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 Therapy (DOTS, 1994-2005) and Stop TB (2006-2014) plans as the coordinating principles for addressing the tuberculosis epidemic worldwide.[1] 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.[2] These goals are consistent with Sustainable Development Goal 3.3, which commits United Nations member countries to end the TB epidemic by 2030.[3] 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.[4]

To achieve the End TB Strategy, the World Health Organization has developed tools for systematically building health system capacity[5–7] and has tracked progress and shared best practices through the release of annual Global Tuberculosis Reports.[3] 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 with HIV, the majority of high-burden countries are not on track to meet the End TB 2030 targets.[3,8] 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.[3] Among its many benefits, successful tuberculosis surveillance allows policymakers in high-burden countries to develop equitable interventions for those with TB burden. According 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 the influence of sex, age, socioeconomic status and differential access to health care on the risks and consequences of TB infection and disease.[3]

The recent history of TB control in Uganda, one of the 30 high-burden countries recognized by the End TB strategy, exemplifies both the possibilities for health system progress and the remaining challenges to TB elimination in low-resource settings. In Uganda, TB control is 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 strategy, calling for a 5% reduction in TB incidence as well as a 30% increase in case notifications over a four-year time period.[9] While the 2019-2020 NTLP report indicated impressive gains in TB treatment since 2015, the case notification target was barely missed (26% increase as opposed to 30%), while the estimated number of total TB cases increased slightly from 2017 to 2019.[10] Two of the principle challenges to universal TB detection and treatment identified in the 2020 report were (1) 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 capacity and adherence to TB treatment regimens.[10]

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 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.[11] TB prevalence and incidence 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.[12]

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.[13] 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.[3,13]

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.[6,14] 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.[15] Even clinically-diagnosed TB cases, particularly those diagnosed in private settings, may be missed in national statistics due to gaps in the reporting cascade.[11,13,16]

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.[17] 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 underreporting. 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.[18] 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

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 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.[9,15,19] As of 2019, Uganda was divided among 128 administrative districts grouped into four regions. Figure 1 shows Uganda's 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 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 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 survey conducted in 2014-2015, a time series of TB case notifications from 2015-2019, and predictive spatial covariates.

The number of administrative districts across Uganda increased from 122 to 128 between 2015 and 2019, with six districts being subdivided throughout this period. To simplify analysis and avoid unnecessary assumptions about the distribution of TB burden prior to district splits, I assessed burden across 122 analytical districts that correspond to districts of Uganda as of 2015. Cases and population counts recorded in split districts were grouped and summed to match the totals that would have been recorded under the 2015 administrative groupings.

I downloaded and extracted cluster-level results from the 2014-15 Uganda National TB Prevalence Survey, including place names associated with each cluster.[20] Using spatial information from Google Maps and OpenStreetMap, I associated each cluster with the GPS 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 setting.[21] Of the 122 analytical 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 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
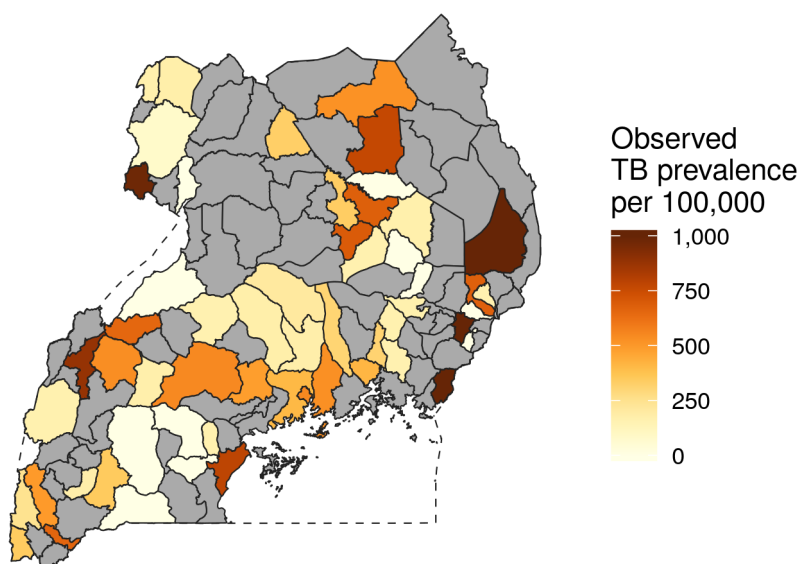
Figure 2: Unadjusted prevalence data from the 2014-2015 Uganda National TB Prevalence Survey, aggregated by district.

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 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 program region.[22] 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.[23] Figure 3 shows the case notification rate for pulmonary TB among adults aged 15 and above for the two years where district-level case notifications were reported.

