

Supplementary Materials

Contents

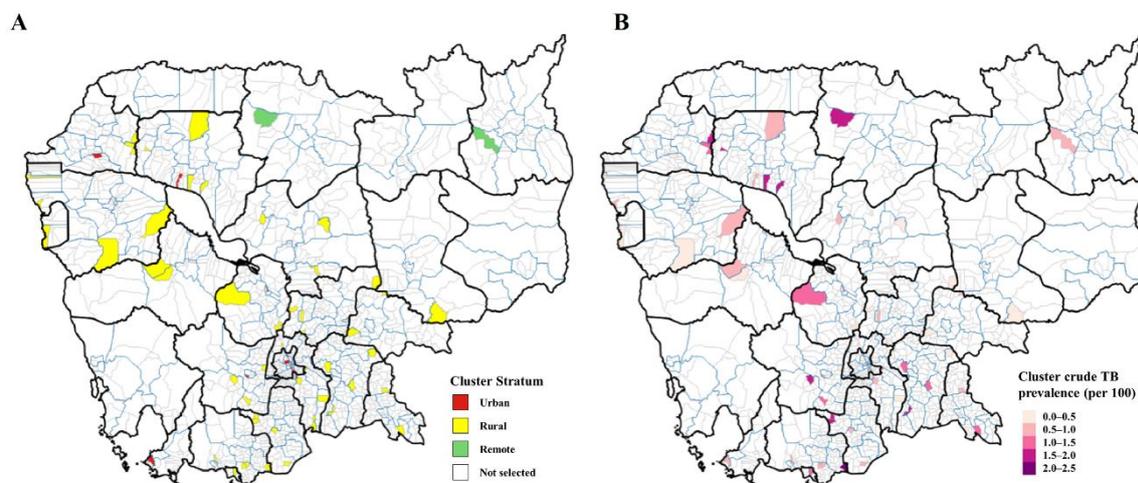
I. Active TB point prevalence model (TB-NTP model)	1
2011 National Tuberculosis Prevalence Survey Data	1
TB-NTP model	2
II. Demographic Epidemiological Model of Cambodia (DEMOKH).....	3
Spatial projections	3
Cambodian population spatial data.....	3
Temporal projections	4
1. Data for the fertility rate model.....	4
2. Data for the mortality rate model	5
3. Urbanisation model	6
III. Future TB projections	7
IV. Access to TB healthcare services	8
Government healthcare facilities	8
List of abbreviations	10
References	10

I. Active TB point prevalence model (TB-NTP model)

2011 National Tuberculosis Prevalence Survey Data

A total of 62 survey sites, or spatial clusters, were randomly selected across Cambodia with population proportionate multistage cluster sampling for the Second National Tuberculosis Prevalence (NTP) survey 2011 from December 2010 to September 2011 (**Supplementary Figure 1**). In total, 37,417 (92.6%) out of 40,423 eligible participants were selected and they met the following criteria:

1. adults (at least 15 years old)
2. resided in one of the 62 survey sites sampled for the Second National Tuberculosis Prevalence (NTP) Survey 2011 for 2 weeks or longer at the time of survey



Supplementary Figure 1: National Tuberculosis Prevalence Survey 2011 survey sites and crude prevalence of bacteriologically positive tuberculosis by site. 62 sites (spatial clusters) were randomly selected by population proportionate multistage cluster sampling for the Second National Tuberculosis Prevalence Survey 2011 from December 2010 to September 2011. As of 2016, Cambodia is divided into 25

provinces (demarcated by black borders) where each province contains several districts (blue borders) and is then further divided into communes (grey borders). The cluster-specific crude prevalence (per 100 population) of bacteriologically positive tuberculosis among individuals aged ≥ 15 years was computed by taking the ratio of the number of bacteriologically positive TB cases and the number of interviewed individuals in the cluster. The crude prevalence ranged from 0 (Thmei in Ta Khamau commune and Phum 2 in Tuol Tumpung Ti Pir commune) to 2.24 (Tnoat Chong Srang commune).

