

---

# Mapping the relationship between tuberculosis burden and case notifications in Uganda

---

Nathaniel Henry

## 1 Introduction

In 2015, the World Health Assembly committed its member states to the ambitious goal of ending the tuberculosis (TB) epidemic by 2035. In response, the World Health Organization (WHO) developed the global End TB strategy, which succeeded the Directly Observed  
5 Therapy (DOTS, 1994-2005) and Stop TB (2006-2014) plans as the coordinating principles for addressing the tuberculosis epidemic worldwide.<sup>1</sup> The End TB strategy outlines three ambitious targets to achieve by 2030: compared to 2015, the plan aims to reduce the number of TB deaths by 90%, reduce the number of incident TB cases by 80%, and ensure that 100% of families are protected from catastrophic costs due to TB.<sup>2</sup> These goals are consistent  
10 with Sustainable Development Goal 3.3, which commits United Nations member countries to end the TB epidemic by 2030.<sup>3</sup> Timely and accurate TB case detection serves as a critical component for two of the three pillars central to the End TB strategy: early detection enables for adequate patient-centered care, while mandatory and universal registration of TB disease allows countries to properly allocate resources and track progress towards goals.<sup>4</sup>

15 To achieve the End TB Strategy, the World Health Organization has developed tools for systematically building health system capacity<sup>5-7</sup> and has tracked progress and shared best practices through the release of annual Global Tuberculosis Reports.<sup>3</sup> Both these reports and external analyses have concluded that while many countries have increased their capacity to identify and treat tuberculosis, particularly drug-resistant TB and TB among people living  
20 with HIV, the majority of high-burden countries are not on track to meet the End TB 2030 targets.<sup>3,8</sup> Developing national health system capacity to detect and treat active TB cases is a necessary condition for the eradication of the TB epidemic in high-burden countries.<sup>3</sup> Among its many benefits, successful tuberculosis surveillance allows policymakers in high-burden countries to develop equitable interventions for those with TB burden. According  
25 to the WHO:

Disaggregation [can] inform analysis of within-country inequalities and associated assessments of equity, with findings used to identify areas or subpopulations where progress is lagging behind and greater attention is needed. Such disaggregation is also an important consideration for the TB community, given  
30 the influence of sex, age, socioeconomic status and differential access to health care on the risks and consequences of TB infection and disease.<sup>3</sup>

The recent history of TB control in Uganda, one of the 30 high-burden countries recognized by the end TB program, exemplifies both the possibilities for health system progress and the remaining challenges to TB elimination in low-resource settings. In Uganda, TB control is  
35 overseen by the National Tuberculosis and Leprosy Programme (NTLP) within the Ugandan Ministry of Health. The NTLP's 2015/16-2019/20 strategic plan mirrors the core objectives of the global End TB plan, calling for a 5% reduction in TB incidence as well as a 30% increase in case notifications over a four-year time period.<sup>9</sup> While the 2019-2020 NTLP report indicated impressive gains in TB treatment since 2015, the case notification target  
40 was barely missed (26% increase as opposed to 30%), while the estimated number of total TB cases increased slightly from 2017 to 2019.<sup>10</sup> Two of the principle challenges to universal TB detection and treatment identified in the 2020 report were (1) stubbornly low case detection rates in the far eastern and southwestern regions of the country; and (2) the threat of service disruptions related to health crises such as COVID-19, which could reduce health system  
45 capacity and adherence to TB treatment regimens.<sup>10</sup>

Tuberculosis control programs in high-burden countries need tools to track progress towards TB case detection and treatment goals across the country using routinely-collected data. In this chapter, I present a model for tracking both tuberculosis prevalence and case notification completeness at the district level across Uganda. This approach synthesizes multiple data  
50 sources available to the Uganda NTLP to provide more precise and reliable estimates of district-level TB burden than existing methods.

## 1.1 TB mapping in high-burden settings

In recent years, TB control programs in lower-resource settings have shown increasing interest in small-area and spatial analyses of tuberculosis burden.<sup>11</sup> TB prevalence and incidence  
55 are expected to vary locally within a country based on differences in underlying risk factors for the disease; by identifying these differences, programs can more efficiently allocate resources to target locations and sub-populations with the highest TB burden.<sup>12</sup>

In countries with high TB burden, primarily in low-income settings, two data sources have historically been used to generate estimates of TB burden. TB prevalence surveys, consid-

ered to be the “gold standard” of TB data collection, estimate TB prevalence at the national level based on sampled cluster locations across a country. These surveys use a well-defined screening process to identify potential TB cases within the eligible population, typically 15 years of age and above, and then conduct lab tests on the sputum of potential cases to confirm the presence of TB.<sup>13</sup> Depending on the underlying TB prevalence in a country and the overall sample size, each TB prevalence survey may identify as few as several dozen TB cases. Given the high costs associated with conducting TB prevalence surveys, most high-burden countries have conducted no more than two of these surveys over the past two decades.<sup>3,13</sup>

A more abundant data source for TB burden is available to countries in the form of aggregate case notifications reported to national TB control programs. Most countries mandate hospitals to report diagnosed TB cases, along with select demographic and diagnostic information, at regular intervals; in countries using electronic health information systems such as the DHIS2, reporting and aggregation of notified TB cases may happen continuously.<sup>6,14</sup> Aggregate case notifications are a potentially appealing source of subnational data due to their broad geographic coverage and wide availability. However, in many cases, counts of active TB diagnoses do not fully capture true TB incidence due to lack of access to health care, insufficient capacity to diagnose, or misdiagnosis with another lung condition.<sup>15</sup> Even clinically-diagnosed TB cases, particularly those diagnosed in private settings, may be missed in national statistics due to gaps in the reporting cascade.<sup>11,13,16</sup>

Although previous studies have explored spatial variation in TB incidence and prevalence within high-burden countries, no widely-accepted standards exist for assessing subnational variation in TB. In a recent systematic review, Shaweno and colleagues identified 168 studies that conducted spatial analyses of TB, including 57 in high-burden settings.<sup>17</sup> Of the 168 identified studies, 161 (96%) used TB case notifications as the data underlying the spatial analysis, although no studies accounted for spatial variation in case notification under-reporting. This oversight may be deeply problematic in high-burden settings: in another spatial assessment conducted in Bangladesh, Nepal, and Pakistan, Van Gurp and colleagues found that the estimated completeness of TB case notifications were correlated with programmatic factors such as the testing rate, facility density, and test positivity rate.<sup>18</sup> The authors concluded that subnational TB case notification rates were more likely to reflect subnational variation in healthcare access and programmatic factors than underlying TB burden.