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
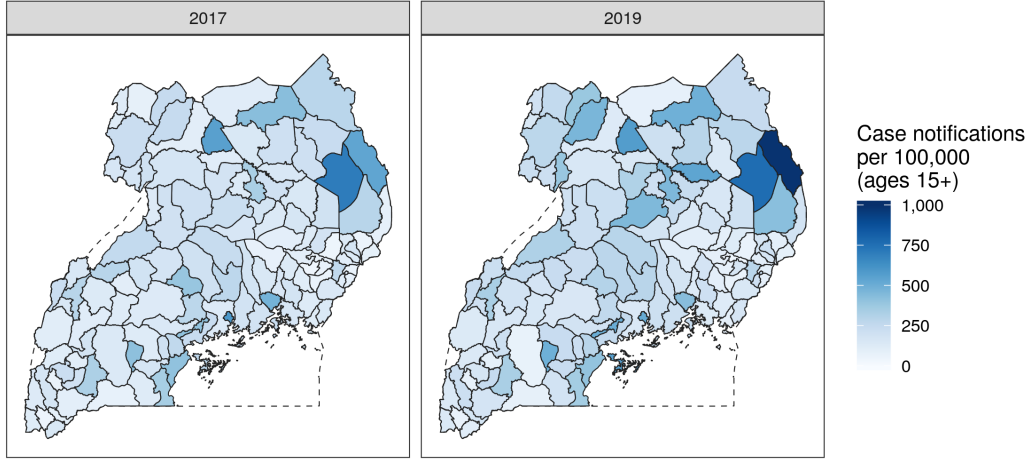
Figure 3: Data from the 2014-2015 Uganda National Tubserculosis Prevalence Survey, geo-positioned and aggregated to the district level. Out of 122 analysis districts, 56 contained underyling prevalence survey clusters.

of the Human Development Index as well as cattle per capita, used as a proxy to identify pastoral communities that may have different TB transmission dynamics.[24] Following Van Gurp et al.,[18] 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.[25,26] All gridded covariates were aggregated to the district level using population weights derived from the WorldPop project.[23]

Because the reporting completeness surface is fit based on relationships between case notifications and estimated prevalence, the three covariates used to fit reporting completeness were truncated to the range of values observed in the 56 districts with both recorded case notifications and TB prevalence survey data. This trimming process, conducted before model fitting and analysis, removed unobserved covariate values that might have yielded extreme model predictions for completeness in districts without prevalence survey data.

2.2   Estimated duration and incidence-prevalence ratio of TB

In populations where disease burden is relatively stable, the prevalence and incidence of a disease are related by the expected (mean) duration of the disease: *Prevalence = Incidence * E[Duration]*.[27] Estimates of TB duration by district are therefore necessary to crosswalk between data sources designed to measure TB prevalence and incidence.

The mean duration of tuberculosis in a population can vary based on factors such as HIV status, rates of case detection and treatment, and delays between the onset of active TB and the beginning of treatment. If an active TB case is immediately treated, the preferred

7

treatment strategy for drug-susceptible TB in HIV negative adults includes 26 weeks, or approximately 6.5 months, of chemotherapy.[28] At the upper range of duration, natural history studies suggest that the mean duration of untreated pulmonary TB among HIV negative adults is approximately three years.[29] The average duration of clinically active TB among adults with HIV is likely to vary based on antiretroviral therapy (ART), possible interactions between drugs used to treat TB and HIV, and higher mortality among individuals coinfected with HIV-TB.[29,30] For drug-resistant TB patients, The WHO recommends a much longer treatment regimen lasting at least 20 months.[31] A reasonable estimate for the expected duration of active TB in a population should combine the experiences of untreated and treated individuals, accounting for HIV coinfection and possible treatment delay, and should fall between the extremes of 6.5 months and three years.

Two research groups at the Institute for Health Metrics and Evalution (IHME) and the WHO produce annual estimates of global TB burden,[3,32] and these groups apply differing approaches to estimate the expected duration of TB by country. The IHME TB group estimates duration as a function of health care access and quality in each country, quantified by a Health Access and Quality (HAQ) Index.[33] Average duration is calculated as a linear interpolation between six months (for the highest possible HAQ score) and three years (for the lowest possible HAQ score). Using this approach, IHME estimates the average duration of active tuberculosis to be 2.15 (95% uncertainty interval [UI]: 2.11 to 2.18) years as of 2019.[32] Conversely, the WHO TB group estimates plausible duration ranges of TB disease by HIV and treatment status, then estimates duration at the national level as a population-weighted combination of these groupings.[34] These groupings, plausible duration distributions for individuals within each grouping, and expected duration across each grouping are shown in Table 1, below. The expectation of duration across each grouping is calculated as $E[Uniform(a,b)] = (a + b)/2$.

