Title:

Subnational burden estimation in Tuberculosis: generation and application of a new tool in Indonesia

Authors:

Cicilia Gita Parwati,[1] Muhammad Noor Farid,[2] Helmi Suryani Nasution,[3] Sulistyo,[3] Carmelia Basri,[2] Dina Lolong, [4] Agnes Gebhard,[1] Edine W Tiemersma, [5] Imran Pambudi,[3] Asik Surya,[3] Rein MGJ Houben[6, 7]

Affiliations

1. KNCV Tuberculosis Foundation, Jakarta, Indonesia
2. National TB Expert Committee, Jakarta, Indonesia
3. National TB Control Program, Ministry of Health, Jakarta, Indonesia
4. National Health Institute or Research and Development, Jakarta, Indonesia
5. KNCV Tuberculosis Foundation, The Hague, The Netherlands
6. TB Modelling Group, TB Centre, London School of Hygiene and Tropical Medicine, UK.
7. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, UK.

Corresponding Author

Cicilia Gita Parwati

KNCV Tuberculosis Foundation

Postbus 146, 2501 CC Den Haag

Netherlands

Tel. +62(0)81804957955

cparwati@gmail.com

Running head: Subnational Tuberculosis burden estimation

Word count abstract: 193

Word count text: 2499

No. of references: 33

Number of tables: 2

Number of figures: 3

Keywords

Target-setting, model, decision-making, district-level, incidence

Abstract

Setting In many high tuberculosis (TB) burden countries, there is substantial geographical heterogeneity in TB burden. In addition, decisions on TB funding and policy are highly decentralised. Subnational estimates of burden however are usually unavailable for planning and target-setting.

Objective and Design We developed SUBsET to distribute national TB incidence through a weighted score using selected variables, and applied for the 514 districts in Indonesia, which have substantial policy and budgetary autonomy in TB. Estimated incidence was compared to reported facility and domicile-based notifications to estimate the case detection rate (CDR). Local stakeholders led model development and dissemination.

Results The final SUBsET model included district population size, level of urbanisation, socio-economic indicators (living floor space and high school completion), HIV prevalence and air pollution. We estimated district-level TB incidence between 201 and 2,485/100,000/year. The facility-based CDR varied between 0 and 190% with high variation between neighbouring districts, e.g. suggesting strong cross-district health utilisation, which was confirmed by domicile-based CDR estimation. SUBsET results informed district-level TB action plans across Indonesia.

Conclusion Applying SUBsET to estimate the subnational burden can be important for high-burden countries and inform TB policy-setting at the relevant, decentralised administrative level.

INTRODUCTION

TB remains the leading cause of death from a single infectious agent and funding to fight the disease remains limited.[1] The burden of TB is widely assumed to be heterogeneously distributed within countries,[2] and policy decision-making, including setting TB care and prevention planning and budgeting, often takes place at the subnational level. To inform decision making at this level, and tailoring of TB care and prevention efforts to local epidemiology, subnational estimates of TB burden are key.

While many high TB burden countries have conducted national TB prevalence surveys to obtain a better estimate of their TB burden,[1] these surveys do not provide estimates on relevant subnational administrative levels. Various studies have reported subnational estimation of disease burden,[3–10] though few in TB, which often used complex methods that cannot be easily understood by local policy makers.[11–16] As such, subnational policy makers are usually left without estimates to inform planning. Data on TB notifications is usually available at subnational level, but provide a poor reflection of disease burden.[2]

Indonesia, with a total population of around 260 million people in 2017, consists of 34 provinces and 514 districts.[17,18] Since 1999, local (i.e. Provincial and District) governments have full autonomy to manage health, financing, planning, and budgeting.[19] Health care is provided by the public and a large private sector.[20] Although TB notification is mandatory, only 53% of all estimated incident cases were notified to the National TB Program (NTP).[1]

Following a recent inventory study, Indonesia is estimated to have approximately 842,000 incident TB cases a year in 2017.[1] To achieve ambitious targets for ending the TB epidemic by 2030, the Indonesia NTP has encouraged local governments to develop a district action plan,[21] that is linked to the National Strategic Plan but tailored to the local challenges, including estimated local burden and health system utilisation.

Our aim was to develop a tool to estimate district-level incidence and health system utilization, balancing detail and granularity with simplicity, so both method and result could be effectively disseminated to local government, and adapted for other high burden countries. We describe the development, findings and dissemination of the SUBsET (SUBnational Burden Estimation for Tb) tool.

METHODS

Principle of method

To promote acceptability and application of the results by policy makers, we worked from the principle that the model should be as simple as possible, use widely available software, and involve a limited number of calculation steps while still utilising available data in an efficient way. Data to inform the model was required to be available in 95% of districts and have an association with TB burden.

