

1 Title:

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

3

4 Authors:

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

7

8 Affiliations

9 1. KNCV Tuberculosis Foundation, Jakarta, Indonesia

10 2. National TB Expert Committee, Jakarta, Indonesia

11 3. National TB Control Program, Ministry of Health, Jakarta, Indonesia

12 4. National Health Institute or Research and Development, Jakarta, Indonesia

13 5. KNCV Tuberculosis Foundation, The Hague, The Netherlands

14 6. TB Modelling Group, TB Centre, London School of Hygiene and Tropical Medicine, UK.

15 7. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical
16 Medicine, UK.

17

18 Corresponding Author

19 Cicilia Gita Parwati

20 KNCV Tuberculosis Foundation

21 Postbus 146, 2501 CC Den Haag

22 Netherlands

23 Tel. +62(0)81804957955

24 cparwati@gmail.com

25

26 **Running head:** Subnational Tuberculosis burden estimation

27 **Word count abstract:** 193

28 **Word count text:** 2499

29 **No. of references:** 33

30 **Number of tables:** 2

31 **Number of figures:** 3

32

33 Keywords

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

35 **Abstract**

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

39

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

45

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

52

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

55

56

57 **INTRODUCTION**

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

63

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

70

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

76

77 Following a recent inventory study, Indonesia is estimated to have approximately 842,000 incident TB
78 cases a year in 2017.¹ To achieve ambitious targets for ending the TB epidemic by 2030, the Indonesia
79 NTP has encouraged local governments to develop a district action plan,²¹ that is linked to the National
80 Strategic Plan but tailored to the local challenges, including estimated local burden and health system
81 utilisation.

82

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

87

88 **METHODS**

89 **Principle of method**

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

94

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

100

101 **Data**

102 *Burden estimates*

103 The national level incidence estimate from WHO Global TB report was used as the starting point.¹ In
104 2014, the prevalence survey found substantial differences in burden between 3 regions (Sumatera, Java-
105 Bali, and Others, i.e. regions other than Sumatera and Java-Bali).²³ We applied the same distribution to
106 the national incidence estimate.

107

108 *Variables for model*

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

112

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

117

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

121

122 **Model**

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

125

126 Step 1: Regional incidence

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

129
$$I_r^{(case)} = \frac{P_r^{(case)}}{P^{(case)}} \times I^{(case)}$$

130 where:

131 $I_r^{(case)}$ = Estimated TB incident cases in region r

132 $P_r^{(case)}$ = TB prevalent cases (absolute value) in region r

133 $I^{(case)}$ = National TB incident cases (absolute value)

134 $P^{(case)}$ = National TB prevalent cases (absolute value)

135

136 Step 2: Variable weight

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

140

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

143
$$S_d^{(v)} = (Pr(v_d) \times RR_r^{(v)}) + (1 - Pr(v_d))$$

144 where,

145 $S_d^{(v)}$ = weight for variable (v) in district d

146 $Pr(v_d)$ = proportion variable (v) among population in district d

147 $RR_r^{(v)}$ = TB relative risk ratio for variable (v) in region r

148 $1 - Pr(v_d)$ = 1 – proportion of variable (v) in district d

149

150 Step 3: Calculation of total weight score per district

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

152
$$S_d = N_d \times s_d^{(floor/kapita < 8m^2)} \times s_d^{(urban)} \times s_d^{(low\ education)} \times s_d^{(HIV)} \times s_d^{(air\ pollution)}$$

153 where:

154 S_d = total score for district d

155 N_d = number of population in district d

156 $S_d^{(floor/kapita < 8m^2)}$ = weight score for variable living floor space in district d

157 $S_d^{(urban)}$ = weight score for variable level of urbanisation in district d

158 $S_d^{(low\ education)}$ = weight score for variable junior high school completion in district d