The limitations identified in case notifications present a a major challenge to assessing spatial variability in TB burden across low-resource settings. Given the small sample sizes and low

95 spatial coverage of many TB prevalence surveys, data from these surveys alone may be  
insufficient to estimate subnational variation in TB burden without the addition of other  
predictive spatial data. In the current chapter, I investigate whether case notification data  
can be combined with TB prevalence data to estimate spatial variation in TB burden under  
the assumption that notification completeness also varies across space and time. I develop  
100 this exploration in the context of Uganda, a high-burden country with a single prevalence  
survey conducted in 2014/15 as well as an annual series of subnational notifications.

Uganda's diverse geography and population both imply that the risk factors for tuberculosis  
are unevenly distributed across the country. Previous studies and the experiences of the  
NTLP confirm that TB burden can vary widely across Uganda's districts and regions.<sup>9,15,19</sup>  
105 As of 2019, Uganda was divided among 128 administrative districts grouped into four regions.  
Figure 1 shows the regions and major population centers to provide context for the data  
sources and results presented below.

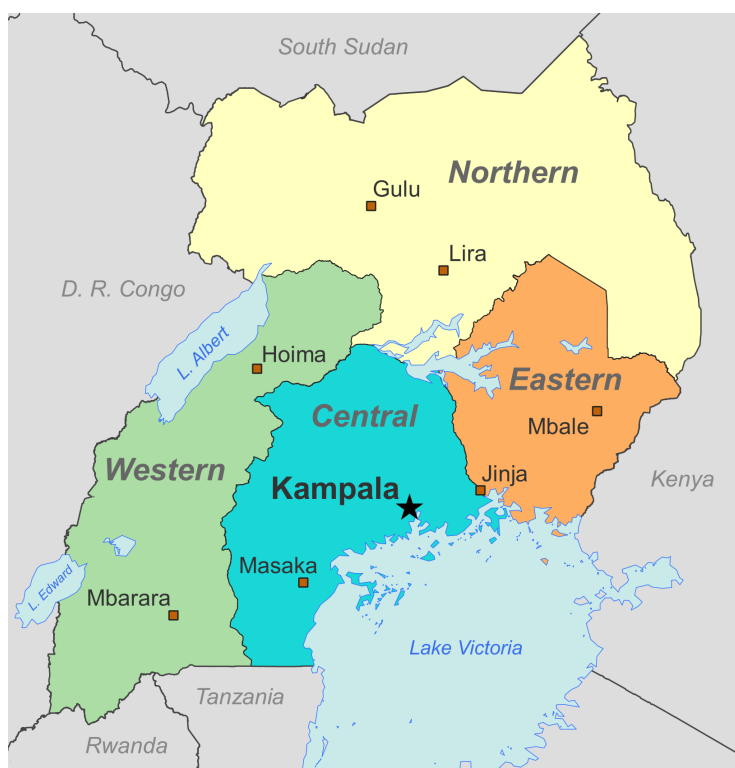


Figure 1: Regions and population centers of Uganda.

## 2 Methods

In this chapter, I develop a geostatistical modeling approach that simultaneously estimates  
110 TB prevalence and case notification completeness by district across Uganda. This approach

synthesizes data from a single national TB prevalence survey an annual time series of case notifications, two data sources that are widely used for TB surveillance in high-burden settings. I first develop crosswalks between the differing age groups, case definitions, and indicators reported across annual notifications and the national TB prevalence survey. I  
115 then present a geostatistical model that incorporates TB prevalence survey data directly as a sample of underlying TB prevalence, and incorporates TB case notifications indirectly as a sample of underlying TB incidence modified by spatially-varying incompleteness.

## 2.1 Data preparation

The joint geostatistical modeling approach combines data from a national TB prevalence  
120 survey conducted in 2014-2015, a time series of TB case notifications from 2015-2019, and predictive spatial covariates.

I downloaded and extracted cluster-level results from the 2014-15 Uganda National TB Prevalence Survey, including place names associated with each cluster.<sup>20</sup> Using spatial information from Google Maps and OpenStreetMap, I associated each cluster with the GPS  
125 location of the sampled township. I then extracted the number of lab-confirmed TB cases as well as the total tested population within each cluster, and summed these values to get district-level totals for each cluster. The prevalence survey tested sputum samples using both a smear test and a culture test; I included individuals who tested positive for either test to approximate the case definition for a confirmed pulmonary TB case in a clinical  
130 setting.<sup>21</sup> Of the 128 districts in Uganda, 56 were associated with at least one cluster from the 2014-15 prevalence survey. Figure 2 shows these raw prevalence estimates by district.

Total case notifications were reported by district across Uganda from 2015 through 2019 in annual reports published by the Uganda National TB and Leprosy Programme. I extracted the total number of diagnosed individuals associated with each district and year from these  
135 reports. Although district-level case notifications were reported for all age groups combined, including both pulmonary and extra-pulmonary cases, another set of tables reported total cases by age group and pulmonary/extra-pulmonary cases across 12 program regions in the country. I associated each district with its associated TB program region, and then corrected the total number of cases downwards based on the estimated fraction of total TB  
140 cases were pulmonary cases among adults aged 15 and over. Across the 12 TB program regions, the proportion of diagnosed extra-pulmonary cases ranged from 2% in the Hoima program region, near Lake Albert, to 19% in the Moroto program region, which borders Kenya in the far east. Meanwhile, the proportion of cases diagnosed in children under age 15 varied from 6% in the Lira program region in Northern Uganda to 19% in the Moroto

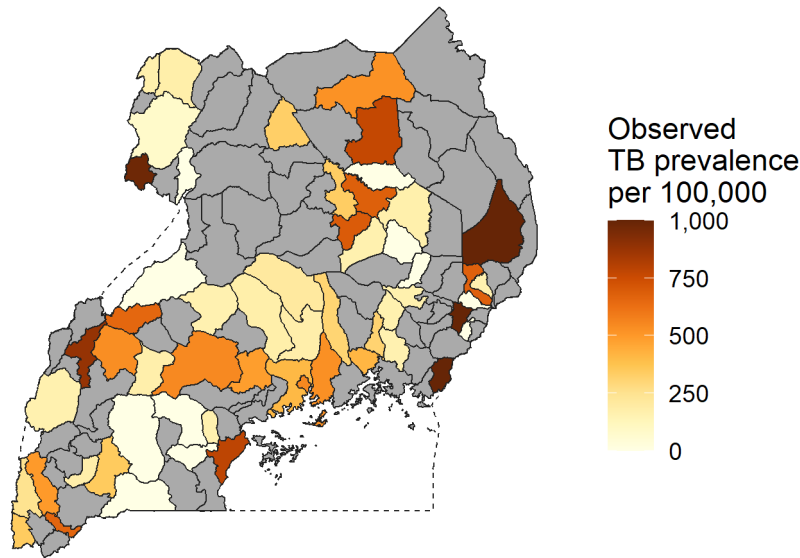


Figure 2: Unadjusted prevalence data from the 2014-2015 Uganda National TB Prevalence Survey, aggregated by district.