No separate ethics approval was obtained as all data were publicly available or anonymised at time of analysis. Model development, including the selection of variables, was inclusive, with direct input from the NTP, relevant partners and representatives from local academia. Taking into account that program indicators and milestones for the End TB strategy were set on incidence rather than prevalence, we chose TB incidence as our outcome.[22]

Data

Burden estimates

The national level incidence estimate from WHO Global TB report was used as the starting point.[1] In 2014, the prevalence survey found substantial differences in burden between 3 regions (Sumatera, Java-Bali, and Others, i.e. regions other than Sumatera and Java-Bali).[23] We applied the same distribution to the national incidence estimate.

Variables for model

Population size for each district was based on estimates from the Central Statistics Agency (BPS) that released a 2010-2020 district population projection for each province based on 2010 National Population survey.[24]

Additional variables were extracted from the National Socio-Economic Survey, an annual socio-demographic survey which covers the whole nation and is powered for district-level estimates.[25] We identified urbanisation, floor space, and education level (see table 1 for definitions), which were also measured in the prevalence survey. We also included HIV burden,[26,27] and air pollution levels,[28–30].

To inform current health system performance or utilisation, and to check estimated values of burden, the NTP provided both domicile-based (according to patient's address) and health-facility-based (according to facility address) notification data for each district.[31]

Model

The SUBsET tool combined all available data to distribute National TB burden through a weighted score for each of the 514 districts, through the steps outlined below.

Step 1: Regional incidence

Incidence estimate of those three regions was calculated by applying the distribution of absolute TB prevalence across the respective regions among 2017 Indonesia population in the respective regions:

$$I_r^{(case)} = \frac{P_r^{(case)}}{P^{(case)}} \times I^{(case)}$$

where:

$I_r^{(case)}$	= Estimated TB incident cases in region r
$Pr^{(case)}$	= TB prevalent cases (absolute value) in region r
$I^{(case)}$	= National TB incident cases (absolute value)
$P^{(case)}$	= National TB prevalent cases (absolute value)

Step 2: Variable weight

For the socioeconomic variables, through conducting multivariable logistic regression we were able to estimate the relative risk directly, by region, from the 2014 prevalence survey.[23] For HIV prevalence and air pollution, values from the literature were used.[27–30]

We then calculated a weight for each variable by multiplying the regional relative risk with the proportion in that district (e.g. proportion living in an urban area):

$$S_d^{(v)} = \left(Pr(v_d) \times RR_r^{(v)}\right) + (1 - Pr(v_d))$$

where,

$S_d^{(v)}$	= weight for variable (v) in district d
$Pr(v_d)$	= proportion variable (v) among population in district d
$RR_r^{(v)}$	= TB relative risk ratio for variable (v) in region r
$1 - Pr(v_d)$	= 1 – proportion of variable (v) in district d

Step 3: Calculation of total weight score per district

A total score for each district was calculated by multiplying all variable weights with the population size:

$$S_d = N_d \times s_d^{(floor/kapita<8m^2)} \times s_d^{(urban)} \times s_d^{(low\ education)} \times s_d^{(HIV)} \times s_d^{(air\ pollution)}$$

where:

S_d = total score for district d

N_d = number of population in district d

$s_d^{(floor/kapita<8m^2)}$ = weight score for variable living floor space in district d

$s_d^{(urban)}$ = weight score for variable level of urbanisation in district d

$s_d^{(low\ education)}$ = weight score for variable junior high school completion in district d

$s_d^{(HIV)}$ = weight score for variable HIV prevalence in district d

$s_d^{(air\ pollution)}$ = weight score for variable air pollution prevalence in district d

Step 4: Distribution of burden

Total weight score per region was calculated by adding up the total weight score per district by respective region, and then distributing the estimated burden across districts-based total on district score from step 3:

$$I_d^{(case)} = \frac{S_d}{S_r} \times I_r^{(case)}$$

where:

$I_d^{(case)}$ = Estimated TB incident cases in district d

S_d = Total weight score in district d

$I_r^{(case)}$ = Estimate TB incident cases in region r

S_r = Total weight score all districts in region r

Calculation of district-level Case Detection Rate (CDR)

To estimate the district-level CDR, the estimated burden in each district was compared to both domicile- and health-facility-based reported notifications. Comparing both domicile- and health-facility-based notifications within and between surrounding districts allowed assessment of district health system performance and cross-district health utilisation.

Validation of SUBsET results

While model validation with data is desirable,[32] neither the prevalence survey or inventory study enabled a district-level comparison. The prevalence survey did not cover complete districts, and the inventory study was powered to provide a national, not district-level estimates. An attempt to use inventory study data at the district level would lead to extremely wide uncertainty intervals around the therefore non-informative point estimates.