Table 1: Plausible duration of clinically active tuberculosis by HIV coinfection and treatment status, as published alongside the WHO Global Tuberculosis Report 2020.

Group	Case category	Distribution of duration (years)	E[duration]
a	Treated, HIV negative	Uniform(0.2-2)	1.100
b	Not treated, HIV negative	Uniform(1-4)	2.500
c	Treated, HIV positive	Uniform(0.01-1)	0.505
d	Not treated, HIV positive	Uniform(0.01-2)	1.005

In this chapter, I apply the WHO approach to estimate TB duration by Ugandan district based on expected rates of TB-HIV coinfection and TB treatment. The formula I apply to estimate expected duration in each district is a weighted average across the four groups identified in Table 1, adjusting for treatment delay:

$$E[D_{Total}] = (E[D_a] + L_a)Prop_a + E[D_b]Prop_b + (E[D_c] + L_c)Prop_c + E[D_d]Prop_d$$

In this equation, $E[D_{<a,b,c,d>}]$ represents the expected (average) duration within each WHO grouping specified in Table 1, while $Prop_{<a,b,c,d>}$ represents the proportion of individuals with active TB that fall within each grouping. The variables L_a and L_c denote the average

215 delay between the onset of active TB and the beginning of treatment for HIV negative (a) and HIV positive (c) adults who receive treatment. Based on a 2014 study of two districts in Uganda which estimated an average treatment delay of four weeks in which the median delay to treatment for active TB was found to be four weeks (0.077 years),[35] I set $L_a = L_c = 0.077$.

Given uncertainty around district-level variation HIV-TB coinfection and treatment rates

220 across Uganda, I made simplifying assumptions to estimate $Prop_{<a,b,c,d>}$ as a function of TB case notification reporting. The proportion of HIV-TB among all TB cases is likely to vary based on factors including HIV prevalence, which has been estimated to vary 10-fold across the districts of Uganda.[36] However, the relationship between HIV prevalence, TB prevalence, and HIV-TB coinfection at the subnational level is not well-understood. The

225 Global Burden of Disease Study 2019 estimated that 41.3% of all prevalent TB cases in Uganda were HIV-TB coinfections as of 2017,[32] and I applied this ratio uniformly across the country to estimate the relative prevalence of HIV positive and HIV negative TB cases. Proportions of treated versus untreated cases were estimated by multiplying the TB case detection rate by the treatment success rate, which was calculated to be 71% nationwide in

230 the Uganda NTLP annual report.[22] Finally, treatment rates were assumed to be constant for TB cases among individuals with and without HIV, so that $Prop_a = Prop_{HIV} * Prop_{Treat}$, and so on. These assumptions allow for the estimation of the proportion of all prevalent TB cases that fall into each WHO category based on the case detection rate, denoted as π:

$$Prop_a = Prop_{Treat} * Prop_{HIV} = .71\pi * .413 = .29\pi$$
$$Prop_b = (1 - Prop_{Treat}) * Prop_{HIV} = (1 - .71\pi) * .413 = .413 - .29\pi$$
$$Prop_c = Prop_{Treat} * (1 - Prop_{HIV}) = .71\pi * .587 = .42\pi$$
$$Prop_d = (1 - Prop_{Treat}) * (1 - Prop_{HIV}) = (1 - .71\pi) * .587 = .587 - .42\pi$$
$$Prop_a + Prop_b + Prop_c + Prop_d = 1$$

235 Using these simplifying assumptions, the average duration of TB cases in a district can be estimated as a function of the case reporting completeness π:

$$D(\pi) = E[D_{Total}]$$

$$= (E[D_a] + L_a)Prop_a + E[D_b]Prop_b + (E[D_c] + L_c)Prop_c + E[D_d]Prop_d$$

$$= (1.1 + .077)(.29\pi) + 2.5(.413 - .29\pi) + (.505 + .077)(.42\pi) + 1.005(.587 - .42\pi)$$

$$= 1.62 - .81\pi$$

I deploy this formula below to estimate the average duration of TB cases by district, and therefore the relationship between TB incidence and prevalence, as a function of TB case

250 notification completeness. I conducted a sensitivity analysis on the effect of varying treatment delays, HIV-TB proportions, and differential treatment between HIV positive and HIV negative TB cases on expected duration, the results of which are presented in Appendix C.

2.3 Small area model incorporating prevalence data and incomplete notifications

245 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, π, is represented as a space-time surface that varies according to underlying covariates as well as a spatially-structured

250 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). Prevalence varies according to an intercept, α, 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

255 fixed effects. The latent spatial surface is parameterized using the BYM2 spatial model formulation described by Riebler and colleagues.[37] The superscript (p) differentiates these terms from a similar spatial field used to estimate reporting completeness (π).