TB-NTP model

Let N be the survey size and N_{clus} the number of clusters sampled during the Second NTP 2011. Let p_i be the probability of active TB infection of an individual i , where $Y_i = 1$ if i has active TB and 0 otherwise. We model active TB point prevalence as $Y_i \sim \text{Bernoulli}(p_i)$ where

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta'x_i + \gamma_{c_i}.$$

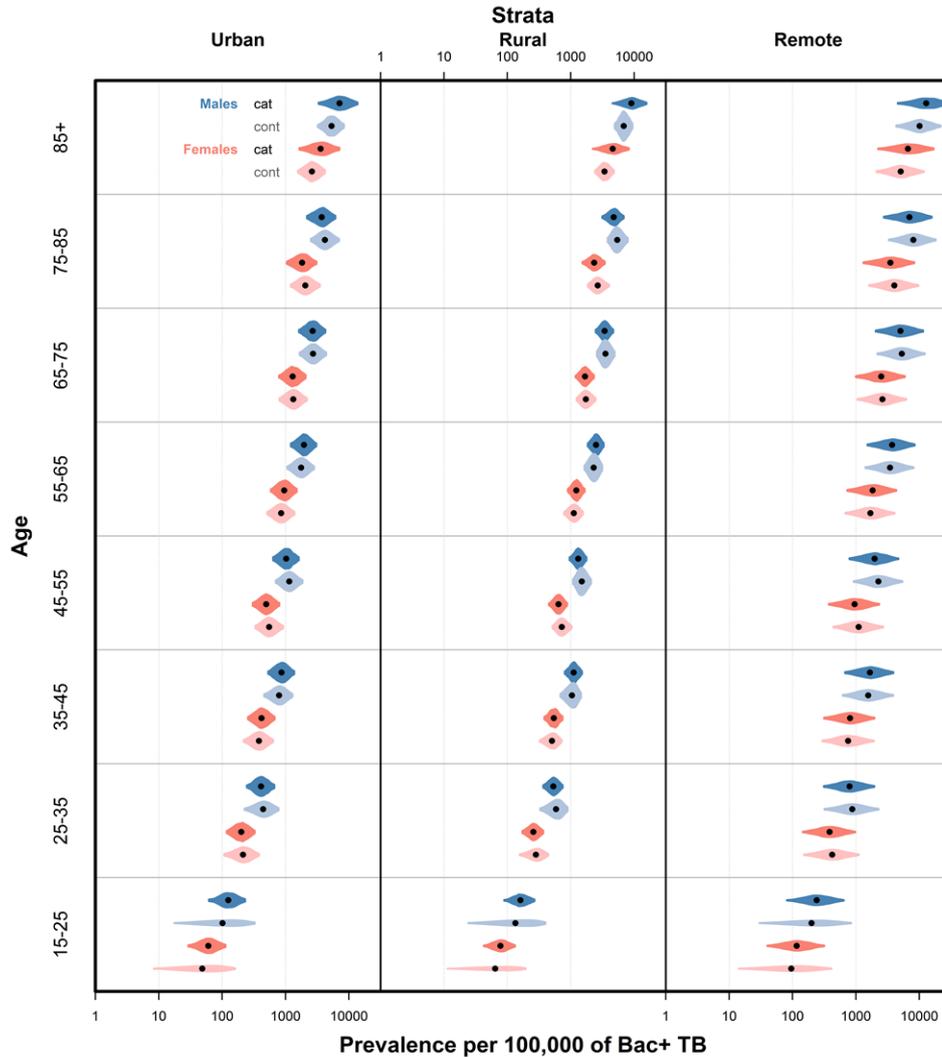
Here, x_i is a vector of individual-level risk factors of active TB infection—including age (in categories or continuous, details as follows), sex, and urbanisation level, while γ_{c_i} is a spatial random effect shared by all individuals in the commune c_i to which i belongs. We wish to determine the parameters β and γ_c . The latter is an effect of clustering of the survey sites. Weakly informative priors were assumed for all the parameters of in model, namely: $\beta \sim \text{Normal}(0, 100^2)$ for each component k in the individual risk factors vector, $\gamma_c \sim \text{Normal}(0, \sigma^2)$ for commune c , and $\sigma \sim \text{Gamma}(1, 0.01)$. Because very few sampled communes border each other, there was insufficient information to fit a model with spatial correlations in the random effects, which were therefore modelled to be conditionally independent of each other.

