1 **The burden of tuberculosis in Ho Chi Minh City, Vietnam: A spatial analysis of**

2 **drug-susceptible and multi-drug resistant cases between 2020 and 2023**

3 **Authors: Ruan Spies[1], Hanh Nguyen Hong[2], Phu Phan Trieu[2], Luong Kim Lan[2], Kim Lan[2], Ngo**

4 **Ngoc Hue[2], Nguyen Thi Le Huong[2], Tran Le Thi Ngoc Thao[2], Nguyen Le Quang[2], Thu Do Dang**

5 **Anh[2], Truong Van Vinh[3], Dang Thi Minh Ha[3], Phan Thuong Dat[3], Nguyen Phuc Hai[3], Le Hong**

6 **Van[2], Guy E Thwaites[2,4], Nguyen Thuy Thuong Thuong[2,4], James A Watson[2,4*], Timothy M**

7 **Walker[2,4*]**

8 ***Contributed equally**

9 **Affiliations:**

10 [1] Nuffield Department of Population Health, University of Oxford, Oxford, UK

11 [2] Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

12 [3] Pham Ngoc Thach Hospital, Ho Chi Minh City, Vietnam

13 [4] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine,

14 University of Oxford, Oxford, UK

15 **Corresponding author: Dr Ruan Spies, Brasenose College, Radcliffe Square, Oxford, OX1**

16 **4AJ, email: ruan.spies@ndm.ox.ac.uk, phone: +447542981035**

17 **Article summary line:** Drug-susceptible and multi-drug resistant tuberculosis in Ho Chi Minh

18 City Vietnam exhibit substantial spatial heterogeneity with overlapping hot spots of disease

19 incidence concentrated in the centre of the city.

20   **Running title:** TB spatial heterogeneity Ho Chi Minh City

23   **Author bio**

24   Dr Ruan Spies is a South African medical doctor and PhD candidate at the Nuffield

25   Department of Medicine at the University of Oxford and the Oxford University Clinical

26   Research Unit (OUCRU) in Ho Chi Minh City, Vietnam. His doctoral work investigates the

27   molecular epidemiology of multi-drug resistant tuberculosis. His main research interests

28   include tuberculosis epidemiology.

29

30   Abstract

31   We characterised the spatial distribution of drug-susceptible (DS) and multi-drug resistant

32   (MDR) tuberculosis (TB) cases in Ho Chi Minh City (HCMC), a major South-East Asian

33   metropolis, and explored demographic and socioeconomic factors associated with local TB

34   burden. Hot spots of DS- and MDR-TB incidence were observed in the central parts of

35   HCMC, with substantial heterogeneity observed across wards. Positive spatial

36   autocorrelation was observed for both DS- and MDR-TB. Ward-level TB incidence was

37   associated with HIV prevalence (incidence rate ratio [IRR] 1.77, 95% CI 1.54-2.03) and the

38   male proportion of the population (IRR 1.05, 95% CI 1.02-1.08). No ward-level demographic

39   and socioeconomic indicators were associated with MDR-TB case count relative to total TB

40   case count. Our findings may inform spatially-targeted TB control strategies and provide

2

41   insights for generating hypotheses about the nature of the relationship between DS- and

42   MDR-TB in HCMC, Vietnam and the wider South-East Asia region.

43   **Word count:** 147

Introduction

44

45 Tuberculosis (TB) causes more deaths worldwide than any other pathogen. Progress in

46 reducing the global burden of TB stalled during the COVID-19 pandemic and an estimated

47 10.6 million people fell ill from TB in 2021, with 1.6 million dying[1]. The number of people

48 with multi-drug resistant TB (MDR-TB), defined by resistance to rifampicin and isoniazid, is

49 estimated to have grown by 3.1% from 2020[1], with an estimated 450,000 incident cases in

50 2021. MDR-TB remains underdiagnosed and is associated with worse treatment outcomes

51 than drug-susceptible TB (DS-TB) [1,2].

52 TB is not only spatially heterogeneous globally, but also locally. Just 30 low-and-middle-

53 income countries (LMICs) account for nearly 90% of the global burden of disease[1], but an

54 unequal distribution of disease has been described more locally too [3-12]. Although poorly

55 understood, the drivers of geographical heterogeneity in TB are thought to reflect the

56 complex interplay between the infectious and susceptible host, the infecting organism, the

57 physical environment and distal determinants like poverty[13].