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

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

In this formulation, reporting completeness π is a logit-linear surface that is indexed by

260 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^π that is parameterized in the same manner as Z_s^p, according to the adjacency structure of Ugandan districts. Reporting 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:

$$Y_i^{prev} = Poisson(N_i^{prev} * p_{s(i)})$$

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:

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

The observed number of case notifications are estimated to be centered around the true 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, $1/D(\pi_{s(i),t(i)})$, which is calculated based on the duration formula presented in the previous section. 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.[38,39] 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.[38,40]

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 completeness varied from less than 40% to nearly 80% as of 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 122 analysis districts of Uganda, estimated pulmonary TB prevalence in adults ranged from 161 per 100,000 people (95% uncertainty interval: 132 to 193) in the district of Butaleja in the southeast of Uganda, to a maximum of 2,300 per 100,000 (uncertainty 1645 to 3186) in Moroto district which borders Kenya in the northeast. As shown in Panel A of Figure 4, estimated prevalence was generally lowest in Uganda's Eastern region and highest in the far northeastern districts of the country. The densely-populated districts of Kampala and Arua experienced moderately high tuberculosis burden relative to the rest of the country, with estimated TB prevalence of 631 (519 to 765) and 550 (477 to 636) 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 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 one 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 every district, although the ratio between the case notification rate and estimated TB prevalence varies by district. This difference reflects the combined effects of case notification incompleteness and estimated average TB duration that fell above 1 in most districts, both of which increased estimated prevalence relative to recorded case notifications.

12

A. Modeled TB Prevalence (Mean Estimate)

B. Comparison to TB Prevalence Survey

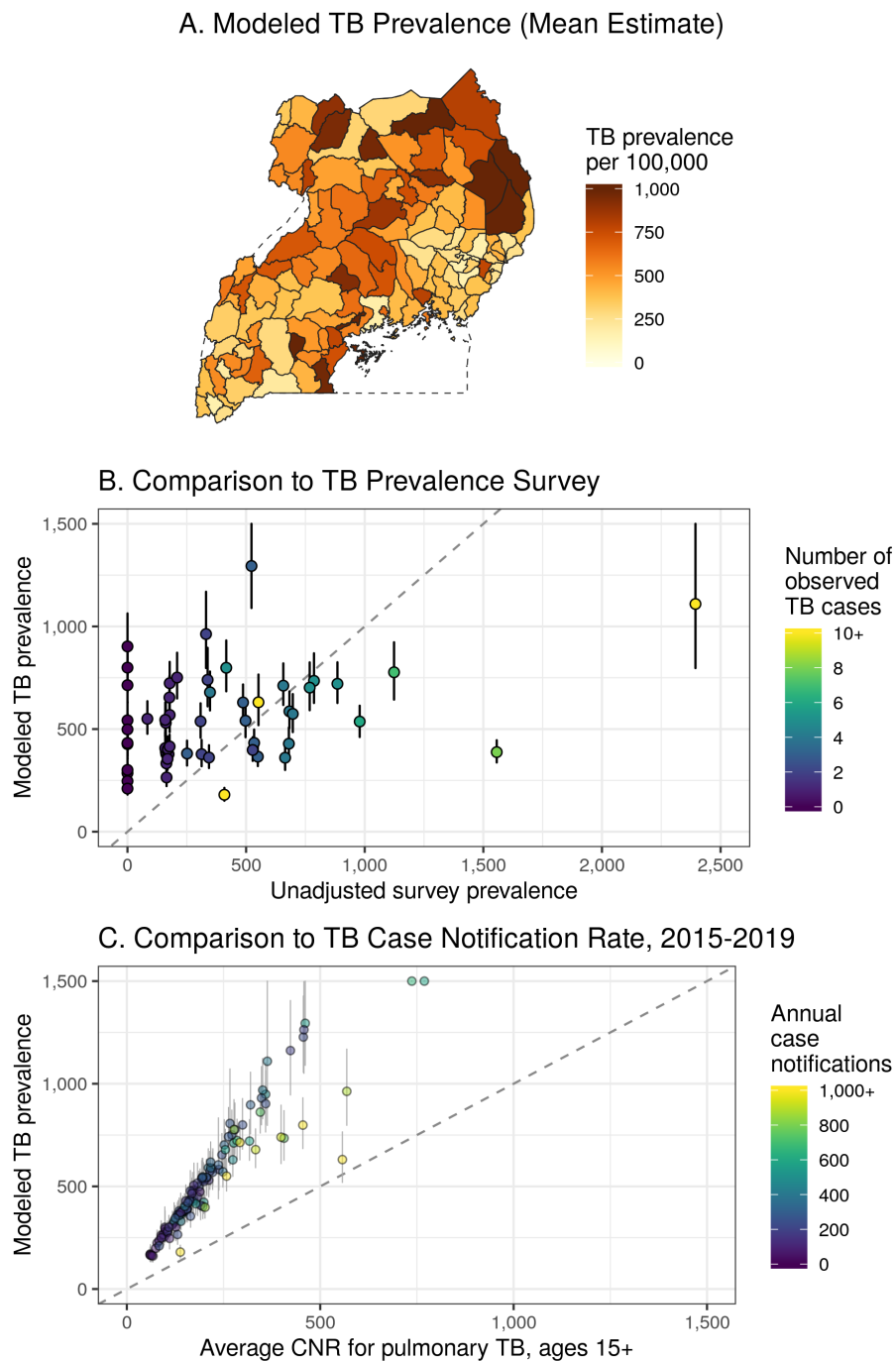## C. Comparison to TB Case Notification Rate, 2015-2019