Dissemination and adoption of model

The model was disseminated and discussed at provincial and district levels, followed by a round of revisions during a national-level stakeholder meeting. The final development step resulted in the addition of two variables to capture strong heterogeneity in HIV prevalence, and measured air pollution between districts.

Uncertainty intervals

Uncertainty intervals were calculated by generating 10,000 random draws from the distribution for both the regional incidence estimate as well as relative risks for included variables.[23,27-30]

Sensitivity analysis

To understand the heterogeneity captured by our model, we compared the results of our calculation with an estimate based on regional incidence and population size alone. We also performed a calculation where we removed each individual variable and compared the results with the full model.

The model was set up in Microsoft Excel, multivariate analyses for region specific TB relative risks were conducted in STATA version 14. We used spmap ado file in STATA version 14 to create the maps which visualise the distribution of the district TB burden estimates and CDR throughout Indonesia, particularly within provinces, thus allowing us to better understand the connection or relationship between one area to another.

RESULTS

Model

Relative risks for the model variables used in step 2 are shown in Table 2.

The range of values across districts for each risk factor was wide (see Table 2, column 4 and 5). When the relative risks were combined with the data for each risk factor, differences in population weight for districts were found in each region i.e. median (range) relative weights Sumatera 2.52 (2.29-2.75), Java-Bali 1.50 (1.37-1.64), and Others 2.10 (1.91-2.29).

District-level TB Incidence

Fig 1 shows the distribution of the SUBsET estimated TB incidence across the 514 districts in Indonesia. The estimated point values for TB incidence ranged between 201 and 2,485/100,000/year. The estimated

TB incidence rates was lowest at Java-Bali region (average median 242/100,000, range 201-787) compared to Sumatera (373/100,000, 295-918) and Others (350/100,000, 280-2485). However, considering that 58% of the total population of Indonesia resides in Java-Bali,[20] this region has the highest absolute number of TB cases.[23]

District-level CDR

Fig 2 shows the distribution of the estimated facility-based CDR throughout all districts. While some districts have very low CDR (0-20%, dark red colour) some others have very high CDR (>100%, green colour) with a range of 0 to 190%. Among 24 (5%) districts with an estimated facility-based CDR of more than 100%, 15 were urban and suburban districts, surrounded by rural districts, which usually have fewer or lower quality TB services (Fig 2, pull outs). Twenty-one districts (4%) had an estimated facility-based CDR between 80 and 100%.

For domicile-based CDR, 9 (2%) districts had an estimated CDR of more than 100%. A further 24 (5%) districts had a domicile-based CDR between 80% and 100% and 51 (10%) had a domicile-based CDR below 20%. At the district level, there was considerable contrast between facility and domicile-based CDR. As an example, for the year 2017, Salatiga city, Surakarta and Magelang city had 121%, 129% and 170% facility-based CDR while the domicile-based CDR were only 32%, 39% and 33% respectively (Fig 2, pulls out).

Uncertainty analyses

Uncertainty analyses provided ranges for incidence rate per 100,000/year population at district level as well as at regional level. For Sumatera Region, this resulted in value (95% uncertainty interval) of 413.4 (305.3-530.8), for Java-Bali 268.0 (212.3-321.0), and for Others 380.1 (277.8-495.9). District-level uncertainty intervals are shown in figure 3.

Sensitivity analyses

Figure 3 shows the additional variation in estimated incidence introduced by the variables in our model, by comparing with a model including population size and regional differences in prevalence. We found that 30% of the districts had a higher and 70% had a lower point estimate for TB incidence rates compared to previous estimates. The newly estimated TB incidence rates were more than 10% different (higher or lower) from the previous TB incidence estimate for 73% of the districts.

Removing a single variable had no relevant impact on the distribution of the estimated burden in the model which shows that there is no single model variable that dominates the differentiation between districts. Considering the dominant influence of population size in the burden distribution across districts, a lower or higher value of a relative risk in a single variable would lead to a lower or higher value of the uncertainty interval.

Model dissemination

The district- and provincial-level TB burden estimates were fed into the development of District and Provincial Action Plans, particularly to inform policy decisions on budget, resource allocation, and intervention planning. Estimates were also incorporated in the 2016-2020 TB National Strategic Plan, and have fed into joint AIDS, TB and Malaria policy meetings at the national level.[33]

DISCUSSION

The SUBsET tool approach was found to provide an accessible and intuitive model for subnational burden estimation. Our final model used five variables to distribute TB incidence from three regional estimates across 514 districts in Indonesia. The model provided substantial differentiation, estimating an incidence ranging between 201 and 2,485/100,000/year. The facility-based CDR varied between 0 and 190%, highlighting low-performing districts, and cross-district health utilisation. Dissemination of the SUBsET tool showed rapid uptake and acceptance of results.