58 The World Health Organisation (WHO) recognises Vietnam as a high-burden country for TB

59 and MDR-TB, with an estimated incidence of 173 (95% CI 112-247) and 9.1 (95% CI 5.5-13)

60 per 100,000 population, respectively [1,14]. The highest incidence is seen in the southern

61 parts of the country, especially in Ho Chi Minh City (HCMC)[15,16]. As anywhere else,

62 patients with MDR-TB in HCMC can have acquired their disease in one of two ways: through

63 selection of drug resistance mutations whilst receiving first-line TB drug treatment, or

64 directly from others through transmission[17]. Comparison of the spatial distributions of DS-

4

65   and MDR-TB across this high-incidence city has the potential to offer insights into the

66   relative contributions of each to MDR-TB burden. For example, the observation of distinct

67   spatial distributions of DS- and MDR-TB may support the hypothesis that MDR-TB is

68   transmitted in networks independent from circulating DS-TB. Alternatively, sporadic MDR-TB

69   cases among clusters of DS-TB cases may be more indicative of *de novo* emergence of MDR-

70   TB through inadequate treatment and selection. Better understanding hyper-local patterns

71   of disease may also contribute to spatially-targeted interventions, such as active case finding

72   and healthcare facility planning (18-21), and to the design of and recruitment into clinical

73   trials and other studies.

74   In this study we aimed to characterise the spatial distribution of DS- and MDR-TB in HCMC

75   and to explore demographic and socioeconomic factors associated with local TB burden.

Methods

Study setting

78   HCMC is the largest metropolis in Vietnam and one of the largest in South-East Asia with a

79   total population approximating 10 million people. It is subdivided into 24 districts, 19 urban

80   and 5 rural **(Figure S1)**, of which 3 were combined to form a municipal city, Thủ Đức City, in

81   2021. HCMC's districts are further subdivided into 322 administrative sub-units consisting of

82   wards, townlets and communes with a median population of approximately 22,000 people

83   (hereafter collectively referred to as wards). This study includes data prior to 2021 and

84   therefore references the former, 24-district sub-division of HCMC.

85 Public sector community-based TB care in HCMC is coordinated through 24 district

86 treatment units (DTUs) where individuals with suspected TB are referred for testing and

87 treatment. Once diagnosed with TB, patients are registered with the National TB Program

88 (NTP). All individuals diagnosed with MDR-TB in the public sector initiate treatment through

89 the city's lung hospital, Phạm Ngọc Thạch (PNTH), and then continue outpatient care

90 through the DTUs. PNTH is the regional centre for MDR-TB treatment in Southern Vietnam

91 and provides treatment to approximately 80% of all MDR-TB cases in Vietnam(22).

Study population

93 The study population included all individuals who registered for TB treatment in the public

94 sector in 23 of HCMC's districts from 1 January 2020 to 30 April 2023. The study excludes TB

95 cases from Cần Giờ, a rural district comprised of 7 wards with a population of 71,527 people

96 (0.8% of HCMC's population)(23), as data were not available for this district. For the

97 ecological analysis, 315 residential wards constituting 23 of HCMC's districts formed the

98 units of analysis.

Data sources

100 Data for participants with DS-TB were accessed from the Vietnam TB Information

101 Management Electronic System (VITIMES), a web-based surveillance system which records

102 TB notifications and treatment outcomes for the NTP(24). VITIMES includes data on all

103 people in HCMC initiated on first-line TB therapy in the public sector. At treatment initiation,

104 patient details are added to a paper-based register which is electronically transcribed by

105 DTU staff at monthly intervals. Data extracted from the electronic register for this study

106   included participant age, sex, home address, HIV status and history of previous TB. Data for

107   participants with MDR-TB were obtained from an ongoing cohort study conducted through

108   the Oxford University Clinical Research Unit (OUCRU). Participants included all individuals

109   initiating treatment for MDR-TB at PNTH. The OUCRU cohort study database was selected

110   over the NTP-based register as the data source for MDR-TB cases as it provided identical

111   case coverage to the NTP-based register, with less missing data.