145 program region.<sup>22</sup> While the age structure of districts could affect the proportion of children under age 15 diagnosed within each district, a preliminary investigation revealed that the proportion of under-15s varied by less than 5% across any two districts within the same program region. Gridded estimates of the over-15 population from the WorldPop project were aggregated to the district level and used as population denominators.<sup>23</sup> Figure 3 shows  
 150 the case notification rate for pulmonary TB among adults aged 15 and above for the two years where district-level case notifications were reported.

In populations where disease burden is relatively stable, the prevalence and incidence of a disease are related by the average duration of the disease:  $Prevalence = Incidence * Duration$ .<sup>24</sup> When treated, the duration of TB is widely accepted to be six months,<sup>3</sup> while  
 155 the average duration of untreated pulmonary TB is estimated to be approximately three years.<sup>25</sup> In a study of two districts in Uganda, the median delay between the onset of active tuberculosis and the beginning of treatment was estimated at four weeks.<sup>26</sup> An estimate for the average duration of disease would combine the experiences of untreated individuals and treated individuals after accounting for treatment delay, and would fall somewhere  
 160 between the two extremes of six months and three years. In their most recent published estimates that include estimates of both TB prevalence and incidence, both the WHO and the Institute for Health Metrics and Evaluation (IHME) estimate that the ratio between

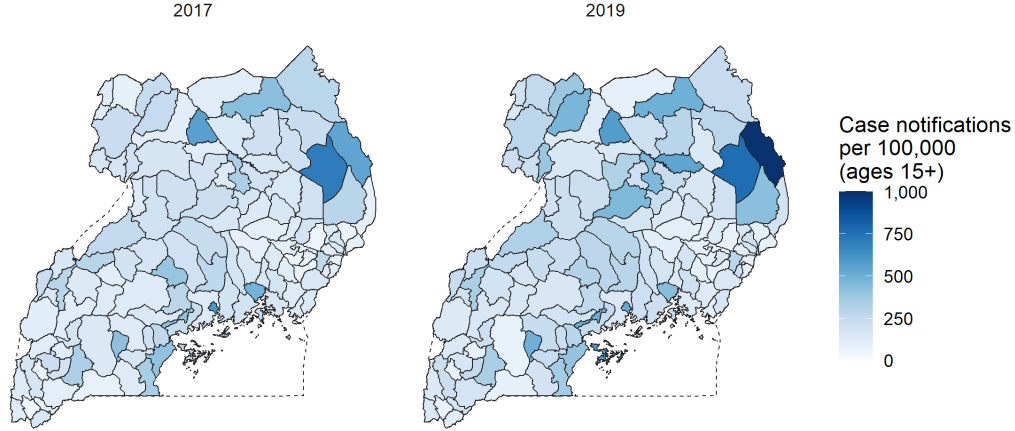


Figure 3: Data from the 2014-2015 Uganda National Tuberculosis Prevalence Survey, geopositioned and aggregated to the district level. Out of 128 districts, 56 contained underlying prevalence survey clusters.

TB prevalence and incidence in Uganda is approximately 1, indicating an estimated average duration of approximately one year.<sup>3,8</sup> I use this estimate to crosswalk between measures of TB incidence and prevalence in the joint geostatistical model.

I also include spatial covariates in this model to inform variation in both the TB burden and case reporting completeness surfaces. The TB prevalence surface varies according to known risk factors for tuberculosis in the context of Uganda: these included subnational estimates of the Human Development Index as well as cattle per capita, used as an estimate of a pastoral lifestyle.<sup>27</sup> Following Van Gorp and colleagues,<sup>18</sup> I included covariates for the reporting completeness surface that are associated with health system access: these included average travel time to health facilities, nighttime light intensity, and urban land cover.<sup>28,29</sup> All gridded covariates were aggregated to the district level using population weights derived from the WorldPop project.<sup>23</sup>

## 2.2 Small area model incorporating prevalence data and incomplete notifications

I developed a geostatistical model to jointly estimate TB prevalence and TB case notification reporting completeness by district between 2015 and 2019. TB prevalence,  $p$ , is represented by a time-invariant spatial surface that varies according to underlying covariates as well as a spatially-structured latent process. Notification completeness,  $\pi$ , is represented as a space-time surface that varies according to underlying covariates as well as a spatially-structured latent process and fixed effect with a random slope by year.

Specifically, the TB prevalence surface is estimated to vary as follows:

$$\log(p_s) = \alpha^p + \overrightarrow{\beta^p} X_s^p + Z_s^p$$

Here, TB prevalence  $p$  is a log-linear surface that is indexed by district (s). Prevalences  
 185 varies according to an intercept,  $\alpha$ , fixed effects  $\overrightarrow{\beta^p}$  on spatial covariates  $X_s^p$  that vary by  
 district, and a structured random effect  $Z_s^p$  that fits excess variation not captured by fixed  
 effects. The latent spatial surface is parameterized using the interpretable BYM2 spatial  
 model formulation described by Riebler and colleagues.<sup>30</sup> The superscript (p) differentiates  
 these terms from a similar spatial field used to estimate reporting completeness ( $\pi$ ).

190 Simultaneously, the reporting completeness surface is estimated to vary as follows:

$$\text{logit}(\pi_{s,t}) = \alpha^\pi + \overrightarrow{\beta^\pi} X_{s,t}^\pi + (\gamma^\pi + \delta_s^\pi) t' + Z_s^\pi$$

In this formulation, reporting completeness  $\pi$  is a logit-linear surface that is indexed by  
 district (s) and year (t). Completeness varies according to fixed effects on covariates that  
 vary by district and year, as well as a structured random effect  $Z_s^\pi$  that is parameterized  
 in the same manner as  $Z_s^p$ , according to the adjacency structure of Ugandan districts. Re-  
 195 porting completeness also varies according to a fixed effect on time with a random slope  
 ( $\gamma^\pi + \delta_s^\pi$ ). This term is multiplied by the year ( $t'$ ) of a particular observation, where the  
 years 2015-2019 have been rescaled to have mean zero and range one.