159 $S_d^{(HIV)}$ = weight score for variable HIV prevalence in district d

160 $S_d^{(air\ pollution)}$ = weight score for variable air pollution prevalence in district d

161

162 Step 4: Distribution of burden

163 Total weight score per region was calculated by adding up the total weight score per district by respective

164 region, and then distributing the estimated burden across districts-based total on district score from step 3:

165
$$I_d^{(case)} = \frac{S_d}{S_r} \times I_r^{(case)}$$

166 where:

167 $I_d^{(case)}$ = Estimated TB incident cases in district d

168 S_d = Total weight score in district d

169 $I_r^{(case)}$ = Estimate TB incident cases in region r

170 S_r = Total weight score all districts in region r

171

172 Calculation of district-level Case Detection Rate (CDR)

173 To estimate the district-level CDR, the estimated burden in each district was compared to both domicile-

174 and health-facility-based reported notifications. Comparing both domicile- and health-facility-based

175 notifications within and between surrounding districts allowed assessment of district health system

176 performance and cross-district health utilisation.

177

178 Validation of SUBsET results

179 While model validation with data is desirable,³² neither the prevalence survey or inventory study enabled

180 a district-level comparison. The prevalence survey did not cover complete districts, and the inventory

181 study was powered to provide a national, not district-level estimates. An attempt to use inventory study

182 data at the district level would lead to extremely wide uncertainty intervals around the therefore non-

183 informative point estimates.

184

185 **Dissemination and adoption of model**

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

190

191 **Uncertainty intervals**

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

194

195 **Sensitivity analysis**

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

199

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

205

206 **RESULTS**

207 **Model**

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

209

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

214

215 **District-level TB Incidence**

216 Fig 1 shows the distribution of the SUBsET estimated TB incidence across the 514 districts in Indonesia.

217 The estimated point values for TB incidence ranged between 201 and 2,485/100,000/year. The estimated

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

222

223 **District-level CDR**

224

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

231

232 For domicile-based CDR, 9 (2%) districts had an estimated CDR of more than 100%. A further 24 (5%)
233 districts had a domicile-based CDR between 80% and 100% and 51 (10%) had a domicile-based CDR
234 below 20%. At the district level, there was considerable contrast between facility and domicile-based
235 CDR. As an example, for the year 2017, Salatiga city, Surakarta and Magelang city had 121%, 129% and
236 170% facility-based CDR while the domicile-based CDR were only 32%, 39% and 33% respectively (Fig
237 2, pulls out).

238

239 **Uncertainty analyses**

240 Uncertainty analyses provided ranges for incidence rate per 100,000/year population at district level as
241 well as at regional level. For Sumatera Region, this resulted in value (95% uncertainty interval) of 413.4
242 (305.3-530.8), for Java-Bali 268.0 (212.3-321.0), and for Others 380.1 (277.8-495.9). **District-level**
243 **uncertainty intervals are shown in figure 3.**

244

245 **Sensitivity analyses**

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

251

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

257

258 **Model dissemination**

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

263

264 **DISCUSSION**

265 The SUBsET tool approach was found to provide an accessible and intuitive model for subnational
266 burden estimation. Our final model used five variables to distribute TB incidence from three regional
267 estimates across 514 districts in Indonesia. The model provided substantial differentiation, estimating an
268 incidence ranging between 201 and 2,485/100,000/year. The facility-based CDR varied between 0 and
269 190%, highlighting low-performing districts, and cross-district health utilisation. Dissemination of the
270 SUBsET tool showed rapid uptake and acceptance of results.

271

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

276

277 **Limitations**

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

286

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

290

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

295

296 **Advantages**

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

303

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

308

309 **CONCLUSIONS**

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

314

315 **Acknowledgements**

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

320 Authors contribution: Cicilia Gita Parwati, Muhammad Noor Farid and Rein MGJ Houben contributed in
321 model conceptualization, formal analysis, validation and writing the manuscript. Agnes Gebhard and
322 Edine Tiemersma contributed in editing the manuscript. All authors provided critical feedback on
323 methodology and manuscript development and declare no conflict of interest.