On district-level, the SUBsET facilitated the comparison of facility-based and domicile-based-CDR which highlighted previously unrecognised cross-district health system utilization. These insights encouraged such districts to improve their own health care system and case detection, as well as improve collaboration with neighbouring districts.

Limitations

Our work has several limitations. Both the regional distribution of incidence and associations between TB burden and socioeconomic variables are based on the 2014 national TB prevalence survey, not on directly measured incidence. While those associations may be slightly different if directly calculated for incidence, we feel they are a reasonable approximation and the limited bias is outweighed by the ability to calculate the relative risks directly for the population and time period. For HIV, the association matches the range of the relative risk of developing TB in HIV-positive infected persons in concentrated and low-level HIV prevalence area; likewise, the association between air pollution and risk of developing TB corresponds with results found in various studies from low to middle income settings.[27–30]

Second, we acknowledge the likelihood that there may be a residual or uncaptured variation of TB incidence beyond that captured by the model, e.g. due to differential levels of malnutrition, or in additional sub-categories within the variables included, but data was not available to include in the model.

Third, we recognize the inability of conducting results validation due to unavailability of data. This prevents the assessment of consistency between the results of our model and other evidence and/or the true burden at district level; however, with future availability of data, the model can be continuously updated and be validated.

Advantages

Within these limitations we achieved our main aim to keep the SUBsET tool simple and intuitive, enabling the rapid dissemination and further country-led adaptation of the model. Using publicly available data also helped the results to be acceptable to the autonomous District Health Office staff. While it is theoretically possible that a more complicated (and effectively 'black box' model[11,13]) approach could have been equally successful as our intuitive and open approach, input from Indonesian stakeholders at the start, and local feedback throughout the process, suggests our judged approach was correct.

Through the above, SUBsET filled an urgent need within the Indonesia NTP to help inform with- and between-districts discussions. Furthermore, adding variables, or new districts is relatively easy, and shows how SUBsET provides a template for other countries to consider when looking for subnational advocacy, provided data are available.

CONCLUSIONS

The transparent modelling approach applied in SUBsET enabled understanding, ownership, and acceptance among the sub-national decision makers in Indonesia. Our approach shows how local data can be utilised to estimate subnational burden, thus providing a template for adaptation in other high burden countries to enable them to inform TB policy at the relevant, decentralised administrative level.

Acknowledgements

We would like to express our gratitude to all TB staffs, policy makers, and academicians who attended the workshop both at National and Provincial level and contributed to model development and results disseminations. We also gratefully acknowledge Rizka Nurfadila who had prepared the data required for this study.

Authors contribution: Cicilia Gita Parwati, Muhammad Noor Farid and Rein MGJ Houben contributed in model conceptualization, formal analysis, validation and writing the manuscript. Agnes Gebhard and Edine Tiemersma contributed in editing the manuscript. All authors provided critical feedback on methodology and manuscript development and declare no conflict of interest.
This work was supported by USAID and Global Fund against AIDS, Tuberculosis and Malaria. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

References

1. World Health Organization. Global tuberculosis report 2018. Geneva: World Health Organization, 2018.

2. Trauer JM, Dodd PJ, Gomes MGM, et al. The Importance of Heterogeneity to the Epidemiology of Tuberculosis. Clin Infect Dis. 2018.

3. Larmarange J, Bendaud V. HIV estimates at second subnational level from national population-based surveys. AIDS Lond Engl. 2014 Nov;28 Suppl 4:S469-76.

4. Marty L, Cazein F, Panjo H, Pillonel J, Costagliola D, Supervie V. Revealing geographical and population heterogeneity in HIV incidence, undiagnosed HIV prevalence and time to diagnosis to improve prevention and care: estimates for France. J Int AIDS Soc. 2018 Mar;21(3):e25100.

5. Benova L, Awad SF, Miller FD, Abu-Raddad LJ. Estimation of hepatitis C virus infections resulting from vertical transmission in Egypt. Hepatology. 2015 Mar;61(3):834–42.

6. Song P, Wang J, Bucan K, Theodoratou E, Rudan I, Chan KY. National and subnational prevalence and burden of glaucoma in China: A systematic analysis. J Glob Health. 2017 Dec;7(2).

7. Barron S, Balanda K, Hughes J, Fahy L. National and subnational hypertension prevalence estimates for the Republic of Ireland: better outcome and risk factor data are needed to produce better prevalence estimates. BMC Public Health. 2014 Jan 10;14:24.

8. Virani S, Bilheem S, Chansaard W, et al. National and Subnational Population-Based Incidence of Cancer in Thailand: Assessing Cancers with the Highest Burdens. Cancers. 2017 Aug 17;9(12):108.

