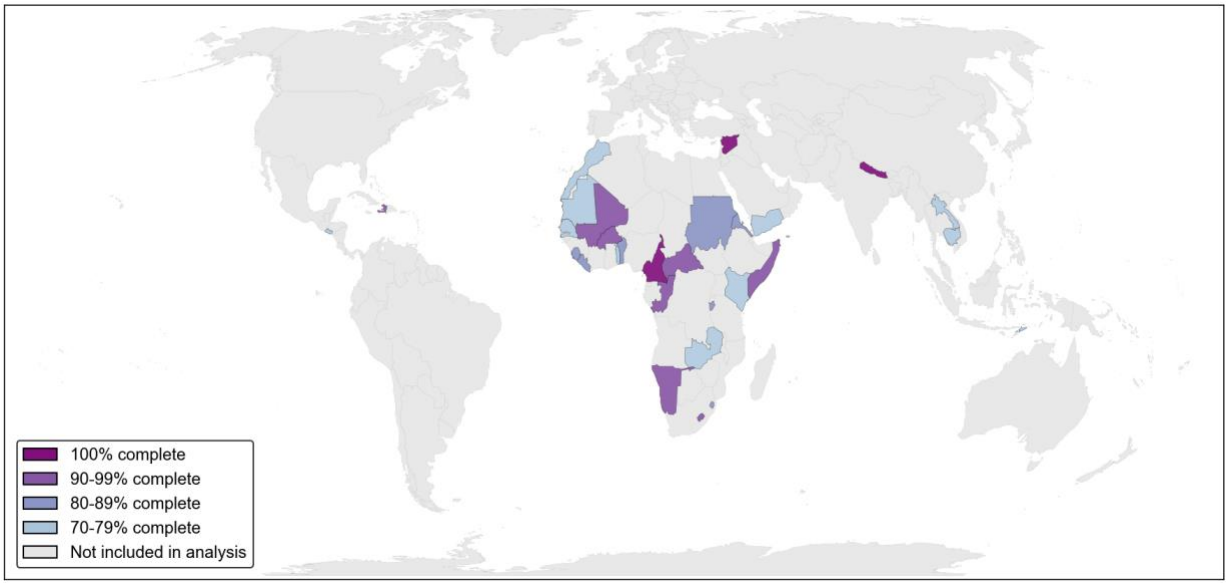

Figure 2 Selected countries with OpenStreetMap road networks over 70% complete as estimated by Barrington-Leigh and Millard-Ball (2019)

2.1. Sensitivity analysis

Distance to healthcare for each country was estimated using measurements from a set of randomly selected population points to their nearest healthcare centre. The sample size was proportional to the total population of the country, with one point for every 100,000 people. Distance was initially measured via the road and path network using the Dijkstra shortest path algorithm. Concurrently the Euclidean distance between the population point and healthcare centre was also calculated. When network degradation reached the stage that routes could no longer be calculated, the Euclidean distance was used instead. The distances between each population point and its nearest healthcare centre was summed, and a mean for each country calculated.

After each iteration of distance calculations, the road and path network was degraded by 5% until completeness was 0%. To ensure each road type (primary, secondary, tertiary, unclassified and path)

[1] https://pgrouting.org/
[2] https://www.osmosis.org/
[3] https://osm2po.de/

was degraded equally, the number of edges of each road type in the country was summed and 5% of its total calculated. The value was then used to turn off 5% of edges of each road type in the network. This degraded the network in total by 5% each time. The analysis was run 100 times for each country to account for variations in routing from the random selection of population points, and random degrading of the network.

3. Initial findings: Benin

Analysis shows that on average, as road and path network completeness increases mean distance to healthcare also increases (Figure 3). The largest increases in distance measured are observed between 40-50% completeness, and begin to plateau at 70%. On average, mean distance measured was 23.41% higher at 80% completeness than at 0%, equalling a 20.64 km increase in mean distance to healthcare. Though some variability was observed (minimum increase of 7.68% (5.7 km) and maximum of 33.17% (33.17 km)), our initial findings indicate that, for Benin, incomplete road and path networks can lead to a substantial underestimation of distance to healthcare.