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.

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.[10] At the national level, case notification completeness increased from 49% (45% to 53%) in 2017 to 55% (51% to 59%) in 2019. In 2017, estimated case notification completeness across districts varied from 27% (23% to 31%) in the rural Otuke district, in the Northern region, to 86% (74% to 94%) in Kampala. In 2019, case notification completeness ranged from an estimated 36% (32% to 41%) in the southwestern Kiruhira district to 80% (68% to 88%) reporting completeness in Kampala. Population was associated with greater reporting completeness, and the populous districts of Kampala, Wakiso, and Gulu all had an estimated reporting completeness of 65% or greater in 2019.
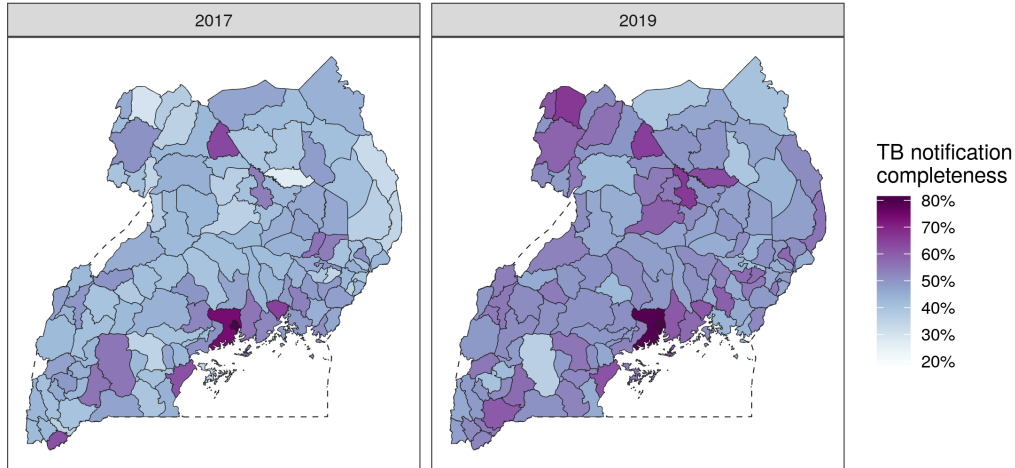

Figure 5: Mean estimated completeness for district-level TB case notifications in 2017 and 2019, as predicted by the joint data model. Completeness is estimated for pulmonary TB cases in people aged 15 and above.

The change in estimated reporting completeness from 2017 to 2019 in Kampala is not significant within the bounds of uncertainty. On the other hand, many districts with the lowest reporting completeness experienced significant increases between 2017 and 2019. As of 2017, 21 of 122 districts were highly likely to have reporting completeness of less than 45%, meaning that the upper bounds of the 95% uncertainty interval fell below the 45% mark; by 2019, only Kiruhira district's reporting completeness fell below the 45% threshold with high certainty.

The large majority of districts, 100 out of 122, experienced an increase in reporting completeness over the duration of the study, with an average increase of six percentage points.

Although the joint model estimated that the remaining 22 districts experienced a decline in case notifications, that decline was not significant within the bounds of uncertainty for any district.