112   District and ward-level demographic and socioeconomic indicators were obtained from

113   published regional data collected as part of the 2019 Vietnam census(23). Extracted

114   indicators which were available at only the district level included: population age structure,

115   unemployment rate, the proportion of households that owned a computer and the number

116   of people living with HIV. All wards within a district were assigned the district value for

117   indicators available only at the district level. For example, District 1 had an HIV prevalence of

118   1.5% and this value for HIV prevalence was subsequently assigned to each of District 1's

119   constituent wards. Extracted indicators which were available at the ward level included total

120   population, population by sex, population density, average number of people per household,

121   literacy rate and residence type (urban or rural). Location was labelled as 'city centre' if

122   wards were located in HCMC's central commercial, commuting and socialising hubs and

123   labelled as 'peripheral' if wards were located outside these areas **(Supplementary methods)**.

124   ## Design and analysis

125   Individual-level data were used for a descriptive, cross-sectional analysis of the burden of TB

126   in HCMC and the characteristics of TB cases. An ecological design, using ward-level data, was

7

127   used to describe ward-level factors associated with TB burden. The outcomes for the

128   ecological analysis were a) total TB incidence and b) burden of MDR-TB relative to total TB.

Descriptive analysis

130   Participant characteristics were summarised with mean and standard deviation (SD) for

131   continuous variables and as counts and proportions for categorical variables. Participant

132   home addresses were deidentified and converted to latitude and longitude coordinates

133   using the Google geocoding service and the "tidygeocoder" package in R(25). Spatial

134   polygons for the administrative units of HCMC were obtained from the Database of Global

135   Administrative Areas(26). Individual TB cases were mapped and aggregated by ward, and

136   average annual incidence of DS- and MDR-TB calculated for each.

Spatial autocorrelation

138   The presence, strength and direction of spatial autocorrelation over the entire study area

139   was assessed separately for DS- and MDR-TB incidence through the calculation of the global

140   Moran's I statistic. Local spatial autocorrelation in these parameters was assessed through

141   the calculation of the Getis-Ord Gi* statistic and Anselin Local Moran's I. The Getis-Ord Gi*

142   statistic was used to define spatial hot spots and cold spots relative to the null hypothesis of

143   spatial randomness over the entire study area. In this analysis, each ward was considered in

144   the context of its neighbouring wards, forming a neighbourhood. The local sum of the values

145   for the given parameter (e.g., DS-TB incidence) for each of the wards in a neighbourhood

146   was then compared proportionally to the sum of the parameter values for all the wards in

147   the study area. Neighbourhoods with statistically significant higher parameter values than

148 the entire study area were designated hot spots and neighbourhoods with statistically

149 significant lower parameter values than the entire study area were designated cold

150 spots(27). The analysis using Anselin Local Moran's I value further compared each ward to its

151 neighbourhood. Wards with high parameter values within neighbourhoods with high values

152 were designated high-high clusters, wards with high values within neighbourhoods with low

153 values were designated high-low outliers, wards with low values within neighbourhoods

154 with low values were designated low-low clusters and wards with low values within

155 neighbourhoods with high values were designated low-high outliers (28). False Discovery

156 Rate Correction for multiple testing and spatial dependency was applied to both local spatial

157 autocorrelation analyses.

158 Ecological analysis

159 Continuous ward-level indicators were summarised with mean and SD or median and

160 interquartile range (IQR) depending on skew. Categorical indicators were summarised as

161 counts and proportions. Exploratory analyses evaluated the relationship between ward-level

162 demographic and socioeconomic indicators and a) total TB incidence and b) MDR-TB case

163 count relative to total TB case count. Univariate associations between ward-level indicators

164 and the natural logarithm of total TB incidence were assessed through the inspection of

165 scatter plots and the calculation of Spearman's rho for continuous indicators and by the

166 Wilcoxon rank-sum test and ANOVA for categorical indicators. Continuous indicators with

167 non-linear associations with the outcome were categorised into tertiles. Indicators

168 associated with total TB incidence with p-value <0.05 were included in a multivariable

169 negative binomial regression model for each outcome. Ward-level TB incidence was

170 modelled by including ward-level TB case count as the dependent variable with an offset

171 term for ward population. Ward-level MDR-TB case count as a proportion of all TB cases was

172 modelled using MDR-TB case count as the dependent variable with an offset term for total

173 TB case count. Visualisation of spatial autocorrelation in the residuals for each negative

174 binomial regression model (measured using Moran's I) demonstrated positive spatial

175 autocorrelation in the residuals for both models, violating the assumption of independence.