Data from the TB prevalence survey, with numerators  $Y_i^{prev}$  and denominators  $N_i^{prev}$ , are  
 evaluated directly against the underlying prevalence surface:

$$Y_i^{prev} = \text{Poisson}(N_i^{prev}, p_{s(i)})$$

200 If only data from the TB prevalence survey was incorporated into this model, it would  
 correspond to a standard small-area estimation framework. However, in addition, case  
 notifications with reported cases  $Y_i^{notif}$  and corresponding population denominators  $N_i^{notif}$   
 are evaluated against the fitted surfaces as follows:

$$Y_i^{notif} = \text{Poisson}(p_{s(i)} / D * N_i^{notif} * \pi_{s(i),t(i)})$$

The observed number of case notifications are estimated to be centered around the true  
 205 incident cases in the population multiplied by case notification incompleteness. The true



incidence rate is estimated as the fitted underlying prevalence in a given district,  $p_{s(i)}$ , multiplied by the inverse of the duration,  $1/D$ , which has been fixed at one. The true underlying incidence of pulmonary TB is multiplied by the over-15 population in a given district-year,  $N_i^{notif}$ , to obtain the true estimated incident cases in each district and year. Finally, the true estimated incident cases are multiplied by the fitted case reporting completeness in a given district and year,  $\pi_{s(i),t(i)}$ , to obtain an estimate of the number of observed cases in a particular district and year.

I assigned priors to all model parameters and then fit the model using the Laplace approximation for mixed-effect parameter estimation.<sup>31,32</sup> I fit two versions of the model: one incorporating both TB prevalence survey results as well as case notifications, and a prevalence-only model incorporating just data from the TB prevalence survey. The model was fit in R v.4.0.3 using the package Template Model Builder v.1.7.18.<sup>31,33</sup>

### 3 Results

The joint spatial model identified that TB prevalence varies over 10-fold across districts in Uganda, while the completeness of TB case notification varied from less than 1 in 4 cases reported to near-complete reporting in 2019, the final year of analysis. A joint model that incorporated both TB case notifications and prevalence data predicted TB prevalence with substantially greater precision than a model using only data from the 2014-15 national TB prevalence survey.

#### 3.1 Spatial variation in tuberculosis prevalence across Uganda

Across the 128 districts of Uganda, estimated pulmonary TB prevalence in adults ranged from 140 per 100,000 (95% uncertainty intervals: 134 to 147) in the district of Wakiso bordering Kampala, to a maximum of 1,468 per 100,000 (1,108 to 1,977) in the northeastern district of Napak, which borders the Bokora Wildlife Reserve. As shown in Panel A of Figure 4, estimated prevalence was generally lowest in the southern districts bordering Lake Victoria and highest in the northern and northwestern regions of the country, mirroring programmatic experience within the NTLP.<sup>27</sup> The densely-populated districts of Kampala and Jinja experienced relatively high tuberculosis burden relative to the rest of the country, with estimated TB prevalence of 557 (547 to 568) and 537 (423 to 656) per 100,000, respectively.

Panel B of Figure 4 shows the relationship between unadjusted TB prevalence data from the 2014-2015 National TB Prevalence Survey (on the x axis) and prevalence estimates generated