Two variants of the model were considered: *age-categorical* and *age-continuous*. In the *age-categorical* model, age measured in years was categorised in to eight groups: 25 and younger, (25, 35], (35, 45], (45, 55], (55, 65], (65, 75], (75, 85], older than 85, before being treated as binary predictors with 25 and younger as the reference group. In the *age-continuous* model, the parameters θ_1 , θ_2 , and θ_3 were introduced where the term corresponding to age was estimated as follows:

$$[I(15 \leq a \leq 25) \cdot (a - 15) \theta_1] + [I(25 \leq a < 37.5) \cdot (10 \theta_1 + (a - 25) \theta_2)] + [I(37.5 \leq a < 50) \cdot (10 \theta_1 + 12.5 \theta_2 + (a - 37.5) \theta_3)] + [I(a \geq 50) \cdot (10 \theta_1 + 12.5 \theta_2 + 12.5 \theta_3 + (a - 50)\theta_4)].$$

The posterior distribution of the parameters was estimated via Markov chain Monte Carlo simulation. The inference was implemented using Just Another Gibbs Sampler (JAGS) within the R statistical environment using 50 000 iterations for each of the models. The model was specified in the JAGS dialect of the BUGS language and later constructed in R via the `rjags` package. With the model, data and initial values properly specified, we initialised the model using the `jags.model` function found in `rjags`.

Both models yield similar results as depicted in **Supplementary Figure 2**, only results of the *age-categorical* model are presented in the main paper (**Figure 2**).



Supplementary Figure 2: Differences in prevalence of bacteriologically positive TB across exposures as estimated by the *age-categorical* and *age-continuous* model. Modelled bacteriologically positive TB prevalence per 100,000 population estimates are represented in the log base 10 scale and are based on individual-level data from the 2011 NTP survey. The estimates from the *age-continuous* model are presented below the *age-categorical* model. The prevalence of active TB varies across age and sex of the individual, and stratum level of his/her living environment. The interpretation of violin plots is like box plots, they display the probability density of the prevalence estimates at different values. Points are posterior median prevalence estimates, curves are posterior distributions of the parameters truncated to within 95% credible interval (CI).

II. Demographic Epidemiological Model of Cambodia (DEMOKH)

The Demographic Epidemiological Model of Cambodia (DEMOKH), an individual-level model that represents the population of Cambodia over space and time was used to project the future prevalence of active TB in Cambodia to the year 2030. DEMOKH was constructed over space and time through the following submodels.

Spatial projections

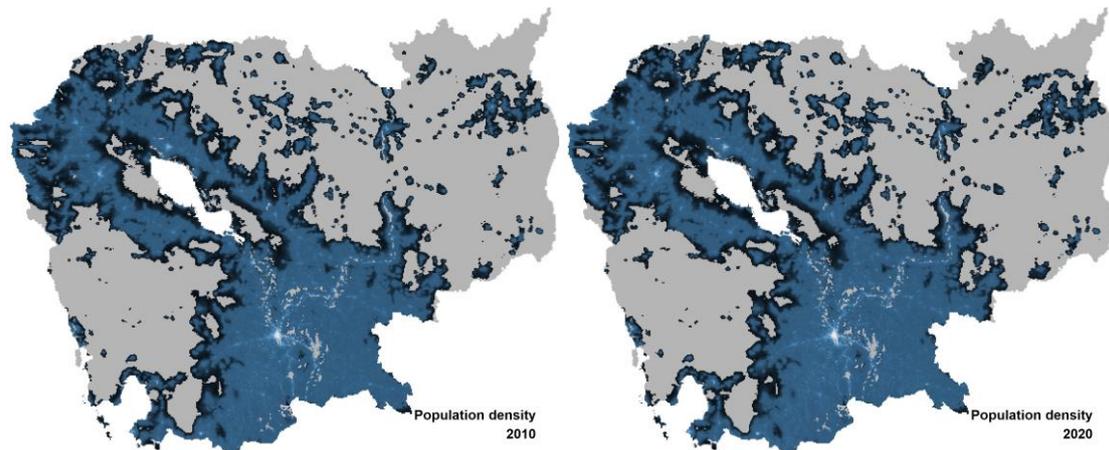
Cambodian population spatial data

Gaughan and colleagues¹ developed high resolution estimates/projections of the spatial distribution of the population using the United Nations' adjusted population distribution of Cambodia for the years 2010, 2015, and 2020, which were extracted from the WorldPop Project². Data were downloaded as raster files and processed in QGIS,³ ArcMap,⁴ and R⁵ to be partitioned into one-kilometre (km) grid squares.