9. Techasaensiri C, Radhakrishnan A, Als D, Thisyakorn U. Typhoidal Salmonella Trends in Thailand. Am J Trop Med Hyg. 2018 Sep;99(3_Suppl):64–71.

10. Raskind-Hood C, Hogue C, Overwyk KJ, Book W. Estimates of adolescent and adult congenital heart defect prevalence in metropolitan Atlanta, 2010, using capture-recapture applied to administrative records. Ann Epidemiol. 2018 Dec 5.

11. Rood E, Khan A, Modak P, et al. A Spatial Analysis Framework to Monitor and Accelerate Progress towards SDG 3 to End TB in Bangladesh. ISPRS Int J Geo-Inf. 2018;8(1):14.

12. Parker JD, Kruszon-Moran D, Mohadjer LK, et al. National Health and Nutrition Examination Survey: California and Los Angeles County, Estimation Methods and Analytic Considerations, 1999-2006 and 2007-2014. Vital Health Stat 2. 2017 May;(173):1–26.

13. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016 Oct;388(10053):1545–602.

14. Avilov KK, Romanyukha AA, Borisov SE, Belilovsky EM, Nechaeva OB, Karkach AS. An approach to estimating tuberculosis incidence and case detection rate from routine notification data. Int J Tuberc Lung Dis. 2015 Mar 1;19(3):288–94.

15. Farzadfar F, Delavari A, Malekzadeh R, et al. NASBOD 2013: design, definitions, and metrics. Arch Iran Med. 2014 Jan;17(1):7–15.

16. Begg S, Vos T, Barker B, Stevenson C, Stanley L, Lopez A. The burden of disease and injury in Australia 2003. Phe. 2003;82:337.

17. World Bank. Indonesia country overview. http://www.worldbank.org/en/country/indonesia/overview. Accessed December 2018.

18. BPS-Statistics Indonesia. Statistical Yearbook of Indonesia 2018. Sub-directorate of Statistical Compilation and Publication, Editor.2018.

19. Government of Indonesia. Law No 22/1999 on Local Government (Undang-Undang Republik Indonesia No 22 Tahun 1999 Tentang Pemerintahan Daerah). 1999.

20. Mahendradhata Y, Trisnantoro L, Listyadewi S, Soewondo P, Marthias T, Harimurti P. The Republic of Indonesia Health System Review. 2017; 7: 64–105 p.

21. Indonesia Ministry of Health. Petunjuk Penyusunan Rencana Aksi Daerah untuk Tuberkulosis. Jakarta, 2017.

22. Executive Board, 134. (2014). Global strategy and targets for tuberculosis prevention , care and control after 2015 Report by the Secretariat. http://www.who.int/iris/handle/10665/172828

23. Ministry of Health Indonesia. Indonesia Tuberculosis Prevalence Survey. 2015;(June).

24. BPS-Statistics Indonesia. Proyeksi Penduduk Kabupaten/Kota Provinsi Aceh 2010-2020. Jakarta, 2015.

25. BPS-Statistics Indonesia. National Socio-Economic Survey 2017.

26. Brown T, Peerapatanapokin W. The Asian Epidemic Model: a process model for exploring HIV policy and programme alternatives in Asia. Sex Transm Infect. 80(Suppl 1):i19–24.

27. Getahun H, Gunneberg C, Granich R, Nunn P. HIV Infection – Associated Tuberculosis : The Epidemiology and the Response. 2010;50:201–7.

28. Rajaei E, Hadadi M, Madadi M, et al. Outdoor Air Pollution Affects Tuberculosis Development Based on Geographical Information System Modeling. Biomed Biotechnol Res J. 2018;2(1):39–45.

29. Liu Y, Cui LL, Hou LJ, et al. Ambient Air Pollution Exposures and Newly Diagnosed Pulmonary Tuberculosis in Jinan, China: A Time Series Study. Sci Rep. 2018;8(1):1–11.

30. Kim J. Is ambient air pollution another risk factor of tuberculosis? Korean J Intern Med. 2014 Mar;29(2):170–2.

31. World Health Organization. Epidemiological Review in Indonesia. Jakarta; 2017.

32. Menzies NA, McQuaid CF, Gomez GB, Siroka A, Glaziou P, Floyd K, et al. Improving the quality of modelling evidence used for tuberculosis policy evaluation. Int J Tuberc Lung Dis. 2019 Apr 1;23(4):387–95.

33. Kementrian Kesehatan Republik Indonesia. Panduan Penentuan Beban dan Target Cakupan Penemuan dan Pengobatan Tuberkulosis di Indonesia Tahun 2019-2024. Jakarta; 2018.