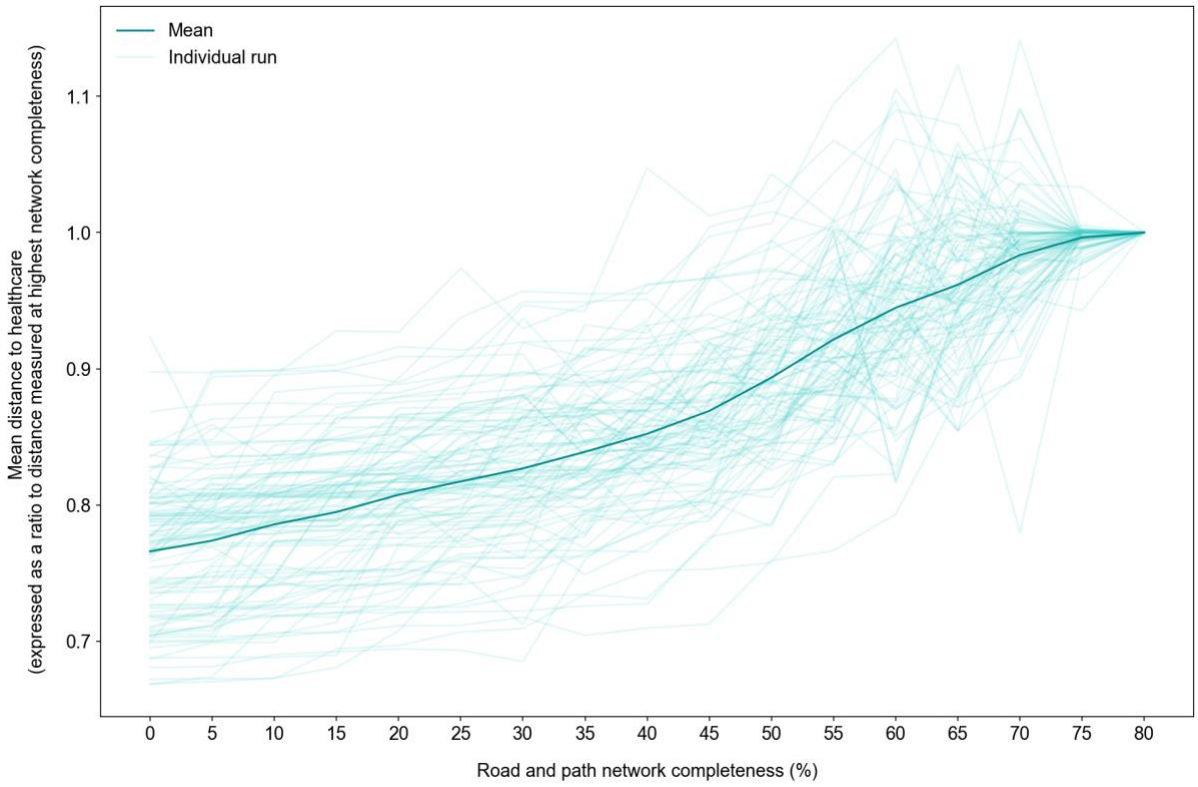

Figure 3 Change in distance to healthcare as the road and path network is degraded for the country of Benin.

4. Conclusion

Preliminary findings support suggestions of underestimation of distance to healthcare using Euclidean distance (Verma and Dash 2020), demonstrating the need for detailed road and path networks for reliable estimation of healthcare availability. Findings suggest a lessening effect around 70%, possibly indicating an acceptable level of 'completeness' at which volunteers can be redirected elsewhere. Further analysis will reveal whether this trend is replicated elsewhere, the sensitivity of analysis to the experiment parameters, and the relationship between the profile of the distance curve and predictor variables including the distribution of health facilities, population and network.

5. References and Citations

BARRINGTON-LEIGH, C. MILLARD-BALL, A. 2019. The world's user-generated road map is more than 80% complete. *PLOS ONE*. 12.