We performed the following steps:

1. We resampled (or re-projected) the original raster files onto the EPSG 3148 coordinate reference system (CRS). This was done in QGIS.
2. We then rasterized the province shape file (in ArcMap) and resampled it to 1km by 1km in CRS EPSG 3148 Indian 1960 / UTM zone 48N. This led to each 1km by 1km grid square having a province label attached to it.
3. We resampled the population raster files to the same extent as the rasterized province file created in Step 2. All population and province raster files had 275 613 elements over a 481 by 573 matrix.
4. We then added an urban layer to the population data in R (via customised R-script). Based on the proportion of the population living in urban areas from census data in 2008 at the province-level, we labelled each 1km by 1km grid cell as urban or rural. To do so, first, we ranked the 1km by 1km grid by population size, and we took the cumulative sum of the population (by province) from most populous to least. Then, if the cumulative sum is smaller than the urban population size for the province, the classified 1km by 1km grid square is urban. Less populous cells were then coded as rural. Explicitly, we did the following within each province:
 - 1) We extracted the density d_i of cell i .
 - 2) We let the total population size of the province be N , urban population size of the province be N_u , and rural population size of the province be N_r .
 - 3) We calculated the population of cell i as $n_i \propto d_i$ such that $\sum_i n_i = N$.
 - 4) We ranked cells by n_i from largest to smallest.
 - 5) We set $\theta = \min_k n_k$ such that $\sum_{i=1}^k n_{(i)} > N_u$.
 - 6) We set $u_i = 1$ if $n_i > \theta$ and $u_i = 0$ if $n_i \leq \theta$.

We then repeated steps 1–6 for other provinces and for each year.



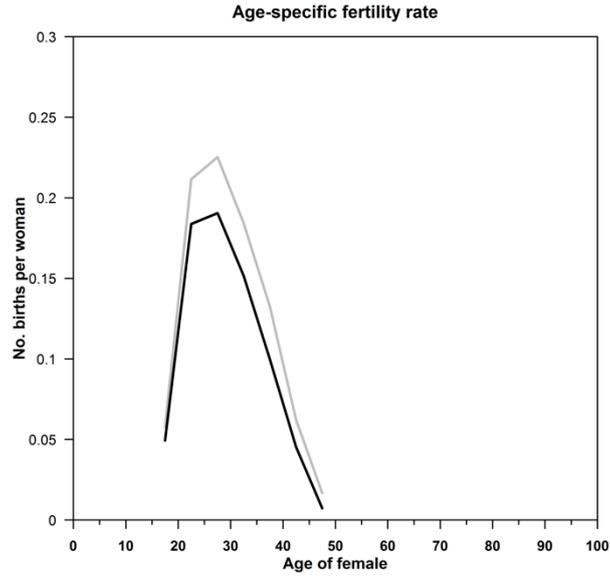
Supplementary Figure 3: Population density of Cambodia for 2010 and 2020. The population density of Cambodia changes over time. The darker regions indicate areas with lower population density and those in grey indicated regions with too few people living there. Upon close inspection,

Temporal projections

The temporal demographic model used to project the future burden of TB in Cambodia was constructed through the following submodels.