Table 1. Variable sources and definitions

Variable	Definition	Range	Data Source
Population size	Number of individuals per district	13,763 to 5,682,911	Projected Population of Regency/City 2010-2020, Statistics Indonesia
Level of urbanisation	Proportion of population that lives in urban area	0% to 100%	National Socio-Economic Survey 2017
Living floor space	Proportion of individuals who live in a house with less than $8m^2$/person	0% to 92%	National Socio-Economic Survey 2017
Junior high school completion	Proportion of individuals who did not complete junior high school or less	29% to 76%	National Socio-Economic Survey 2017
HIV	Proportion of individuals with HIV infection	0% to 23%	National AIDS programme 2012
Air pollution	Proportion of individuals with air pollution exposure	5% to 100%	Meteorological, Climatological, and Geophysical Agency (BMKG) 2017

Table 2. Results from multivariate analysis of 2013/2014 TB Prevalence Survey

RISK FACTORS-TB ASSOCIATIONS				
Variable	**Region**	**Relative Risk**	**Lower***	**Upper***
Living in urban area	Sumatera	1.72	1.22	2.44
	Java-Bali	1.32	0.93	1.88
	Others	1.30	0.92	1.82
Living in a house less than 8m^2/person	Sumatera	1.50	1.03	2.19
	Java-Bali	1.30	0.83	2.06
	Others	1.15	0.79	1.65
Not completing junior high school	Sumatera	1.11	0.78	1.60
	Java-Bali	1.34	0.90	2.00
	Others	1.61	1.10	2.36
HIV prevalence	All regions	30	20	45
Air pollution	All regions	1.47	1.20	1.80

* Lower and upper bounds reflect 95% confidence interval. Note, relative risks for HIV prevalence and air pollution were not available by region, but came from literature.[26–29]

