Estimating subnational TB burden in Pakistan

Kate LeGrand, MPH, CPH
Jennifer Ross, MD, MPH

January 7, 2020
Overview

1. Tuberculosis epidemiology
2. Sources of data
3. Data challenges in the TB landscape
4. IHME geospatial modeling methods
5. Case study in spatial TB modeling
TB Epidemiology

Tuberculosis is the top infectious killer in the world

- Globally, 1 in 4 is infected and at risk of developing TB disease.
- Fell ill with TB in 2018, with 3 million undiagnosed or unreported cases.
- Eight countries account for 2/3 of all incident cases.
- In 2018, 1.5 million died from TB including 251,000 people with HIV.

Source: WHO Global tuberculosis report, 2019
Ending the epidemic requires data

- To identify priority populations for interventions
- To establish rates of progress
- To predict future trends and advocate for resources

Source: ESRI
Data Sources

• Data Library Services
  o Data Curators, Specialists, Analysts
  o Seek and intake data from a variety of sources
    — Public vs. Private availability
    — Data ownership
  o Maintain the Global Health Data Exchange
    — GBD Compare

• Network Engagement
  o Develop key relationships around the world
  o Collaborator network
TB models leverage challenging data

<table>
<thead>
<tr>
<th>Metric</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence</td>
<td>Case notifications are a biased estimate</td>
</tr>
<tr>
<td></td>
<td>• Care seeking, diagnostic limitations, incomplete reporting</td>
</tr>
<tr>
<td>Prevalence</td>
<td>Sparse data, as national prevalence surveys are expensive to implement</td>
</tr>
<tr>
<td></td>
<td>• No point-of-care test</td>
</tr>
<tr>
<td>Mortality</td>
<td>Vital registration systems are not present in most high-burden countries</td>
</tr>
</tbody>
</table>

### FIG. 3.34

National surveys of the prevalence of TB disease, actual (2000–2019) and planned (2020)

- **2000**: China
- **2001**: Cambodia
- **2002**: Cambodia
- **2003**: Malaysia
- **2004**: Indonesia
- **2005**: Eritrea
- **2006**: Thailand
- **2007**: Philippines, Viet Nam
- **2008**: Bangladesh
- **2009**: Myanmar
- **2010**: China
- **2011**: Cambodia, Ethiopia, Lao PDR, Pakistan
- **2012**: Gambia, Nigeria, Rwanda, UR Tanzania, Thailand
- **2013**: Malawi, Ghana, Sudan
- **2014**: Indonesia, Zambia, Zimbabwe
- **2015**: Bangladesh, Kenya, Mongolia, Uganda
- **2016**: DPR Korea, Philippines
- **2017**: Mozambique, Myanmar, Namibia, South Africa, Viet Nam
- **2018**: Nepal, Eswatini
- **2019**: India, Lesotho
- **2020**: Botswana

WHO Global TB Report, 2019
Global vs. Local Burden of Disease

- **Location matters**
  - National (or even subnational) averages may obscure finer-scale geographic trends
IHME Geospatial Modeling

• Small area estimation
  o Administrative areas (polygons)
    o Draw strength to estimate from neighboring regions, years, and ages

• Model-based geostatistics
  o Points
    o Use observed data to make prediction (and uncertainty estimate) for areas that are unobserved

KIT TB Hackathon 2019

- Collaboration between KIT Royal Tropical Institute, Stop TB Partnership, and the Pakistan National Tuberculosis Control Program
- Data: 2010 NTP survey and any publicly available data source
- Goal: Estimate district-level prevalence in 2018
- Teams: IHME, IDM, EPCON, Univ. of Sheffield, and Univ. of Milan
- Presentation: The Union World Conference for Lung Health, India
Epidemiological covariates

• Created a custom covariate for conflict and violence
  o Conflict (riots + protests) and violence against civilians
Epidemiological covariates

- Access: Travel time to nearest settlement

*Figure 1 | Global map of travel time to cities for 2015. The accessibility map has a spatial resolution of approximately $1 \times 1$ km, spans $60^\circ$ south to $85^\circ$ north latitude, and enumerates travel time to the city with the shortest associated journey.*

IHME Pakistan Geostatistical Model Results

• Results
  o National prevalence, age 15+, decreased from 2010 to 2018
  o District-level prevalence range varied greatly across the country
  o Model accurately predicts points with high-prevalence
  o Model over-predicts areas with low prevalence and sample sizes which proved to be a challenge across competitors
  o Wide uncertainty in unsampled areas
HACK TB
Estimating subnational TB burden in Pakistan

Coordination:
- Pakistan National TB Program
- KIT Royal Tropical Institute
- Stop TB Partnership
- International Union Against Tuberculosis and Lung Disease

Team members:
Audrey Batzel
Brigette Blacker
Nat Henry
Kate LeGrand
Bobby Reiner
Jennifer Ross
Emma Spurlock
Mingyou Yang

Internal review:
Zahid Butt
Simon Hay
Hmwe Kyu
Jorge Ledesma
Ali Mokdad

Data curation:
Karly Williams

Funding:
- **JMR:** National Institute of Allergy and Infectious Diseases, National Institutes of Health
- **IHME:** Bill & Melinda Gates Foundation

---

Kate E. LeGrand, MPH, CPH
Geospatial Tuberculosis Data Extraction Analyst
Institute for Health Metrics and Evaluation
kel15@uw.edu

Jennifer M. Ross, MD, MPH
Acting Assistant Professor, International Clinical Research Center
UW Departments of Global Health and Medicine (Infectious Disease)
Visiting Faculty, Institute for Health Metrics and Evaluation
Jross3@uw.edu
Appendix

• Discussion themes
• Model selection and validation
• Model equation
• Data challenges in the TB landscape
• Case study results
Model selection and validation

Correlation between data and modeled estimates
Observed in 2011. R squared: 0.5388

Sample Size
- 1300
- 1500
- 1700
- 1900

Modeled value at pixel (median, per 100k)

Raw data estimate (per 100k)
Model Equation

- Spatially explicit linear regression
- Assumed binomial distribution of bacteriologically-confirmed TB cases detected in a cluster

\[ \text{logit} \left( p_{s,t} \right) = \alpha + \beta X_{s,t} + Z_s \]

- Probability of having TB in a particular space-time location
- Probability of having TB when all covariates are equal to zero
- Vector of fixed-effect coefficients for set of space-time covariates
- Intercept for spatially-correlated residual variation that is not accounted for by fixed effect terms