1. Data for the fertility rate model

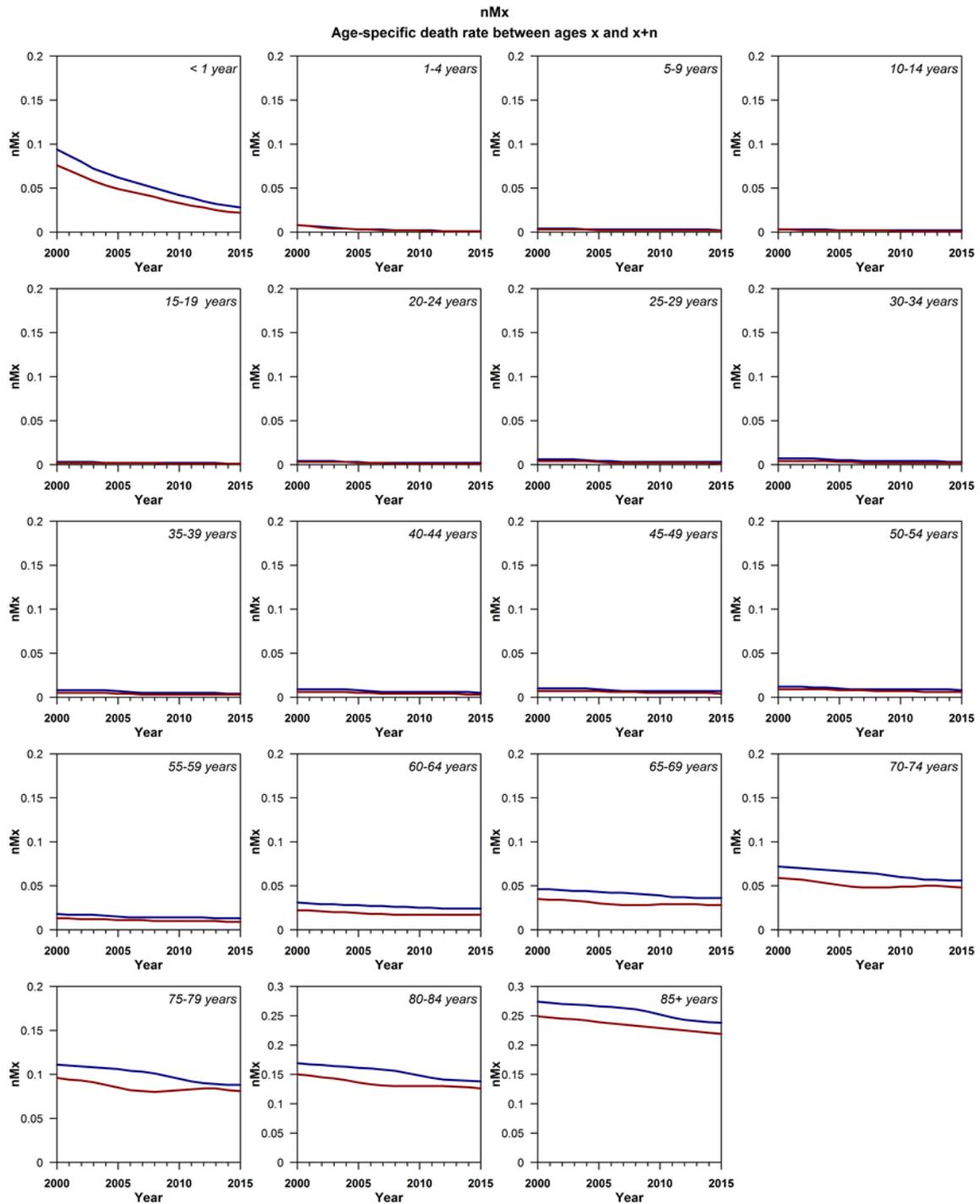
Fertility rates in Cambodia dropped from over 6 in the 1960s to just above replacement by the 2010s. This decline was approximately proportionate across maternal age for the period 1995 to 2000, and 2000 to 2005 (**Supplementary Figure 4**). Because the evolution of fertility rates after reaching replacement levels is not well understood and may vary due to cultural differences between countries, we did not seek to extrapolate the trend in fertility beyond the latest available date. Instead, we held the age-specific fertility rate at 2005 constant until 2030, possibly slightly overestimating fertility as a result.



Supplementary Figure 4: Age-specific fertility rate of Cambodia. The annual age specific fertility rates for five year age-bands across the country as a whole over the period 1995 to 2000 (grey line) and 2000 to 2005 (black line) were obtained from the United Nations Statistics Division.

2. *Data for the mortality rate model*

Historic changes in age- and sex-specific mortality rates were obtained from the Global Health Observatory data repository of the World Health Organization and are plotted in **Supplementary Figure 5**.