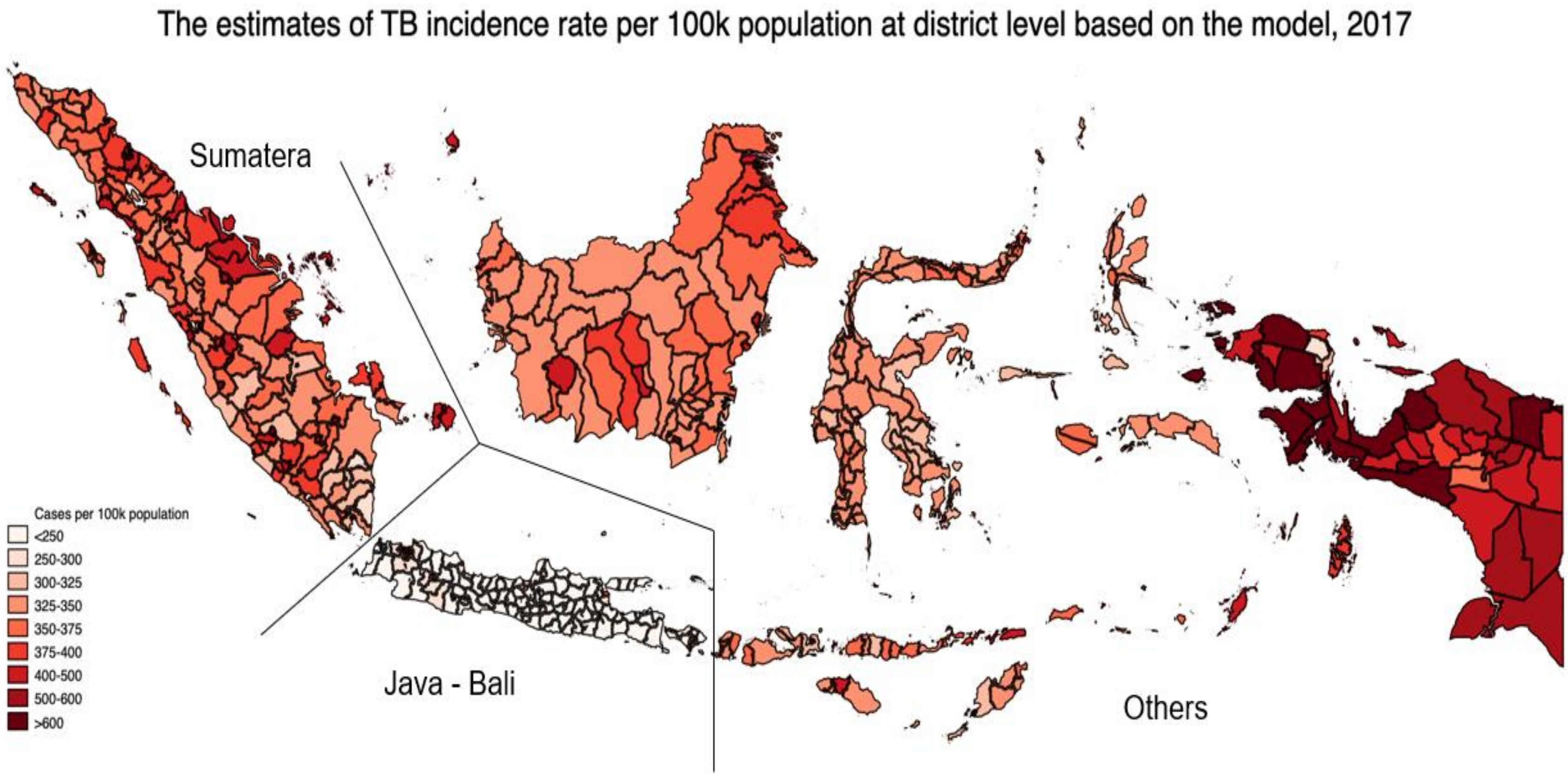

Fig 1. Estimated incidence per 100 000/year by district.

Figure shows three regions (solid lines) and 514 districts with their estimated incidence per 100,000 population.

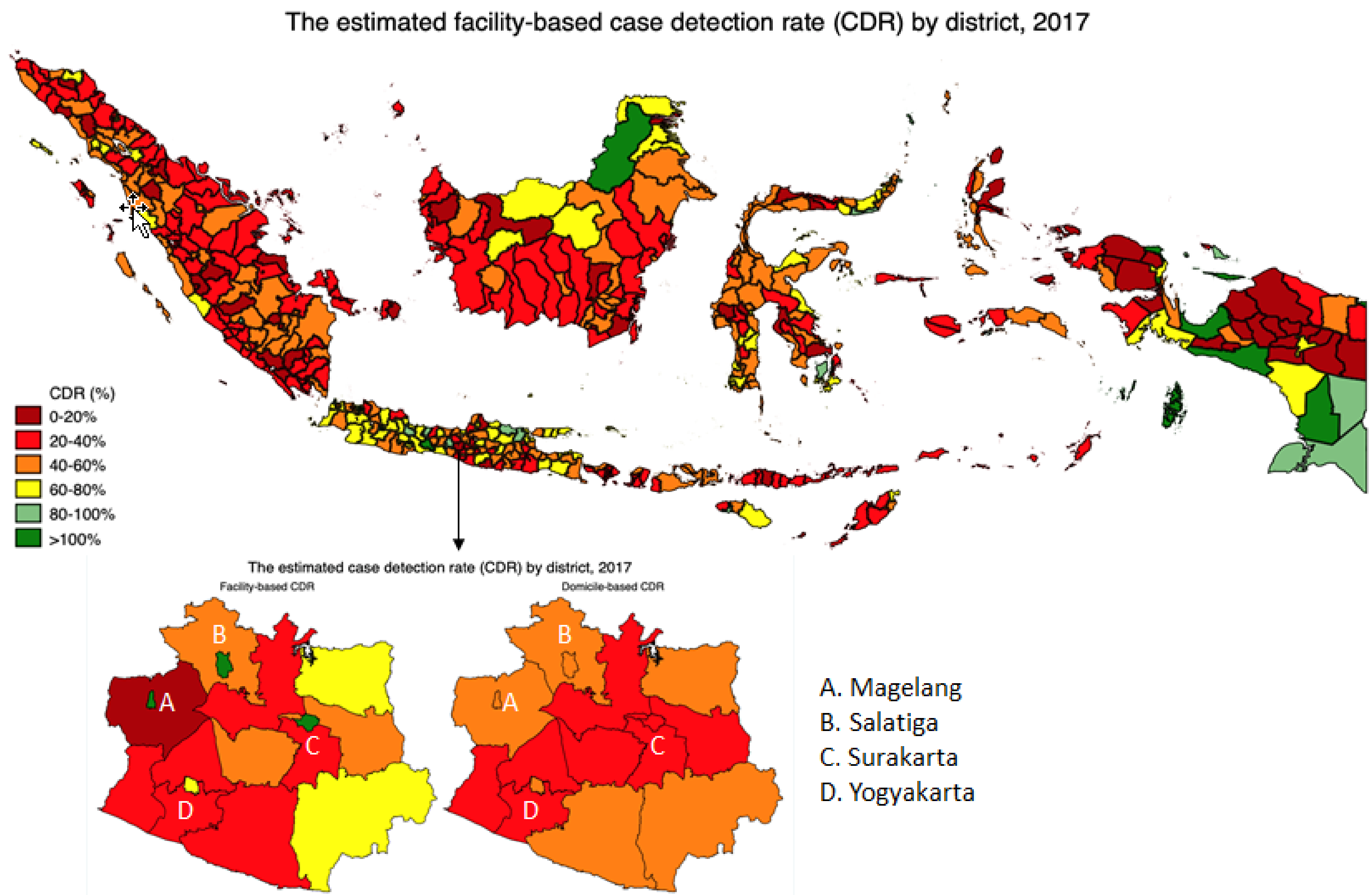

Fig 2. The distribution of the estimated facility-based case detection rate (CDR).