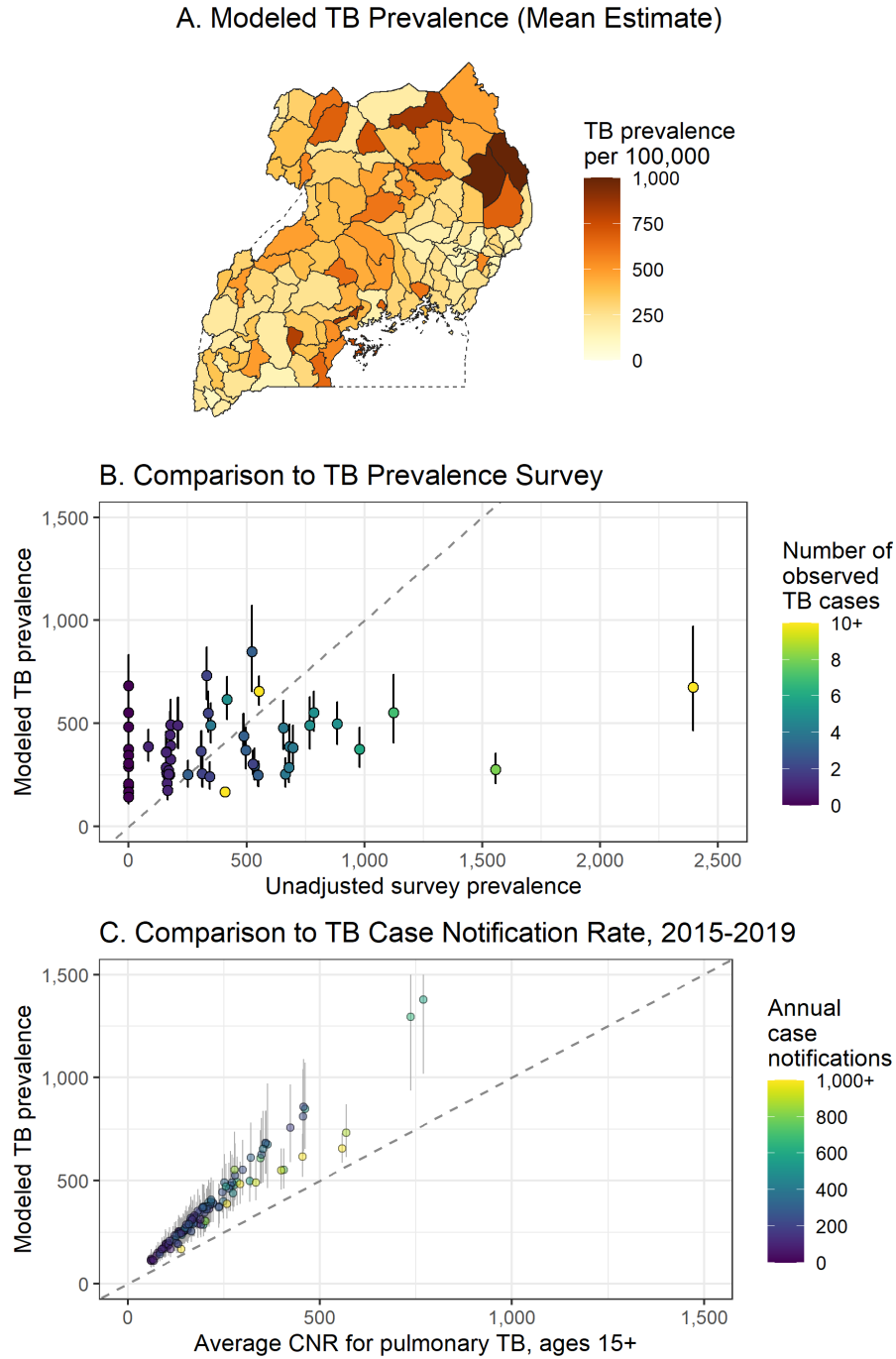


Figure 4: Results from the joint model compared to underlying data sources. Panel A: TB prevalence estimated from the joint data model. Panel B: scatter plot comparing modeled estimates, with uncertainty, against raw TB prevalence survey data in 56 districts. Panel C: scatter plot comparing modeled estimates by district, with uncertainty, against TB case notification rates among adults averaged across the years 2015-2019.

from the joint model (on the y axis). Given the relatively small numbers of observed cases in any individual prevalence survey cluster, the joint model generally smooths extreme observations from the prevalence survey, increasing estimates where only zero or one case was observed, and decreasing the extremely high district prevalence estimate of nearly 2,500 per 100,000 (1 in 40). These modifications are due to the combined effect of the smoothing spatial random effect in the prevalence surface as well as the effects of the notifications data.

Panel C of Figure 4 shows the relationship between the average TB case notification rate by district (on the x axis) and prevalence estimates (on the y axis). Noting the estimated TB incidence-to-prevalence ratio of approximately 1 in Uganda, the model estimates higher prevalence than indicated by case notifications in any district, although the ratio between the case notification rate and estimated TB prevalence varies by district. This difference reflects the variable incompleteness of case notification reporting by district across Uganda.

### 3.2 Completeness of case notifications increases over time

Figure 5, below, shows model estimates of case notification completeness by district for two years, 2017 and 2019. This figure demonstrates the general increasing trend in case notification completeness over time, matching the trend of increasing completeness estimated by the NTLP at the national level.<sup>10</sup> In 2017, estimated case notification completeness varied from 21% (17% to 25%) in the rural Otuke district, in the Northern region, to 99% (92% to 100%) in Kampala. In 2019, case notification completeness ranged from an estimated 26% (21% to 31%) in the southwestern Kiruhira district to, once again, 99% (93% to 100%) reporting completeness in Kampala. Population was associated with greater reporting completeness, as the urban districts of Kampala, Wakiso, Mukono, Gulu, and Jinja all had an estimated reporting completeness of 80% or greater in 2019.

The large majority of districts, 107 out of 128, experienced an increase in reporting completeness over the duration of the study, with an average increase of eight percentage points. Although the joint model estimated that the remaining 21 districts experienced a decline in case notifications, this decline was only significant in one district, Kiruhira, which dropped from 48% completeness (39% to 56%) in 2017 to 26% (21% to 31%) in 2019.

Based on the fitted model parameters, TB case notification rates were negatively associated with travel time to health facilities, and positively associated with nighttime lights, urbanicity, and time. Additionally, the spatially-structured effect in the completeness model identified district groupings with higher or lower completeness than would be expected from the covariate fixed effects alone: this included a group of high-completeness districts

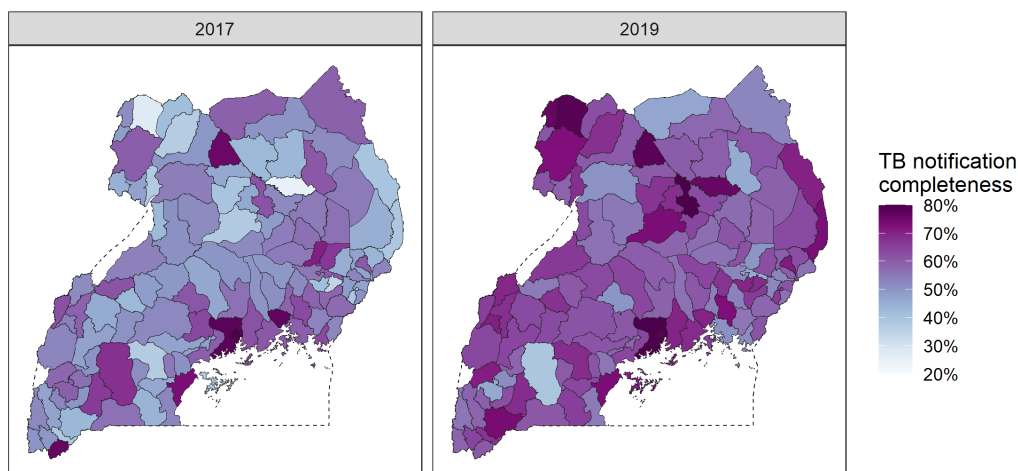


Figure 5: Mean estimated completeness for district-level TB case notifications in 2017 and 2019, as predicted by the joint data model. Completeness is estimated for pulmonary TB cases in people aged 15 and above.

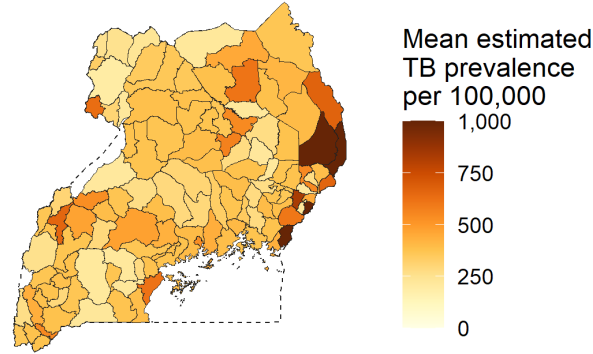
surrounding Kampala in the southern Central region, and a group of lower-completeness districts in the far east of the Northern region.