324 This work was supported by USAID and Global Fund against AIDS, Tuberculosis and Malaria. The
325 funders of the study had no role in study design, data collection, data analysis, data interpretation, or
326 writing of the report. The corresponding authors had full access to all the data in the study and had final
327 responsibility for the decision to submit for publication.

328

329 **References**

330

- 331 1. World Health Organization. Global tuberculosis report 2018. Geneva: World Health
332 Organization, 2018.
- 333 2. Trauer JM, Dodd PJ, Gomes MGM, et al. The Importance of Heterogeneity to the
334 Epidemiology of Tuberculosis. *Clin Infect Dis*. 2018.
- 335 3. Larmarange J, Bendaud V. HIV estimates at second subnational level from national
336 population-based surveys. *AIDS Lond Engl*. 2014 Nov;28 Suppl 4:S469-76.
- 337 4. Marty L, Cazein F, Panjo H, Pillonel J, Costagliola D, Supervie V. Revealing geographical
338 and population heterogeneity in HIV incidence, undiagnosed HIV prevalence and time to
339 diagnosis to improve prevention and care: estimates for France. *J Int AIDS Soc*. 2018
340 Mar;21(3):e25100.
- 341 5. Benova L, Awad SF, Miller FD, Abu-Raddad LJ. Estimation of hepatitis C virus infections
342 resulting from vertical transmission in Egypt. *Hepatology*. 2015 Mar;61(3):834–42.
- 343 6. Song P, Wang J, Bucan K, Theodoratou E, Rudan I, Chan KY. National and subnational
344 prevalence and burden of glaucoma in China: A systematic analysis. *J Glob Health*. 2017
345 Dec;7(2).
- 346 7. Barron S, Balanda K, Hughes J, Fahy L. National and subnational hypertension prevalence
347 estimates for the Republic of Ireland: better outcome and risk factor data are needed to
348 produce better prevalence estimates. *BMC Public Health*. 2014 Jan 10;14:24.
- 349 8. Virani S, Bilheem S, Chansaard W, et al. National and Subnational Population-Based
350 Incidence of Cancer in Thailand: Assessing Cancers with the Highest Burdens. *Cancers*.
351 2017 Aug 17;9(12):108.
- 352 9. Techasaensiri C, Radhakrishnan A, Als D, Thisyakorn U. Typhoidal Salmonella Trends in
353 Thailand. *Am J Trop Med Hyg*. 2018 Sep;99(3_Suppl):64–71.