National map shows distribution of estimated facility-based CDR across the 514 districts. Pull-out figure shows very high facility-based CDR (more than 100%, green colour) in central urban districts, and low facility-based CDRs in surrounding districts. When viewed as domicile-based CDR, these differences in CDR are no longer present, highlighting cross-district health system utilisation.

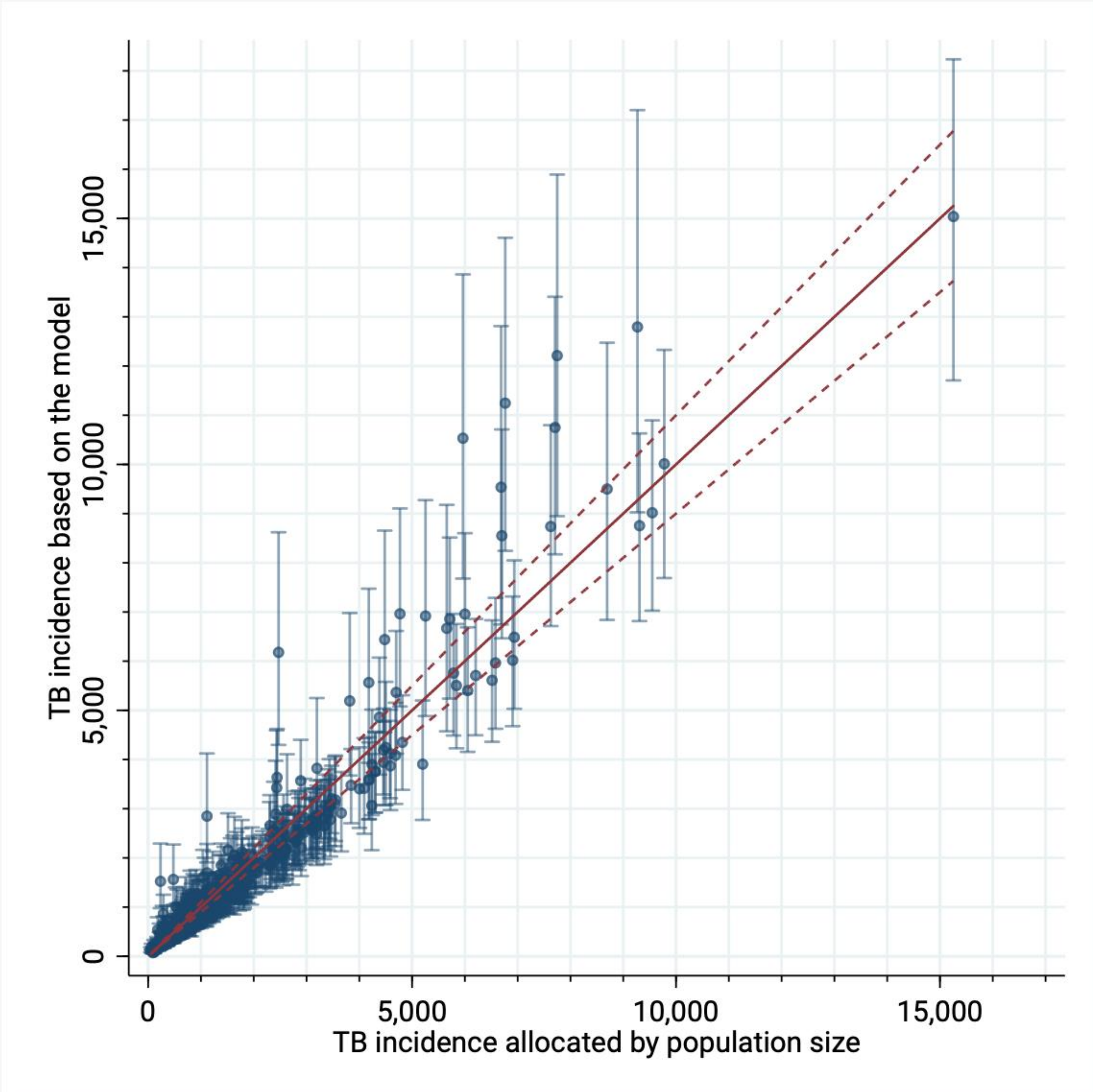

Fig 3. Heterogeneity captured by model variables.

Figure shows change in estimated absolute incidence with 95% uncertainty interval from a model with population size and regional differences in prevalence only (X-axis), and a model from SUBsET (Y-axis). Markers above/lower straight red line indicate districts with a higher/lower estimate based on the full model compared to the simple model.