DOWHANIUK, N. 2021. Exploring country-wide equitable government health care facility access in Uganda. *International Journal for Equity in Health* 20(38).

FONTE, C. ANTONIOU, V. BASTIN, L. ESTIMA, J. JOKAR ARSANJANI, J. LASO BAYAS, J. SEE, L. VATSEVA, R. 2017. Assessing VGI Data Quality. In: *Mapping and the Citizen Sensor*: 137–163.

HERFORT, B. LI, H. FENDRICH, S. LAUTENBACH, S. ZIPF, A. 2019. Mapping Human Settlements with Higher Accuracy and Less Volunteer Efforts by Combining Crowdsourcing and Deep Learning. *Remote Sensing.* 11: 1799.

HERFORT, B. REINMUTH, M. DE ALBUQUERQUE, J. ZIPF, A. 2017. Towards evaluating crowdsourced image classification on mobile devices to generate geographic information about human settlements. In: *Proceedings of the 20th AGILE International Conference on Geographic Information Science*.

HUCK, J.J. PERKINS, C. HAWORTH, B.T. MORO, E.B., NIRMALAN, M. 2021. Centaur VGI: A Hybrid Human–Machine Approach to Address Global Inequalities in Map Coverage. *Annals of the American Association of Geographers.* 111.

IYER, H.S. FLANIGAN, J. WOLF, N.G. SCHROEDER, L.F. HORTON, S. CASTRO, M.C. REBBECK, T.R. 2020. Geospatial evaluation of trade-offs between equity in physical access to healthcare and health systems efficiency. *BMJ Global Health*. 5(:e00349).

KITCHIN, R. PERKINS, C. DODGE, M. 2009. Thinking about maps. In: DODGE, M. KITCHEN, R. PERKINS, C. *Rethinking Maps:* 1-25. Routledge.

LUQMAN, M. KHAN, S.U. 2021. Geospatial application to assess the accessibility to the health facilities in Egypt. *The Egyptian Journal of Remote Sensing and Space Science*.

MOONEY, P. CORCORAN, P. CIEPLUCH, B. 2013. The potential for using volunteered geographic information in pervasive health computing applications. *Journal of Ambient Intelligence and Humanized Computing*. 4: 731–745.

QUINN, S. LEÓN, L.A. 2019. Every single street? Rethinking full coverage across street-level imagery platforms. *Transactions in GIS.* 23: 1251–1272.

VERMA, V. DASH, U. 2020. Geographical accessibility and spatial coverage modelling of public health care network in rural and remote India. *PLOS ONE*.

WATKINSON, K. HUCK, J.J. HARRIS, A. 2021. Centaur VGI: Evaluating a human-machine workflow for increased productivity during humanitarian mapping campaigns. In: *Proceedings of the 29th Annual Geographical Information Science Conference.*

5. Acknowledgements

This research benefited from the following open source software, libraries and databases: PostgreSQL (https://www.postgresql.org/), PostGIS (https://postgis.net/), PG Routing (https://pgrouting.org/), Osmosis (https://www.osmosis.org/), OSM2PO (https://osm2po.de/) and OpenStreetMap

(https://www.openstreetmap.org). Computational facilities were provided by the MCGIS Research Group at the University of Manchester.

Biographies

Kirsty Watkinson is a PhD student in the Department of Geography at the University of Manchester researching the application of emerging technologies to support and improve the production of volunteered geographic information for humanitarian purposes.

Jonathan Huck is a Senior Lecturer in Geographical Information Science in the Department of Geography at the University of Manchester. He is interested in the application of maps, GIS and emergent technologies to geographical problems, particularly in the areas of health, conflict and the environment.

Angela Harris is a Senior Lecturer in Remote Sensing in the Department of Geography at the University of Manchester. Angela is interested in the innovative application of remote sensing to monitor the impact of environmental change on vegetation communities and improve our understanding of ecosystem physical processes.

---

[*] kirsty.watkinson@manchester.ac.uk

[†] jonathan.huck@manchester.ac.uk

[‡] angela.harris@manchester.ac.uk