### 3.3 Effect of including notifications on estimates of TB burden

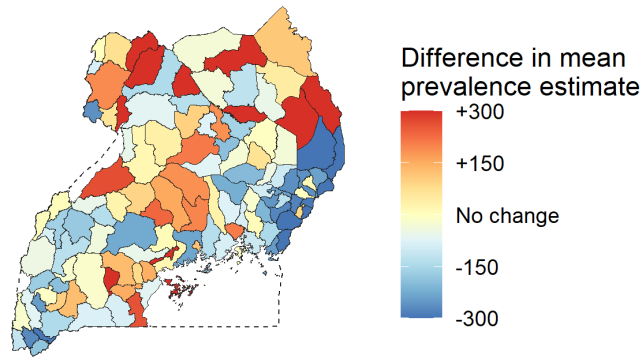
A second, prevalence-only model was run using only raw data from the 2014-15 national  
 275 TB prevalence survey along with the same suite of spatial covariates. This model iteration represents an attempt at district-level TB prevalence estimation without incorporating data from case notifications. Panel A of Figure 6 below shows the mean estimated TB prevalence based on this model. Compared to the raw data shown in Figure 2, the prevalence-only model captures and smooths extreme values within districts observed in the prevalence  
 280 survey, but strongly smooths towards the national mean in unobserved districts. Estimates of prevalence generated from this model are also much more uncertain than the joint model. In the joint model, the width of the 95% UI for prevalence varies from 13 to 869 deaths per 100,000, with an average spread of 209. In the prevalence-only model, the corresponding widths of the 95% UI for estimated prevalence ranges from 423 to 5,079, with an average  
 285 spread of 1,007.

Panels B and C in Figure 6 visualize the effect of indirectly incorporating case notifications into the modeled prevalence surface. As shown in Panel B, incorporating case notifications changes the mean estimates for TB prevalence across Ugandan districts, increasing mean  
 290 estimated prevalence in the northeast and decreasing estimated prevalence in the coastal south. However, given the wide uncertainty associated with the prevalence-only estimates,

### A. Modeled TB Prevalence, No Notifications



### B. Effect of Notifications on Mean Estimate



### C. Effect of Notifications on Uncertainty Intervals

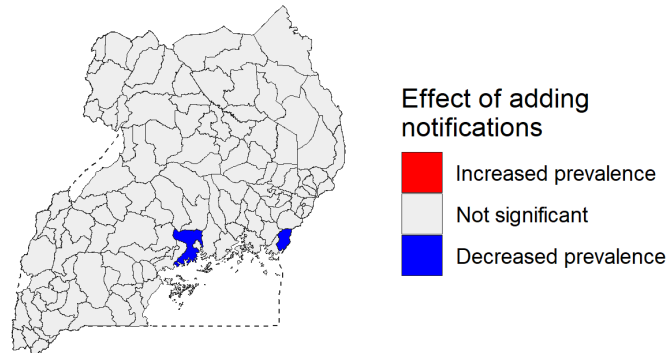


Figure 6: Effect of adding notifications to the joint estimation model. Panel A: Mean estimated TB prevalence based on a spatial model incorporating only data from the Uganda National TB Prevalence Survey. Panel B: Difference in mean estimated TB prevalence between the joint data model and a model that only incorporated prevalence data. Panel C: Effect of adding case notifications, taking uncertainty into account.

just two districts of 128 experience significant changes between the two model formulations. Both of these districts, Wakiso district in the Central region and Busia district in the Eastern region, were sampled in the National TB Prevalence Survey.

Figure 7 demonstrates how the addition of case notification data increases model precision, enabling the identification of low- and high-burden districts with greater confidence. The figure shows model predictions for districts with TB prevalence below 253 per 100,000, the national estimate reported by the 2014-2015 National TB Prevalence Survey, as well as districts with prevalence exceeding 506 per 100,000, or double the national average. In the context of this figure, low-confidence predictions indicate that the model's mean prevalence estimate passed one of these thresholds, while high-confidence predictions indicate that both bounds of the 95% uncertainty interval for prevalence in a given district have passed the threshold. Greater uncertainty in the model without notifications, shown on the right of this figure, precludes the observation of low-burden districts with high uncertainty. In addition, the survey-only model only predicts two districts to have high burden with high confidence: both of these were sampled in the National TB Prevalence Survey. The joint data model, shown on the right, makes more precise estimates due to the inclusion of corrected case notifications that supplement the prevalence survey data. With high confidence, the joint model predicts that 20 of the 128 districts fall below the lower prevalence threshold and 11 of 128 districts exceed the upper prevalence threshold.

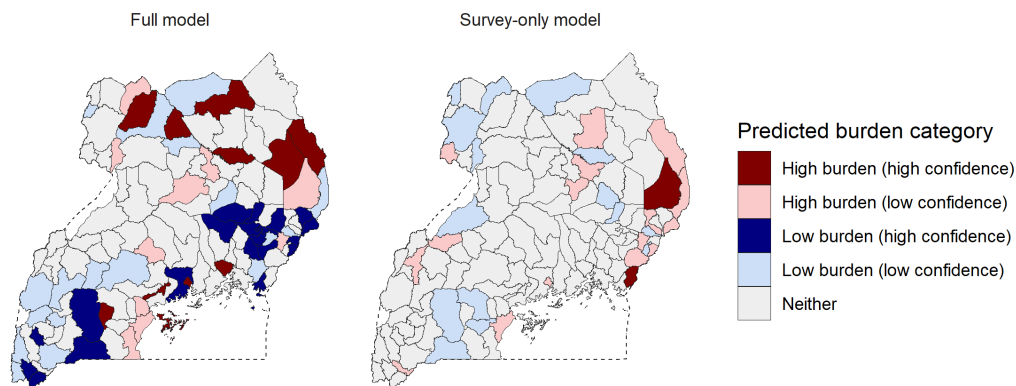


Figure 7: Comparison of estimates from the joint data model (left) and a model using only prevalence survey data (right) against a low prevalence threshold of 253/100,000 and a high threshold of 506/100,000. A high-confidence estimate indicates that the 95% uncertainty bounds uniformly pass a threshold, while a low-confidence estimate indicates that the model's mean estimate in a district passes a threshold but the 95% uncertainty bounds include both sides of that threshold.

