Mapping the denominator: spatial demography in the measurement of progress

Andrew J. Tatem[a,b,c,*]

[a]*Department of Geography and Environment, University of Southampton, UK;* [b]*Fogarty International Center, National Institutes of Health, Bethesda, USA;* [c]*Flowminder Foundation, Stockholm, Sweden*

*Corresponding author: Present address: Department of Geography and Environment, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. Tel: +44 (0)238 059 2975; E-mail: A.J.Tatem@soton.ac.uk

Measuring progress towards international health goals requires a reliable baseline from which to measure change and recent methodological advancements have advanced our abilities to measure, model and map the prevalence of health issues using sophisticated tools. The provision of burden estimates generally requires linking these estimates with spatial demographic data, but for many resource-poor countries data on total population sizes, distributions, compositions and temporal trends are lacking, prompting a reliance on uncertain estimates. Modern technologies and data archives are offering solutions, but the huge range of uncertainties that exist today in spatial denominator datasets will still be around for many years to come.

Keywords: Burden estimation, Census, Demography, Health metrics, Mapping, Population

The field of health metrics has grown substantially in recent years, with new studies on the estimated national, regional and global burden of communicable and non-communicable diseases appearing every week. Methodological approaches to deriving such estimates have become increasingly sophisticated, especially for infectious diseases, where the high resolution mapping of subnational-scale risks, integrated with transmission models are becoming standard practices adopted by international organizations.[1–6] These approaches are generally based on samples, often from household surveys, to produce prevalence estimates at point locations or aggregate areas that are then modelled to produce complete large area coverage using covariates in spatial or space-time frameworks.[7] Such approaches typically require the use of population count data at similar spatial resolutions to provide a denominator, enabling conversion from estimated prevalences to numbers at risk or clinical cases, breakdowns by vulnerable groups and change estimates. Three issues exist, however, in terms of these input population data, relating to the estimation of country population sizes and distributions, population compositions and population change.

First, the question exists on whether we really know how many people there are living in many countries today, or a decade or more ago. The answer for some of the highest burden countries in terms of disease and general ill-health is a clear no. Without censuses undertaken in, for example, Pakistan for 16 years, Madagascar for 21 years and Nigeria for 8 years, current estimates are based on models that rely on multiple assumptions. In contrast, many other high disease-burden countries have undertaken regular and recent censuses (e.g., Chad 2009, India 2011 and Niger 2012). The uncertainties that these variations in availability of data leads to are well illustrated by the size of the variations in population size estimates produced by two of the leading and most widely used sources of country population data, the United Nations Population Prospects[8] and the Central Intelligence Agency World Factbook.[9] In terms of some of the worst cases, estimates made by the two organizations of the population sizes in 2014 of Angola (last census 1970), Democratic Republic of Congo (last census 1984) and Sierra Leone (last census 2004), for example, differ by 16%, 12% and 8%, respectively, while for other countries with more recent censuses, estimates are almost identical. This is also simply the estimation of national population totals, so when subnational distributions are required, the task of producing estimates becomes even more challenging for those countries with outdated or non-existent data. Thus, while uncertainty quantification has become sophisticated in estimating the prevalences of health conditions, only this side of the equation is generally considered when estimating the size of populations at risk, and uncertainty in reality is likely to vary depending on the country, and be substantially larger for some countries when those in the denominator are accounted for.[10,11]

Modern technology is offering solutions to tackling these wide variations in our knowledge of population numbers and distributions in resource-poor regions. High-resolution satellite imagery, processed using sophisticated image analysis techniques, are

enabling the large-scale mapping of built-up areas and individual buildings at unprecedented detail.[12,13] When combined with estimates of occupancy from ground surveys, these offer a 'bottom-up' approach to population-size estimation and mapping that potentially circumvents the requirement for census data. Further, the proliferation of mobile phones across the world provides opportunities for anonimized usage data to form the basis for rapid assessments of population distributions.[14] Finally, those countries that are implementing population censuses are increasingly making use of GPS technology to provide demographic data of unprecedented spatial detail.