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

Using the simplified formula for TB duration presented above, $D(\pi) = 1.62 - .81\pi$, the average duration of TB cases was estimated as a function of case notification completeness. Figure 6 shows the estimated duration of pulmonary TB among adults by district in 2017 and 2019. At the national level, the average duration of pulmonary TB among adults was estimated to have fallen from 1.22 (1.19-1.25) years to 1.17 (1.14-1.21) years between 2017 and 2019 as a result of greater case reporting and subsequent treatment. As of 2019, the average duration of pulmonary TB was estimated to range from 0.98 (.90-1.07) years in Kampala to 1.33 (1.29-1.36) in Kiruhira, a difference of over four months. Given that the relationship between TB incidence and prevalence is defined based on the average duration of disease, $Prevalence = Incidence * E[Duration]$, space-time variation in TB duration mediated the relationship between prevalence estimates from the 2014-2015 National TB Prevalence Surey and incidence estimates from completeness-adjusted case notifications.
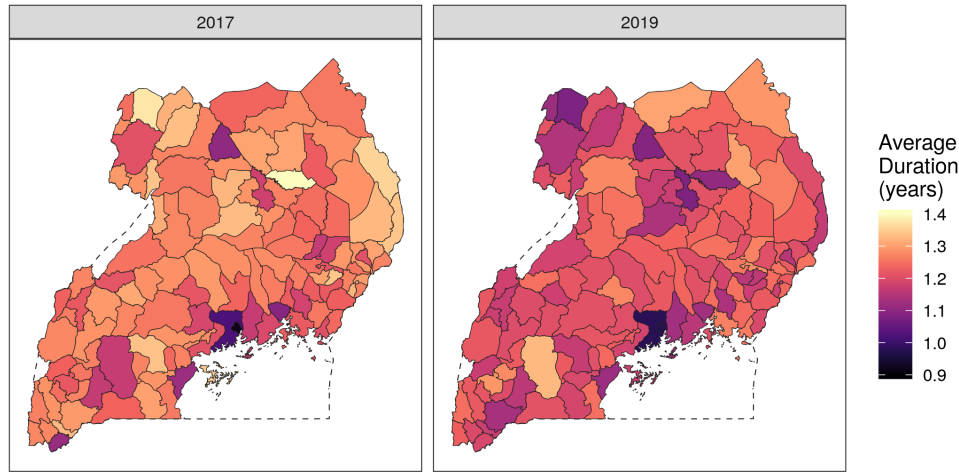

Figure 6: Average duration of active pulmonary TB among adults by Ugandan district, estimated as a function of case notification completeness.
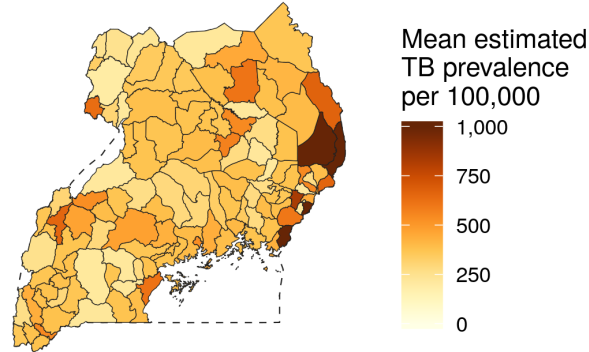
3.3 Effect of including notifications on estimates of TB burden

A second, prevalence-only model was run using only raw data from the 2014-15 national TB prevalence survey along with the same suite of spatial covariates. This model iteration
365 represents an attempt at district-level TB prevalence estimation without incorporating data from case notifications. Panel A of Figure 7 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 survey, but strongly smooths towards the national mean in unobserved districts. Estimates
370 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 61 to 1,541 deaths per 100,000, with an average uncertainty interval width of 197. In the prevalence-only model, the corresponding widths of the 95% UI for estimated prevalence ranges from 423 to 5,079, with an average spread of 1,007.

375 Panels B and C in Figure 7 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 estimated prevalence in the northeast and the districts surrounding Hoima while decreasing estimated prevalence in the coastal south and southeast. However, given the wide uncer-
380 tainty associated with the prevalence-only estimates, just eight districts of 122 experience significant changes between the two model formulations. Of these eight districts, seven were sampled in the National TB Prevalence Survey.

Figure 8 demonstrates how the addition of case notification data increases model precision, enabling the identification of low- and high-burden districts with greater confidence. The
385 figure shows model predictions for districts where TB prevalence falls below 253 per 100,000, the national estimate for Uganda reported by the 2014-2015 National TB Prevalence Survey, as well as districts exceeding a high prevalence threshold of 759 per 100,000, or triple that national estimate. In the context of this figure, low-confidence predictions indicate that the model's mean estimated prevalence passed a given threshold, while high-confidence predic-
390 tions 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 one district to have high burden with high confidence: this district, Nakapiripirit, was the location of the survey cluster
395 with the highest recorded prevalence from the National TB Prevalence Survey. The joint data model, shown on the right, makes more precise estimates due to the inclusion of cor-