- 354 10. Raskind-Hood C, Hogue C, Overwyk KJ, Book W. Estimates of adolescent and adult
355 congenital heart defect prevalence in metropolitan Atlanta, 2010, using capture-recapture
356 applied to administrative records. *Ann Epidemiol.* 2018 Dec 5.
- 357 11. Rood E, Khan A, Modak P, et al. A Spatial Analysis Framework to Monitor and Accelerate
358 Progress towards SDG 3 to End TB in Bangladesh. *ISPRS Int J Geo-Inf.* 2018;8(1):14.
- 359 12. Parker JD, Kruszon-Moran D, Mohadjer LK, et al. National Health and Nutrition
360 Examination Survey: California and Los Angeles County, Estimation Methods and
361 Analytic Considerations, 1999-2006 and 2007-2014. *Vital Health Stat 2.* 2017
362 May;(173):1–26.
- 363 13. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and
364 years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis
365 for the Global Burden of Disease Study 2015. *The Lancet.* 2016 Oct;388(10053):1545–602.
- 366 14. Avilov KK, Romanyukha AA, Borisov SE, Belilovsky EM, Nechaeva OB, Karkach AS. An
367 approach to estimating tuberculosis incidence and case detection rate from routine
368 notification data. *Int J Tuberc Lung Dis.* 2015 Mar 1;19(3):288–94.
- 369 15. Farzadfar F, Delavari A, Malekzadeh R, et al. NASBOD 2013: design, definitions, and
370 metrics. *Arch Iran Med.* 2014 Jan;17(1):7–15.
- 371 16. Begg S, Vos T, Barker B, Stevenson C, Stanley L, Lopez A. The burden of disease and
372 injury in Australia 2003. *Phe.* 2003;82:337.
- 373 17. World Bank. Indonesia country overview.
374 <http://www.worldbank.org/en/country/indonesia/overview>. Accessed December 2018.
- 375 18. BPS-Statistics Indonesia. Statistical Yearbook of Indonesia 2018. Sub-directorate of
376 Statistical Compilation and Publication, Editor.2018.
- 377 19. Government of Indonesia. Law No 22/1999 on Local Government (Undang-Undang
378 Republik Indonesia No 22 Tahun 1999 Tentang Pemerintahan Daerah). 1999.
- 379 20. Mahendradhata Y, Trisnantoro L, Listyadewi S, Soewondo P, Marthias T, Harimurti P. The
380 Republic of Indonesia Health System Review. 2017; 7: 64–105 p.
- 381 21. Indonesia Ministry of Health. Petunjuk Penyusunan Rencana Aksi Daerah untuk
382 Tuberkulosis. Jakarta, 2017.
- 383 22. Executive Board, 134. (2014). Global strategy and targets for tuberculosis prevention , care
384 and control after 2015 Report by the Secretariat.
385 <http://www.who.int/iris/handle/10665/172828>
- 386 23. Ministry of Health Indonesia. Indonesia Tuberculosis Prevalence Survey. 2015;(June).

- 387 24. BPS-Statistics Indonesia. Proyeksi Penduduk Kabupaten/Kota Provinsi Aceh 2010-2020.
388 Jakarta, 2015.
- 389 25. BPS-Statistics Indonesia. National Socio-Economic Survey 2017.
- 390 26. Brown T, Peerapatanapokin W. The Asian Epidemic Model: a process model for exploring
391 HIV policy and programme alternatives in Asia. *Sex Transm Infect.* 80(Suppl 1):i19–24.
- 392 27. Getahun H, Gunneberg C, Granich R, Nunn P. HIV Infection – Associated Tuberculosis :
393 The Epidemiology and the Response. 2010;50:201–7.
- 394 28. Rajaei E, Hadadi M, Madadi M, et al. Outdoor Air Pollution Affects Tuberculosis
395 Development Based on Geographical Information System Modeling. *Biomed Biotechnol*
396 *Res J.* 2018;2(1):39–45.
- 397 29. Liu Y, Cui LL, Hou LJ, et al. Ambient Air Pollution Exposures and Newly Diagnosed
398 Pulmonary Tuberculosis in Jinan, China: A Time Series Study. *Sci Rep.* 2018;8(1):1–11.
- 399 30. Kim J. Is ambient air pollution another risk factor of tuberculosis? *Korean J Intern Med.*
400 2014 Mar;29(2):170–2.
- 401 31. World Health Organization. Epidemiological Review in Indonesia. Jakarta; 2017.
- 402 32. Menzies NA, McQuaid CF, Gomez GB, Siroka A, Glaziou P, Floyd K, et al. Improving the
403 quality of modelling evidence used for tuberculosis policy evaluation. *Int J Tuberc Lung*
404 *Dis.* 2019 Apr 1;23(4):387–95.
- 405 33. Kementerian Kesehatan Republik Indonesia. Panduan Penentuan Beban dan Target Cakupan
406 Penemuan dan Pengobatan Tuberkulosis di Indonesia Tahun 2019-2024. Jakarta; 2018.
- 407