176 To account for this, a spatially autocorrelated random effects term (using the centroid of

177 each ward as latitude and longitude) assuming a Matérn covariance structure was added to

178 each model. Additional assumptions including the absence of multicollinearity and

179 inequality in outcome means and variances were also assessed. Model fit, for the mixed-

180 effects models and standard models, was compared using Akaike information criterion (AIC)

181 and scatter plots of the observed versus fitted values.

182 A sensitivity analysis estimated the association between ward-level demographic and

183 socioeconomic indicators and both outcomes using conditional autoregressive modelling. In

184 contrast to the main analysis, in which spatial information was formatted as point data (i.e.,

185 latitude and longitude coordinates for the centroid of each ward), in the sensitivity analysis

186 spatial information was reformatted as areal data, with each ward represented by a spatial

187 polygon surrounded by an administrative boundary. Ward neighbours were defined by

188 contiguity in administrative boundaries and neighbourhood lists were converted to an

189 adjacency matrix using binary weights to signify the presence (1) or absence (0) of a

190 neighbour. The adjacency matrix was incorporated into the negative binomial regression

191 model as a random effects term in order to account for spatial autocorrelation between

192 neighbouring wards.

10

193   All statistical analyses were conducted with R Studio(29). The calculation of spatial statistics

194   and mapping was conducted with ArcGIS Online(30).

Ethical considerations

196   This study was approved by the Oxford Tropical Research Ethics Committee (reference

197   number 51-19) and the Institutional Review Board at Phạm Ngọc Thạch Hospital (643/PNT-

198   HDDD).

Results

Descriptive analysis

201   Between 1 January 2020 and 30 April 2023, 36,089 people registered for DS-TB treatment

202   and 1,451 people registered for MDR-TB treatment in HCMC. 49 participants with DS-TB

203   (0.1%) and 12 participants with MDR-TB (0.8%) provided residential addresses outside of

204   HCMC and were excluded from the spatial analysis. Most participants were male (N=25,463

205   [68%]), 30,268 (81%) were urban dwelling and the mean age was 45 (SD 16.5) **(Table 1)**. HIV

206   co-infection was present in 5% of all participants (N=1,692), with this proportion similar for

207   both DS- and MDR-TB groups. Previous TB infection was reported by 4,721 (13%) of the

208   participants treated for DS-TB and 795 (55%) of the participants treated for MDR-TB,

209   although it is unknown how many of these previous infections were due to drug-resistant

210   TB. Among 31,999 cases with no history of TB previously, 640 (2%) were registered for MDR-

211   TB treatment while 772 (14%) cases of the 5,516 cases with a history of previous TB were

212   registered for MDR-TB treatment**.** Asymmetric population pyramids demonstrated a greater

11

213   DS- and MDR-TB burden among men in middle to late-middle age, although the sex

214   distributions were more symmetrical below age 40 **(Figure 1)**. The average annual incidence

215   of notified DS-TB and MDR-TB in HCMC during this period was 121.4 (95% CI 119.1-123.7)

216   and 4.8 (95% CI 4.4-5.4) per 100,000, respectively.

217

12

Table 1: Characteristics of individuals registered for TB treatment in Ho Chi Minh City from 1 January 2020 to 30 April 2023, stratified by TB type.

Characteristic	Drug-susceptible TB (N=36089)	Multi-drug resistant TB (N=1451)	Overall (N=37540)
Age, mean (SD)	44.9 (16.6)	45.7 (14.1)	44.9 (16.5)
Sex, n (%)			
Female	11,732 (32.5%)	385 (26.5%)	12,117 (32.3%)
Male	24,357 (67.5%)	1,066 (73.5%)	25,423 (67.7%)
HIV status, n (%)			
Negative	27,745 (76.9%)	1,333 (91.9%)	29,078 (77.5%)
Positive	1,609 (4.5%)	83 (5.7%)	1,692 (4.5%)
Unknown	6,735 (18.7%)	35 (2.4%)	6,770 (18.0%)
Previous TB history, n (%)			
No	31,344 (86.9%)	655 (45.1%)	31,999 (85.2%)
Yes	4,721 (13.1%)	795 (54.8%)	5,516 (14.7%)
Unknown	24 (0.1%)	1 (0.1%)	25 (0.1%)
Residence type, n (%)			
Urban	29,086 (80.6%)	1,182 (81.5%)	30,268 (80.6%)
Rural	6,948 (19.3%)	246 (17.0%)	7,194 (19.2%)