Supplementary Figure 5: Annual, sex-specific, five-year mortality rates for the 16-year period from 2000 to 2015. Data were extracted from the Global Health Observatory data repository of the World Health Organization

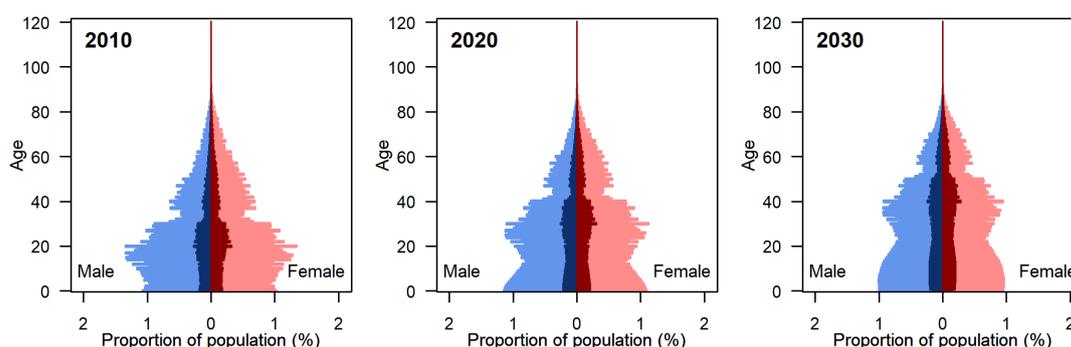
3. Urbanisation model

The proportion urbanised in province p in year t , U_{pt} , was derived for the years $t = 2010, 2015$ and 2020 as described later in step 4 of the *Spatial projection* section. To estimate the proportion in other years, we used linear regression to estimate U_{pt} , given the near linear relationship with time.

Reallocation between rural/urban strata was done to ensure the correct proportion urbanised was maintained over time in each province. Each year, we reallocated the population within a province proportionately to urban and rural strata. The overall demographic model therefore involved iterating the following steps:

1. Reallocating the population within provinces to rural and urban strata, based on the modelled proportion in each stratum for that province.
2. Projecting the number of births within each stratum/province, setting the sex according to the sex ratio.
3. Determining the number of deaths for each age group/sex/stratum/province and removing these from the population.
4. Aging the remaining population by a year.

Within a stratum of a province, the population was assigned to grid cells based on the projected population density in each grid cell, as described earlier in the supplementary information.



Supplementary Figure 6: DEMOKH’s projection of the Cambodian population for 2010–2030. The population pyramids by age and sex of Cambodia is classified by urbanisation extent (urban, the darker area, or rural, the lighter area). As the population of Cambodia ages forward in time, DEMOKH projects the population to be older and more urbanised.

III. Future TB projections

The projections of possible future active TB prevalence resulted from a three-stage procedure:

1. Imputations of the current spatial structure of the population;
2. Implemented DEMOKH on the spatial structure, ageing the existing population, depleting it by deaths, and replenishing it with births;
3. Application existing or reduced (as described in **Section IV**) age-specific point prevalence rates, stratified by sex and urban/rural strata, to the projected future spatial structure.

The annual mean prevalence and total number of cases were calculated for each 1km² grid and decennial projections are presented in **Figure 5** of the main paper.

We projected number of active TB cases to the year 2030 under three scenarios: *No Future Improvement*, *Continual Reduction*. The modelling provides estimates of likely disease burden (i) *No Future Improvement*: in the absence of increased funding for control programmes, (ii) *Continual Reduction*: continued reduction in TB prevalence seen in the NTP surveys of 2002 to 2011 as estimated by Eang and colleagues,⁶ and (iii) *GDP projection*: reduction in TB prevalence seen as Cambodia develops.

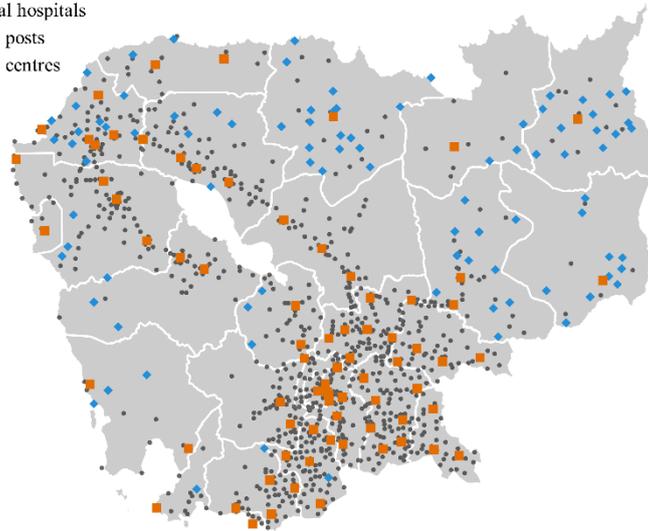
IV. Access to TB healthcare services

Government healthcare facilities

Cambodia has a well-established network of public health systems, consisting of referral hospitals, health centres, and health posts. These government healthcare facilities provide healthcare to the population across the country, from the urban cities to the remote communities (see **Supplementary Figure 7**). While a majority (61%) visit the private healthcare facilities for curative care, preventive care—immunization, TB testing and HIV/AIDS prevention and control—is dominated by the public/government healthcare facilities.⁷