Beyond estimates of population counts and distributions, the second major sticking point in the use of spatial demographic data is that of population composition. Vulnerable groups such as children under 5 years, women of childbearing age and the elderly remain the focus of the majority of international health studies, and are central to the Millennium Development Goals. Here, however, producing estimates of the numbers and spatial distributions of these vulnerable groups results in uncertainties to increase further, as input data becomes even sparser. Previous approaches to estimating vulnerable populations at risk have been limited by data availability and have simply taken existing spatial population count data and applied national level multipliers.[4–6,15–17] Analyses have shown that, on top of the existing issues with total population counts and distributions, such an approach leads to significant differences in vulnerable population at-risk estimates over accounting for the subnational variations that are universal in population age structures.[11] Solutions to these issues are less clear, but the growth in national household surveys, including the availability of cluster-level GPS coordinates are providing new contemporary and more spatially detailed data for improving estimates of vulnerable population distributions.

Measuring change and providing reliable denominators across multiple years represents a final challenge. Substantial population changes in terms of urbanization, migration and demographic shifts have taken place over the past decade and longer, particularly in those countries with the greatest burdens of ill health, yet reliable spatial data on these aspects remains sparse and inconsistent between countries and time periods. Ongoing projects are attempting to assemble what comparable information exists over multiple time points in terms of census, surveys, urban growth and migration data (e.g., The WorldPop project, Internal Migration Around the Globe and Integrated Public Use Microdata Series, International), while health and demographic surveillance systems are providing valuable information on trends over time in high disease-burden countries, and covering a range of geographies through efforts such as the INDEPTH network.[18] However, the reality remains that another significant source of uncertainty comes into the denominator equation when measuring progress in terms of changes in populations at risk, vulnerable groups covered by interventions or numbers vaccinated.

Spatial demographic datasets and production methods are rapidly improving, fuelled by improvements in technology and computing, but substantial limitations and uncertainties remain, particularly for those regions of the world where little data exists on how many people there are and where they live. Such uncertainties inherent in the demographic datasets used to provide denominators and processing steps taken are rarely acknowledged or accounted for, resulting in hidden uncertainties in many high impact disease-burden studies that are guiding international policies. If we want to be able to measure progress in tracking international health issues effectively, we need both methods to quantify the uncertainty inherent in spatial demographic data, and reliable denominator baselines from which to measure from. At present, for many of the resource-poor regions of the world these are still lacking.

References

1   WHO. World Malaria Report 2014. Geneva: World Health Organization; 2014.

2   Gething PW, Patil AP, Smith DL et al. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. Malar J 2011;10:378.

3   Bhatt S, Gething PW, Brady OJ et al. The global distribution and burden of dengue. Nature 2013;496:504–7.

4   Garske T, Van Kerkhove MD, Yactayo S et al. Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. PLoS Med 2014;11:e1001638.

5   Cairns M, Roca-Feltrer A, Garske T et al. Estimating the potential public health impact of seasonal malaria chemoprevention in African children. Nat Commun 2012;3:881.

6   Griffin JT, Ferguson NM, Ghani AC. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. Nat Comms 2014;5:3136.

7   Patil AP, Gething PW, Piel FB, Hay SI. Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol 2011;27:246–53.

8   United Nations Population Division. World population prospects, 2012 revision. Geneva: United Nations; 2012.

9   Central Intelligence Agency. The World Factbook. Washington D.C., USA: US Government Printing Office; 2014.

10 Tatem AJ, Campiz N, Gething PW et al. The effects of spatial population dataset choice on estimates of population at risk of disease. Popul Health Metr 2011;9:4.

11 Tatem AJ, Garcia AJ, Snow RW et al. Millennium development health metrics: where do Africa's children and women of childbearing age live? Popul Health Metr 2013;11:11.

12 Esch T, Taubenböck H, Felbier A et al. The path to mapping the global urban footprint using TanDEM-X data. Proc ISPRS 2011;34.

13 Pesaresi M, Ehrlich D, Caravaggi I et al. Toward global automatic built-up area recognition using optical VHR imagery. IEEE App Earth Obs Rem Sens 2011;4:923–34.

14 Bengtsson L, Lu X, Thorson A et al. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med 2011;8:e1001083.

15 Gething PW, Kirui VC, Alegana VA et al. Estimating the number of paediatric fevers associated with malaria infection presenting to Africa's public health sector in 2007. PLoS Med 2010;7: e1000301.

16 Murray CJ, Rosenfeld LC, Lim SS et al. Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet 2012;379:413–31.

17 Schur N, Vounatsou P, Utzinger J. Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. PLoS Negl Trop Dis 2012;6:e1773.

18 Sankoh O, Byass P. The INDEPTH Network: filling vital gaps in global epidemiology. Int J Epidemiol 2012;41:579–88.