## 4 Discussion

In this chapter, I demonstrate a novel framework for estimating TB prevalence by district across Uganda by synthesizing data from a TB prevalence survey and annual case notifications. A joint estimation model compensates for the weaknesses of each data source: it increases the precision of estimates available from either source alone while also correcting for systematic under-reporting in case notifications data. This approach corrects for weaknesses in previous spatial modeling frameworks for tuberculosis, which have primarily relied solely on just one of these two data types.<sup>17</sup>

TB prevalence surveys are generally not powered to predict subnational variation in TB prevalence.<sup>13</sup> Because of the larger sample sizes and time series associated with TB case notifications, this joint model generates more precise estimates of TB prevalence than a comparable model based on data from a single TB prevalence survey alone. This precision facilitates the identification of exemplars and hot spots for TB, allowing for more targeted and locally-appropriate policymaking. Two southern districts, Wakiso and Butaleja, have 95% uncertainty intervals that fall below the prevalence threshold of 200 per 100,000; on the other hand, the districts of Moroto and Napak have 95% UIs that fall entirely above the high prevalence threshold of 1,000 per 100,000. The stark differences between these predictions, and their high level of confidence, could be used as the basis of future studies to determine the elements of successful TB control in high-performing districts; they could also serve as the evidence base to develop new interventions in the sub-populations with the greatest burden, as recommended by the WHO Global Tuberculosis Report 2019.<sup>3,34</sup>

Previous research has demonstrated that unadjusted TB case notifications are an inappropriate proxy for TB incidence in high-burden settings due to confounding factors related to TB program activity and access to health care.<sup>12,17</sup> The space-time surface of notification completeness generated by the joint model offers a new tool for adjusting future case notifications and for targeting improvements in TB case-finding. Case notification completeness is a key indicator for tuberculosis control in Uganda, as described in NTLP's 2021-2025 strategic plan.<sup>27</sup> Measures of notification completeness have previously been estimated only through expensive audits in a small number of districts; while limited by the model assumptions, this joint modeling approach offers a supplementary method to tracking notification completeness that can be estimated nationwide each year. For example, the model identified 32 districts where case reporting was less than half complete with 95% certainty; four of these, Kiruhura, Butebo, Namayingo, and Mayuge, were predicted with 95% certainty to have estimated case finding of less than 40%. These estimates could be used as the basis for developing programs that strengthen the TB case reporting cascade in the southwestern

345 districts surrounding the Mbarara regional referral hospital, where reporting remains low;  
as well as for increasing access to health services in northeastern districts that are far from  
population centers and their regional referral hospitals. In the near future, this model could  
be applied to 2020 notifications to explore spatially-varying barriers to accessing TB care  
that may have arisen in the wake of the COVID-19 pandemic.<sup>35</sup>

#### 350 4.1 Extension to other contexts

While this chapter focused on the context of Uganda, similar methods could be applied  
in other high-burden countries with a TB prevalence survey. According to the Global  
Tuberculosis Report 2019, 13 countries in Africa and six countries in Asia had completed  
just one TB prevalence survey since 2007.<sup>3</sup> Many of these countries rely on notifications  
355 data to track annual change in TB prevalence: by combining these notifications with data  
from a single prevalence survey, these programs could learn a substantial amount about  
subnational variation in TB control nationwide. In Ethiopia, Pakistan, and Bangladesh,  
previous studies have already explored factors associated with subnational variation in case  
reporting completeness, and this information could be used to develop highly predictive  
360 spatial covariates for a latent completeness surface.<sup>12,18,36</sup>

Although this model makes use of the two most common data sources for TB surveillance  
available in high-burden countries, additional data types could easily be incorporated into  
the model framework to improve the precision of estimates. Countries may conduct case  
inventory studies to estimate under-reporting in particular districts.<sup>11</sup> Capture-recapture  
365 analyses of tuberculosis cases may also reveal spatial variation in TB notification complete-  
ness, subject to the limitations of this method that have been previously discussed in Chapter  
2.<sup>37,38</sup> Completeness estimates generated from these analyses can be directly incorporated  
to improve the precision of case reporting completeness, which is otherwise fit only indirectly  
in the model through the relationship between notifications and prevalence survey data.

370 The paradigm of intermittent nationwide prevalence surveys combined with incomplete case  
notifications data arises in the context of other infectious diseases. In sub-Saharan Africa,  
Malaria Indicator Surveys (MIS) and AIDS Indicator Surveys (AIS) are often considered to  
be the gold standard for surveillance, but small sample sizes and gaps in temporal coverage  
limit their use for routine surveillance. Previous geospatial studies have combined these  
375 surveys with routine surveillance data, using an estimated crosswalk between data types;<sup>39,40</sup>  
these models could be made more flexible by adapting the joint model approach described  
in this chapter.



## 4.2 Limitations

While this model estimates relatively precise estimates of TB prevalence and case notification completeness across districts in Uganda, it relies on an assumption that may not hold over long time periods. Because TB incidence in Uganda has fallen by less than 5% since 2015, as estimated by the NTLP, WHO, and IHME,<sup>3,8,10</sup> the joint model assumes that subnational variation in TB prevalence remains constant over the time period 2015-2019. Conversely, case notification reporting completeness was modeled as a space-time surface to capture documented increases in case notifications over the same time period.<sup>10,41</sup>. While such an assumption simplifies the comparisons between prevalence and incidence data sources, it cannot hold if the End TB goals are to be met. In settings where TB burden is decreasing, the underlying TB prevalence surface should vary in both space and time, with district-specific censuses directly grounding prevalence estimates in the years following a prevalence survey.

## 4.3 Conclusions

In this chapter, I demonstrate that subnational TB case notifications can provide meaningful information about the spatial distribution of TB burden within a country. This study extends the neonatal mortality model developed in Chapter 2 to account for health surveillance data that is expected to be incomplete, with incompleteness varying in both space and time. Whether the measured outcome is all-cause mortality or infectious disease prevalence, the principle of indirect data synthesis offers programmatically useful insights about variation in both underlying risk and data quality. In the next chapter, the same principle is applied to estimate the quality of cause-specific mortality data in Italy, a country where mortality data is typically considered unbiased and complete.

## 5 References

1. Onozaki, I. & Raviglione, M. Stopping tuberculosis in the 21st century: goals and strategies. *Respirology* 15, 32–43 (2010).
2. World Health Organization. The End TB Strategy. <https://www.who.int/teams/global-tuberculosis-programme/the-end-tb-strategy> (2015).
3. World Health Organization. Global Tuberculosis Report 2019. 1–283 (World Health Organization, 2019).