A. Modeled TB Prevalence, No Notifications

B. Effect of Notifications on Mean Estimate
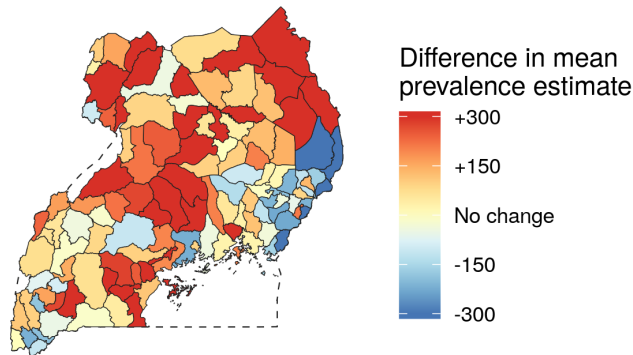
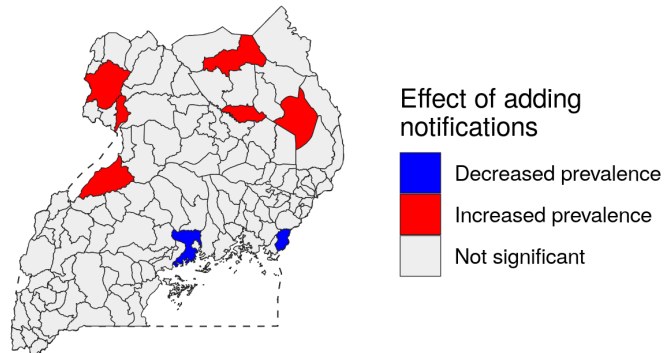## C. Effect of Notifications on Uncertainty Intervals

Figure 7: 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.

rected case notifications that supplement the prevalence survey data. With high confidence, the joint model predicts that 8 of the 122 districts fall below the lower prevalence threshold and 13 of 122 districts exceed the upper prevalence threshold.
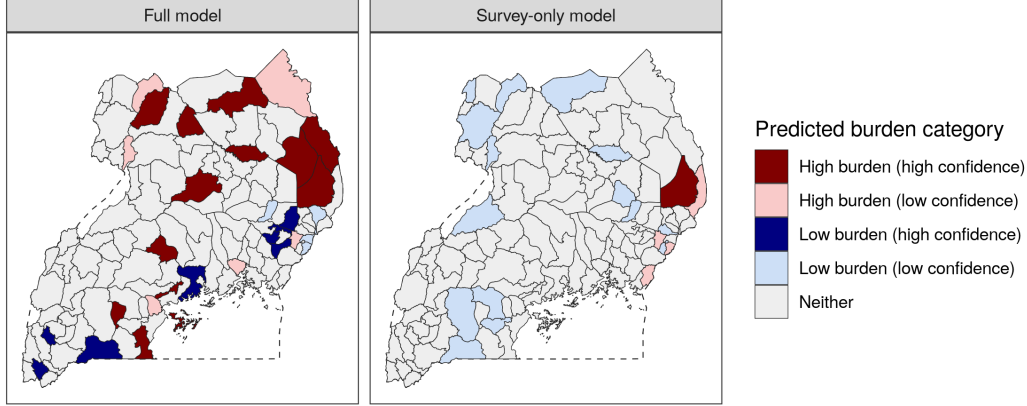

Figure 8: Comparison of estimates from the joint data model (left) and a model using only prevalence survey data (right) against a low prevalence threshold of 253/100,000 and a high threshold of 506/100,000. A high-confidence estimate indicates that the 95% uncertainty bounds uniformly pass a threshold, 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.[17]

TB prevalence surveys are generally not powered to predict subnational variation in TB prevalence.[13] 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. The southeastern district of Butaleja, with an estimated TB prevalence of 161 (132 to 193), is the only district that falls below the prevalence

threshold of 200 per 100,000 with high confidence; on the other hand, the four districts of Moroto and Napak have 95% UIs that fall entirely above the high prevalence threshold of 1,500 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 End TB strategy.[3,41]

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.[12,17] 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.[24] Measures of notification completeness have previously been estimated only through expensive audits in a small number of districts; this joint modeling approach, while limited by assumptions, offers a supplementary method to tracking notification completeness that can be estimated nationwide each year. For example, the model identified 10 districts where case reporting was less than half complete with 95% certainty as of 2019. These estimates could be used as the basis for developing programs that strengthen the TB case reporting cascade in the southwestern 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.[42]

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 2020, of the 30 high TB burden countries classified by the WHO, 16 countries had completed just one TB prevalence survey since 2011.[3] Many high burden countries rely on aggregated case notifications to track annual change in TB prevalence: by combining 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 spatial covariates for a latent completeness surface.[12,18,43]

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.[11] Capture-recapture analyses of tuberculosis cases may also reveal spatial variation in TB notification completeness, subject to the limitations of this method that have been previously discussed in Chapter 2.[44,45] 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.