408 **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

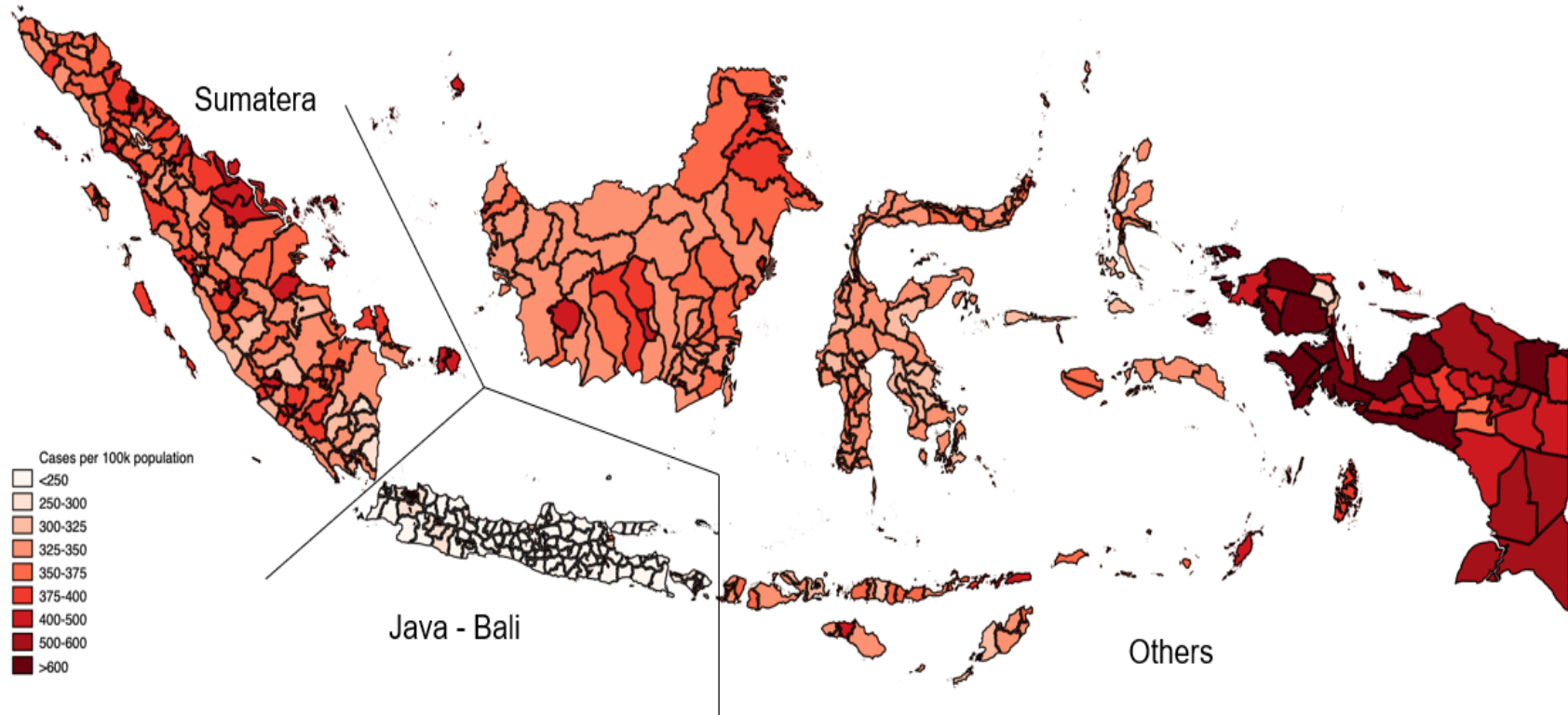
409

410 **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 8m²/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

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

The estimates of TB incidence rate per 100k population at district level based on the model, 2017