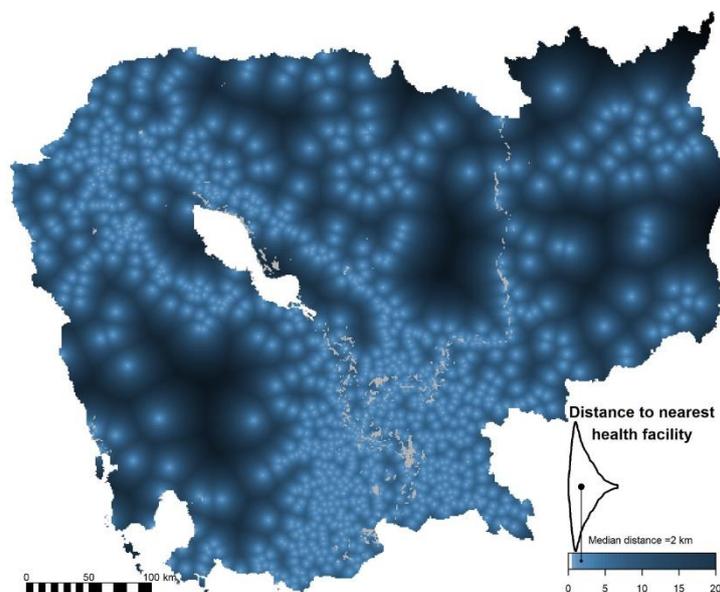
Government Health Facilities

- Referral hospitals
- ◆ Health posts
- Health centres



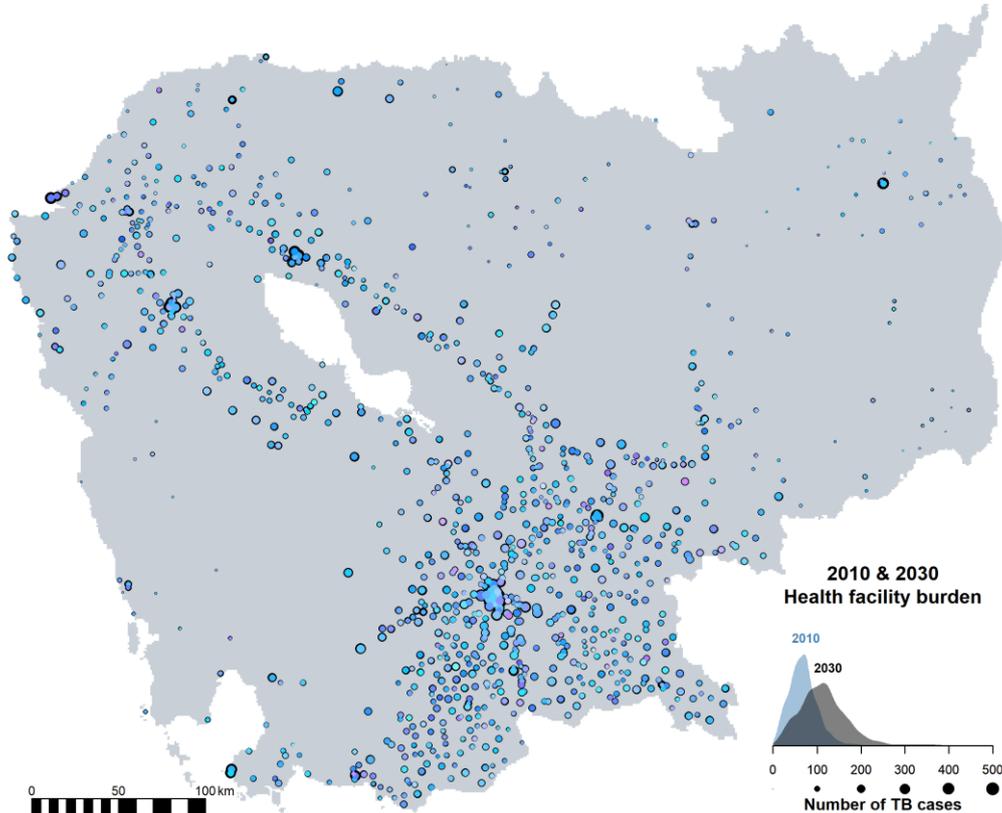
Supplementary Figure 7: Geographical data of government healthcare facilities. Referral hospitals, health posts, and health centres were extracted from United Nations Office for the Coordination of Humanitarian Affairs^{8,9}