4. World Health Organization. Implementing the End TB Strategy. <https://www.who.int/teams/global-tuberculosis-programme/the-end-tb-strategy> (2015).
- 410 5. World Health Organization. Framework and standards for country health information systems (2E). 1–63 [http://www.who.int/healthmetrics/documents/hmn\\_framework200803.pdf](http://www.who.int/healthmetrics/documents/hmn_framework200803.pdf) (2012).
6. World Health Organization. Definitions and reporting framework for tuberculosis: 2013 revision. 1–40 <https://apps.who.int/iris/handle/10665/79199> (2020).
7. World Health Organization. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. vol. 35 1–92 <http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.35.021103.105711> (2010).  
415
8. Kyu, H. H. et al. Global, regional, and national burden of tuberculosis, 1990–2016: results from the Global Burden of Diseases, Injuries, and Risk Factors 2016 Study. *The Lancet Infectious Diseases* 18, 1329–1349 (2018).
9. Uganda National Tuberculosis and Leprosy Programme. National Tuberculosis and Leprosy Control Programme: revised national strategic plan 2015/16-2019/20. 1–52 file:///C:/Users/nyomb/Downloads/Revised-NTLP-Strategic-Plan-2015-2020\_\_25th-Jul2017-final.pdf (2017).
- 420 10. Uganda National Tuberculosis and Leprosy Programme. Uganda National TB and Leprosy Program: July 2019-June 2020 report. 1–72 (2020).
11. Glaziou, P. & Floyd, K. Latest developments in WHO estimates of TB disease burden. 1–13 (2018).
12. Rood, E. et al. A spatial analysis framework to monitor and accelerate progress towards SDG 3 to end TB in Bangladesh. *ISPRS International Journal of Geo-Information* 8, 1–11 (2019).  
425
13. Glaziou, P., Van Der Werf, M. J., Onozaki, I. & Dye, C. Tuberculosis prevalence surveys: Rationale and cost. *International Journal of Tuberculosis and Lung Disease* 12, 1003–1008 (2008).
14. Dehnavieh, R. et al. The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. *Health information management : journal of the Health Information Management Association of Australia* 48, 62–75 (2019).

- 430 15. Karamagi, E. et al. Improving TB case notification in northern Uganda: evidence of a quality improvement-guided active case finding intervention. *BMC Health Services Research* 18, 954 (2018).
16. Uplekar, M. et al. Mandatory tuberculosis case notification in high tuberculosis-incidence countries: Policy and practice. *European Respiratory Journal* 48, 1571–1581 (2016).
17. Shaweno, D. et al. Methods used in the spatial analysis of tuberculosis epidemiology: A systematic review. *BMC Medicine* 16, 1–18 (2018).
- 435 18. Van Gurp, M. et al. Finding gaps in TB notifications: spatial analysis of geographical patterns of TB notifications, associations with TB program efforts and social determinants of TB risk in Bangladesh, Nepal and Pakistan. *BMC Infectious Diseases* 20, 1–14 (2020).
19. Kirirabwa, N. S. et al. A four-year trend in pulmonary bacteriologically confirmed tuberculosis case detection in Kampala-Uganda. *BMC Pulmonary Medicine* 19, 1–8 (2019).
- 440 20. Uganda Ministry of Health. The Uganda National Tuberculosis Prevalence Survey, 2014-2015: survey report. 1–162 <https://www.health.go.ug/cause/the-uganda-national-tuberculosis-prevalence-survey-2014-2015-survey-report/> (2015).
21. Uganda National Tuberculosis and Leprosy Programme. Manual for management and control of tuberculosis and leprosy (3E). 1–162 (2017).
22. Uganda National Tuberculosis and Leprosy Programme. National Tuberculosis and Leprosy Division: July 2017-June 2018 report. 1–49 (2018).
- 445 23. Tatem, A. J. WorldPop, open data for spatial demography. *Scientific Data* 4, 2–5 (2017).
24. Freeman, J. & Hutchison, G. B. Prevalence, incidence, and duration. *American Journal of Epidemiology* 112, 707–723 (1980).
- 450 25. Tiemersma, E. W., Werf, M. J. van der, Borgdorff, M. W., Williams, B. G. & Nagelkerke, N. J. D. Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: A systematic review. *PLoS ONE* 6, (2011).

26. Buregyeya, E., Criel, B., Nuwaha, F. & Colebunders, R. Delays in diagnosis and treatment of pulmonary tuberculosis in Wakiso and Mukono districts, Uganda. *BMC Public Health* 14, 1–10 (2014).
27. Uganda National Tuberculosis and Leprosy Programme. National strategic plan for tuberculosis and leprosy control: 2020/21-2024/25. 1–106 (2020).
28. Weiss, D. J. et al. Global maps of travel time to healthcare facilities. *Nature Medicine* 26, 1835–1838 (2020).
29. Thomson, D. R. et al. Extending Data for Urban Health Decision-Making: a Menu of New and Potential Neighborhood-Level Health Determinants Datasets in LMICs. *Journal of Urban Health* 96, 514–536 (2019).
30. Riebler, A. et al. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research* 25, 1145–1165 (2016).
31. Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. M. TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software* 70, (2016).
32. Thorson, J. T. & Kristensen, K. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fisheries Research* 175, 66–74 (2016).
33. R Core Team. R: A Language and Environment for Statistical Computing. (2018).
34. Glaziou, P., Dodd, P. J., Zignol, M., Sismanidis, C. & Floyd, K. Methods used by WHO to estimate the global burden of TB disease. 1–36 [https://www.who.int/tb/publications/global\\_report/gtbr2018\\_online\\_technical\\_appendix\\_global\\_disease\\_burden\\_estimation.pdf](https://www.who.int/tb/publications/global_report/gtbr2018_online_technical_appendix_global_disease_burden_estimation.pdf) (2018).
35. Togun, T., Kampmann, B., Stoker, N. G. & Lipman, M. Anticipating the impact of the COVID-19 pandemic on TB patients and TB control programmes. *Annals of Clinical Microbiology and Antimicrobials* 19, 1–6 (2020).
36. Shaweno, D., Trauer, J. M., Denholm, J. T. & McBryde, E. S. A novel Bayesian geospatial method for estimating tuberculosis incidence reveals many missed TB cases in Ethiopia. *BMC Infectious Diseases* 17, 1–8 (2017).

- 475 37. Hook, E. B. & Regal, R. R. Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews* 17, 243–264 (1995).
38. Van Hest, R., Grant, A. & Abubakar, I. Quality assessment of capture-recapture studies in resource-limited countries. *Tropical Medicine and International Health* 16, 1019–1041 (2011).
39. Dwyer-Lindgren, L. et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature* 570, 189–193 (2019).
- 480 40. Lucas, T. C. D. et al. Model ensembles with different response variables for base and meta models: malaria disaggregation regression combining prevalence and incidence data. *BioRxiv* 2–7 (2019).
41. Uganda National Tuberculosis and Leprosy Programme. National Tuberculosis and Leprosy Programme: annual report for 2017. 1–72 (2017).