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 spatial modeling studies have combined these surveys with routine surveillance data, using an estimated crosswalk between data types;[36,46] these models could be made more flexible by adapting the synthesis approach described in this chapter.

4.2   Limitations

While this model generates 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 WHO and IHME modeling studies agreee that nationwide TB incidence in Uganda fell by less than 5% between 2015 and 2019,[32,47] the model presented in this chapter assumes that subnational variation in TB prevalence also remains constant over this time period. Conversely, case notification reporting completeness was modeled as a space-time surface to capture documented increases in case notifications over the same time period.[10,48]. 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.

To relate data sources measuring tuberculosis incidence and prevalence, I roughly approximated variation in TB duration by district as a function of case detection rates. This

approximation required strong assumptions about rates of HIV-TB coinfection, drug resistance, and TB treatment and their effects on duration by district in Uganda. While I conducted a sensitivity analysis showing that the effect of relaxing any individual assumption would have a relatively minor effect on estimated duration, further research is needed to better understand subnational variation in HIV-TB and TB treatment across Uganda. The IHME and WHO use differing approaches to calculate disease duration;[32,34] while this chapter extends the WHO approach due to the greater availability of relevant inputs at the district level, additional research is needed to compare and validate these approaches in a variety of settings.

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 widely considered to be 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 2020. 1–208 https://www.who.int/publications/i/item/9789240013131 (2020).

4.   World Health Organization. Implementing the End TB Strategy. https://www.who.int/teams/global-tuberculosis-programme/the-end-tb-strategy (2015).

5.   World Health Organization. Framework and standards for country health information systems (2E). 1–63 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).

515   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).

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-Stragetic-Plan-2015-2020_25th-Jul2017-final.pdf (2017).

520

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).

525   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).

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).

530

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).

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).

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).

23. Tatem, A. J. WorldPop, open data for spatial demography. Scientific Data 4, 2–5 (2017).

24. Uganda National Tuberculosis and Leprosy Programme. National strategic plan for tuberculosis and leprosy control: 2020/21-2024/25. 1–106 (2020).

25. Weiss, D. J. et al. Global maps of travel time to healthcare facilities. Nature Medicine 26, 1835–1838 (2020).

26. 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).

27. Freeman, J. & Hutchison, G. B. Prevalence, incidence, and duration. American Journal of Epidemiology 112, 707–723 (1980).

28. Nahid, P. et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. Clinical Infectious Diseases 63, e147–e195 (2016).

29. 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).

30. Payne, B. & Bellamy, R. Managing tuberculosis in people with HIV. BMJ Clinical Evidence (2007).

31. Falzon, D. et al. WHO guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update. European Respiratory Journal 38, 516–528 (2011).

32. Ledesma, J. R. et al. Global, regional, and national sex differences in the global burden of tuberculosis by HIV status, 1990–2019: results from the Global Burden of Disease Study 2019. The Lancet Infectious Diseases 3099, 1–20 (2021).

33. Fullman, N. et al. Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. The Lancet 391, 2236–2271 (2018).

34. Glaziou, P., Dodd, P. J., Dean, A. S. & Floyd, K. Methods used by WHO to estimate the global burden of TB disease: global tuberculosis report 2020. https://cdn.who.int/media/docs/default-source/hq-tuberculosis/global-tuberculosis-report-2020/tb2020_technical_appendix_20201014.pdf (2020).

35. 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).

36. Dwyer-Lindgren, L. et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nature 570, 189–193 (2019).

37. 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).

38. Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. M. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software 70, (2016).

39. 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).

40. R Core Team. R: A Language and Environment for Statistical Computing. (2018).

41. Glaziou, P., Dodd, P. J., Zignol, M., Sismanidis, C. & Floyd, K. Methods used by WHO to estimate the global burden of TB disease: global tuberculosis report 2018. 1–36 https://www.who.int/tb/publications/global_report/gtbr2018_online_techn ical_appendix_global_disease_burden_estimation.pdf (2018).

42. 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).

43. 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).

44. Hook, E. B. & Regal, R. R. Capture-recapture methods in epidemiology: Methods and limitations. Epidemiologic Reviews 17, 243–264 (1995).

45. 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).

46. 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).

47. World Health Organization. Global Tuberculosis Report 2019. 1–283 (World Health Organization, 2019).

48. Uganda National Tuberculosis and Leprosy Programme. National Tuberculosis and Leprosy Programme: annual report for 2017. 1–72 (2017).