218   *HIV – human immunodeficiency virus, SD – standard deviation, TB – tuberculosis.*

219   **Figure 1: Population pyramids demonstrating age and sex distributions of participants**

220   **registered for drug-susceptible and multi-drug resistant tuberculosis treatment in Ho Chi**

221   **Minh City, 1 January 2020 to 30 April 2023.** DS TB – drug-susceptible tuberculosis, MDR TB –

222   multi-drug resistant tuberculosis.

223   Substantial spatial heterogeneity in both DS- and MDR-TB average annual incidence was

224   observed across HCMC wards **(Figures 2 and 3)**. DS-TB incidence (per 100,000) ranged from

13

225 26.7 in Bình Lợi (District Bình Chánh) to 1,345.3 in An Khánh (District 2). Thirty-two wards

226 recorded 0 MDR-TB cases during the study period while Ward 8 (District 11) demonstrated

227 an MDR-TB incidence of 31.7 per 100,000. In the overall study population, 3.9% (3.7-4.1) of

228 all TB cases were treated for MDR-TB.

229 **Figure 2: Choropleth map displaying the geographic variation in average annual incidence**

230 **per 100,000 for drug-susceptible tuberculosis, subdivided by ward, from 1 January 2020 to**

231 **30 April 2023**. Map does not include Cần Giờ district. TB – tuberculosis.

232 **Figure 3: Choropleth map displaying the geographic variation in average annual incidence**

233 **per 100,000 for multi-drug resistant tuberculosis, subdivided by ward, from 1 January 2020**

234 **to 30 April 2023.** Map does not include Cần Giờ district. TB - tuberculosis.

Spatial autocorrelation

236 The global Moran's I statistic was 0.14 (p <0.001) for both DS-TB incidence and MDR-TB

237 incidence – demonstrating weak positive global spatial autocorrelation for each parameter.

238 This demonstrated that, over the entire study area, wards with similar values for the above

239 parameters (e.g., similar DS-TB incidences) were located closer to each other than would be

240 expected if the wards were randomly arranged (i.e., there was evidence of some spatial

241 clustering for each parameter). The global Moran's I, however, provided no information

242 about where these clustered wards were located or how the clustering of DS-TB related to

243 the clustering of MDR-TB. **Figure 4** presents the results of the hot spot analysis, using the

244 Getis-Ord Gi* statistic. Hot spots were evident in the central parts of HCMC for both DS and

245 MDR-TB. Cold spots were observed to the north of the city centre. Like the hot spots, the

14

246   DS- and MDR-TB cold spots largely overlapped spatially. **Figure 5** uses Anselin Local Moran's I

247   to demonstrate wards in which TB incidence was congruent with the surrounding

248   neighbourhood (clusters) and wards in which TB incidence contrasted the surrounding

249   neighbourhood (outliers). Heterogeneity in incidence, for both DS- and MDR-TB, was evident

250   even within hot spots and cold spots. Indeed, for DS-TB, most of the wards in the city centre

251   hot spot, when considered separately from their neighbourhood, were low-high outliers

252   (wards situated within hot spot neighbourhoods but with low incidences relative to the

253   overall study area). A greater number of the wards that constituted the MDR-TB hot spot

254   were high-high clusters (incidence values congruent with those of the surrounding

255   neighbourhood), indicating more homogeneity within the MDR-TB hot spots.

256   **Figure 4: Spatial clustering of drug-susceptible TB incidence and multi-drug resistant TB**

257   **incidence in Ho Chi Minh City from 1 January 2020 to 30 April 2023, based on the Getis-**

258   **Ord GI* statistic.** DS-TB – drug-susceptible tuberculosis, MDR-TB – multi-drug resistant

259   tuberculosis.

