

# Disease Mapping with Spatially Uncertain Data

Justin Manjourides<sup>\*1</sup>, Ted Cohen<sup>2,3</sup>, Caroline Jeffery<sup>4</sup> and Marcello Pagano<sup>5</sup>

<sup>1</sup>Dept of Health Sciences, Northeastern University, Boston, MA, USA; <sup>2</sup>Div of Global Health Equity, Brigham & Women's Hospital, Boston, MA, USA; <sup>3</sup>Dept of Epidemiology, Harvard School of Public Health, Boston, MA, USA; <sup>4</sup>Intl Health Group, Liverpool School of Tropical Medicine, Liverpool, United Kingdom; <sup>5</sup>Dept of Biostatistics, Harvard School of Public Health, Boston, MA, USA

## Objective

Uncertainty regarding the location of disease acquisition, as well as selective identification of cases, may bias maps of risk. We propose an extension to a distance-based mapping method (DBM) that incorporates weighted locations to adjust for these biases. We demonstrate this method by mapping potential drug-resistant tuberculosis (DRTB) transmission hotspots using programmatic data collected in Lima, Peru.

## Introduction

Uncertainty introduced by the selective identification of cases must be recognized and corrected for in order to accurately map the distribution of risk. Consider the problem of identifying geographic areas with increased risk of DRTB. Most countries with a high TB burden only offer drug sensitivity testing (DST) to those cases at highest risk for drug-resistance. As a result, the spatial distribution of confirmed DRTB cases under-represents the actual number of drug-resistant cases[1]. Also, using the locations of confirmed DRTB cases to identify regions of increased risk of drug-resistance may bias results towards areas of increased testing. Since testing is neither done on all incident cases nor on a representative sample of cases, current mapping methods do not allow standard inference from programmatic data about potential locations of DRTB transmission.

## Methods

We extend a DBM method [2] to adjust for this uncertainty. To map the spatial variation of the risk of a disease, such as DRTB, in a setting where the available data consist of a non-random sample of cases and controls, we weight each address in our study by the probability that the individual at that address is a case (or would test positive for DRTB in this setting). Once all locations are assigned weights, a prespecified number of these locations (from previously published country-wide surveillance estimates) will be sampled, based on these weights, defining our cases. We assign these sampled cases to DRTB status, calculate our DBM, repeat this random selection and create a consensus map[3].

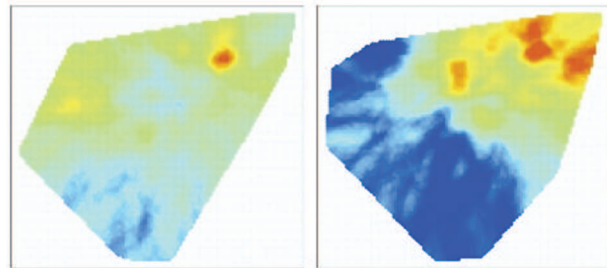
## Results

Following [2], we select reassignment weights by the inverse probability of each untested case receiving DST at their given location. These weights preferentially reassign untested cases located in regions of reduced testing, reflecting an assumption that in areas where testing is common, individuals most at risk are tested. Fig. 1 shows two risk maps created by this weighted DBM, one on the unadjusted data (Fig. 1, L) and one using the informative weights (Fig. 1, R). This

figure shows the difference, and potentially the improvement, made when information related to the missingness mechanism, which introduces spatial uncertainty, is incorporated into the analysis.

## Conclusions

The weighted DBM has the potential to analyze spatial data more accurately, when there is uncertainty regarding the locations of cases. Using a weighted DBM in combination with programmatic data from a high TB incidence community, we are able to make use of routine data in which a non-random sample of drug resistant cases are detected to estimate the true underlying burden of disease.



(L) Unweighted DBM of risk of a new TB case that received DST being positive for DRTB, compared to all new TB cases that received DST. (R) Weighted DBM of the risk of a new TB case that received DST being positive for DRTB, based on lab-confirmed DRTB cases and IPW selected non-DST TB cases, compared to all new TB cases.

## Keywords

surveillance; multiple addresses; distance based

## References

- [1] H Lin, et al. Assessing spatiotemporal patterns of multidrug-resistant and drug-sensitive tuberculosis in a south american setting. *Epi Infect*, 2010.
- [2] C Jeffery. Disease mapping and statistical issues in public health surveillance. PhD thesis, Harvard University, 2010.
- [3] J Manjourides, et al. Identifying multidrug resistant tuberculosis transmission hotspots using routinely collected data. *Tuberculosis*, 92(3), 2012.

\*Justin Manjourides

E-mail: Justin.manjourides@gmail.com