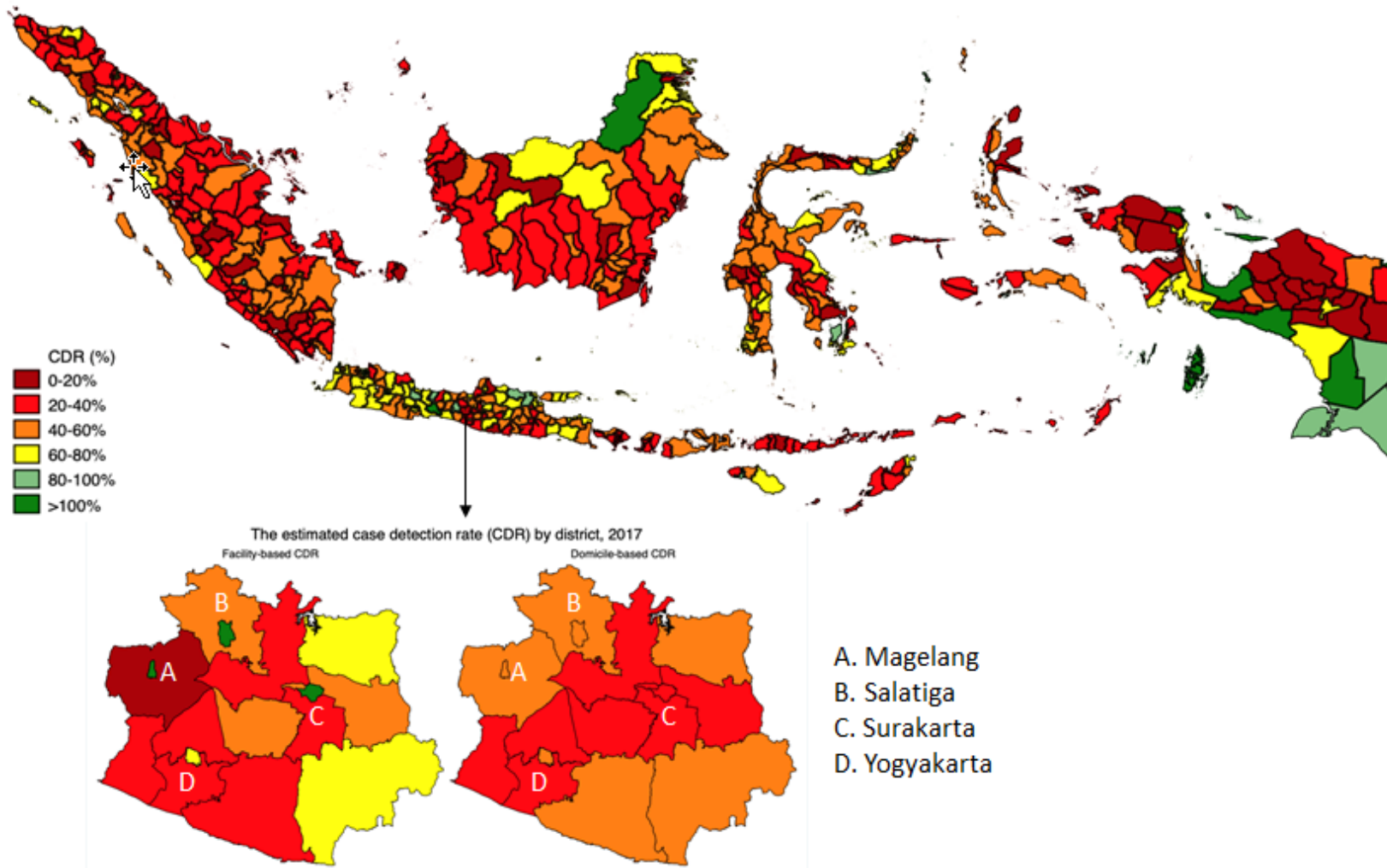


413

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

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

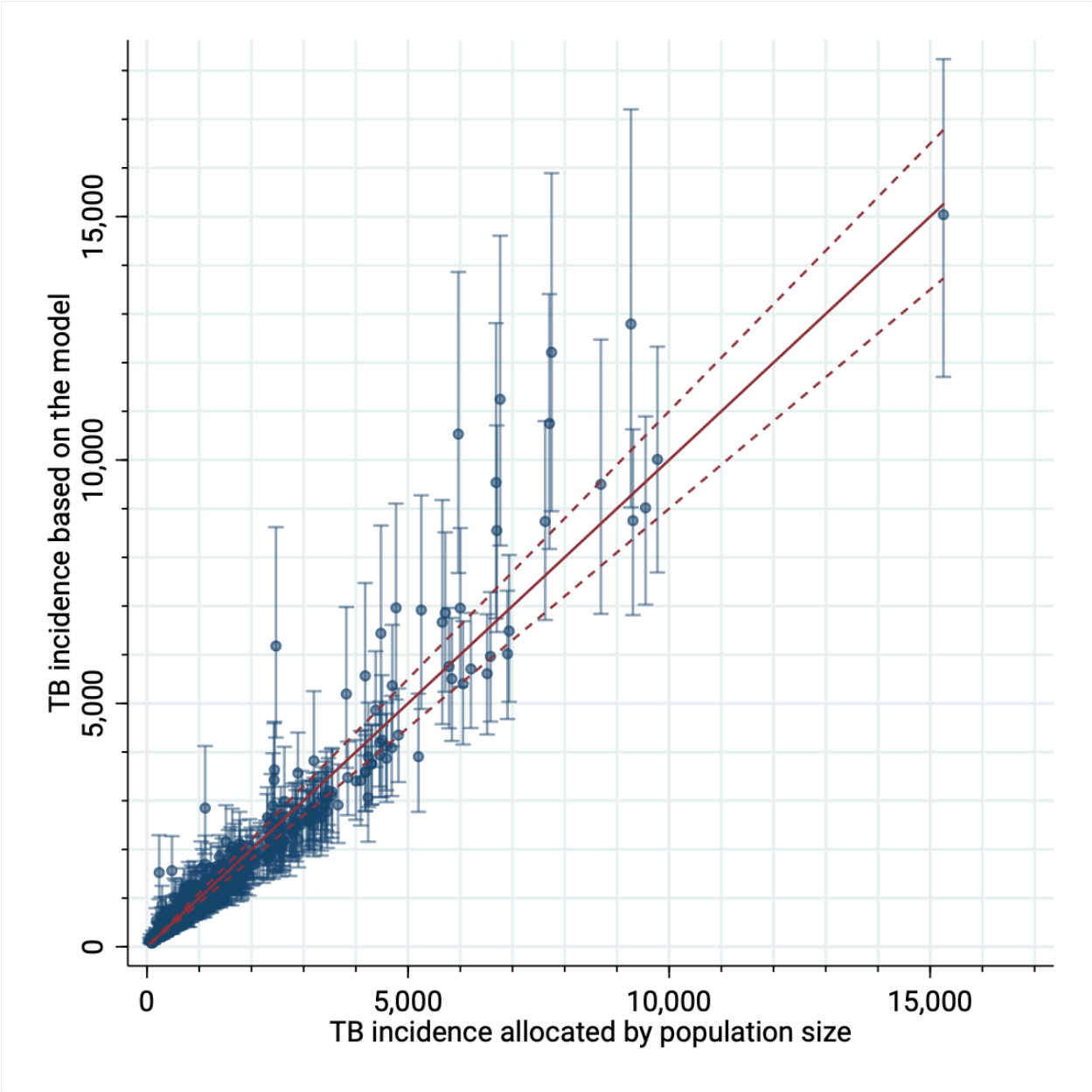
The estimated facility-based case detection rate (CDR) by district, 2017



416

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

418 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
419 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
420 cross-district health system utilisation.



421

422 **Fig 3. Heterogeneity captured by model variables.**

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