The countrywide TB control programme was launched by National Center for Tuberculosis and Leprosy Control (known by its French acronym, CENAT) in 1994.^{10,11} The programme was purposed to set up to control and treat TB. Since then, directly observed therapy short-course (DOTS) was introduced, expanded across the nation, and provided at the community level (community DOTS or cDOTS).



Supplementary Figure 8: Distance to nearest government healthcare facility. The population-weighted median distance to the nearest healthcare facility is 2 km. The lighter regions indicate shorter distances (population-weighted) to the nearest healthcare facility.

Assuming the population visited the nearest healthcare facility for preventive care like TB care, on average, the nearest government healthcare facility is 2 km away (**Supplementary Figure 8**). We then coalesced the spatial healthcare facilities data with the spatial distribution of the active TB cases to project the number of cases visiting or in the catchment of each healthcare facility. The potential number of cases nearest to the government healthcare facility was calculated i.e. projected catchment size of the healthcare facility. In **Supplementary Figure 9**, we compared the projected catchment size in 2010 and 2030 given no future improvement to the TB care since 2011. We also presented the two scenarios for 2030 (**Figure 6** in the main paper): *No Future Improvement* and *Continual Reduction*.



Supplementary Figure 9: Projected catchment size of the government healthcare facility in Cambodia in 2010 versus 2030. The size of the points corresponds to the number of active TB cases within the catchment area of a government healthcare facility in 2010 (points in hues of blue) and 2030 (black points). The projected burden in 2030 by the *No Future Improvement* model is projected to be larger at almost all healthcare facilities.

List of abbreviations

CENAT: National Center for Tuberculosis and Leprosy Control

CRS: coordinate reference system

DEMOKH: Demographic Epidemiological Model of Cambodia

DOTS: Directly observed therapy short-course

EPSG: European Petroleum Survey Group

GIS: geographic information system

JAGS: Just Another Gibbs Sampler

km: kilometres

NTP: National Tuberculosis Prevalence

TB: tuberculosis

References

- 1 Gaughan AE, Stevens FR, Linard C, Jia P, Tatem AJ. High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015. *PLoS ONE* 2013; **8**: e55882.
- 2 The Worldpop Project. <http://www.worldpop.org.uk/> (accessed April 17, 2018).
- 3 QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation, 2009 <http://qgis.osgeo.org>.
- 4 ESRI. ArcGIS Desktop: Release 10. Redlands, California: Environmental Systems Research Institute, 2011.
- 5 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2017 <https://www.R-project.org/>.
- 6 Mao TE, Okada K, Yamada N, *et al.* Cross-sectional studies of tuberculosis prevalence in Cambodia between 2002 and 2011. *Bull World Health Organ* 2014; **92**: 573–81.
- 7 Pacific WHORO for the W. Cambodia-WHO country cooperation strategy 2016-2020. Manila : WHO Regional Office for the Western Pacific, 2016 <http://iris.wpro.who.int/handle/10665.1/13372> (accessed April 17, 2018).
- 8 United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Humanitarian Data Exchange. <https://data.humdata.org/> (accessed April 18, 2018).
- 9 Open Development Cambodia (ODC). Open Development Cambodia (ODC). <https://opendevdevelopmentcambodia.net/> (accessed April 18, 2018).
- 10 Hill PS, Tan Eang M. Resistance and renewal: health sector reform and Cambodia's national tuberculosis programme. *Bull World Health Organ* 2007; **85**: 631–6.
- 11 Khieu K, Ito K, Hamajima N. Experience in tuberculosis treatment through directly observed therapy short course in health centres and communities in Cambodia. *Public Health* 2007; **121**: 696–9.