260   **Figure 5: Spatial clusters and outliers of drug-susceptible TB incidence and multi-drug**

261   **resistant TB incidence in Ho Chi Minh City from 1 January 2020 to 30 April 2023, based on**

262   **the Anselin Local Moran's I statistic.** DS-TB – drug-susceptible tuberculosis, MDR-TB – multi-

263   drug resistant tuberculosis.

264   ## Ward-level factors associated with TB burden

265   Surprisingly, wards in the highest tertile of TB incidence had the lowest male proportion of

266   the population (47.6%) although the range of male proportion of the population between

15

267   wards in the highest and lowest tertiles of TB incidence was small (47.6-48.1%). Literacy rate

268   (98.8%), the proportion of homes that owned a computer (65.2%), but also the lowest

269   unemployment rate (2.8%) were also lowest in these wards **(Table 2)**. These wards had the

270   highest proportion of the population aged 30 to 59 (45.5%), population density (32,117

271   people/km^2), number of people per household (3.6 people) and HIV prevalence (0.9%).

272   Indicators significantly associated with TB incidence in the univariate analyses, and

273   subsequently included in the final multivariable models, included the male proportion of the

274   population, proportion of the population aged 30 to 59, average number of people per

275   household, literacy rate, unemployment rate and HIV prevalence **(Supplementary Results)**.

276   In a multivariable negative binomial regression model with mixed-effects, in contrast to the

277   unadjusted association, the male proportion of the population was significantly associated

278   with total TB incidence (incidence rate ratio [IRR] 1.05, 95% CI 1.02-1.08) while each

279   percentage increase in HIV prevalence was associated with a 77% increase in TB incidence

280   (IRR 1.77, 95% CI 1.54-2.03) **(Table 3)**. None of the selected indicators were significantly

281   associated with MDR-TB case counts relative to total TB case counts. The mixed-effects

282   models including spatially autocorrelated random effects terms demonstrated better fit than

283   the standard models and estimates from the sensitivity analysis were similar to the main

284   analysis **(Supplementary Results)**.

Table 2: Ho Chi Minh City ward-level demographic and socioeconomic indicators stratified by tertiles of overall tuberculosis incidence

	Overall TB incidence			
	1st tertile Median incidence = 84/100,000 (N=105)	**2nd tertile** Median incidence = 120/100,000 (N=105)	**3rd tertile** Median incidence = 187/100,000 (N=105)	**Overall** (N=315)
Male proportion of population, mean (SD)	48.1 (1.86)	48.2 (1.82)	47.6 (2.78)	48.0 (2.21)
Proportion of population aged 30 to 59, mean (SD)	44.1 (2.23)	44.2 (2.02)	45.5 (1.52)	44.6 (2.04)
Residence type, n (%)				
Urban	86 (81.9%)	89 (84.8%)	88 (83.8%)	263 (83.5%)
Rural	19 (18.1%)	16 (15.2%)	17 (16.2%)	52 (16.5%)
Location, n (%)				
City centre	23 (21.9%)	22 (21%)	24 (22.9%)	69 (22%)
Peripheral	82 (78.1%)	83 (79%)	81 (77.1%)	246 (78%)
Total population	26,050 (12,402-40,289)	25,575 (13,354-42,067)	16,911 (11,190-25,068)	22,383 (12,397-36,880)
Population density (people/km^2), median (IQR)	27,537 (9,203-44,241)	20,810 (6,323-41,535)	32,117 (13,005-46,854)	27,781(8,233-44,812)
Average number of people per household, mean (SD)	3.51 (0.319)	3.55 (0.269)	3.62 (0.321)	3.56 (0.307)
Literacy rate, median (IQR)	99.3 (98.7-99.6)	99.3 (98.7-99.6)	98.8 (97.7-99.3)	99.2 (98.5-99.6)
Unemployment rate, mean (SD)	3.25 (1.17)	3.10 (1.05)	2.76 (0.747)	3.04 (1.02)
Proportion of homes that own a computer, median (IQR)	71.7 (55.4-77.6)	71.0 (59.3-76.5)	65.2 (59.3-73.2)	71.0 (59.3-76.3)
HIV prevalence, median (IQR)	0.49 (0.34-0.84)	0.48 (0.34-0.93)	0.93 (0.45-1.29)	0.49 (0.36-0.95)

Table 2: Ho Chi Minh City ward-level demographic and socioeconomic indicators stratified by tertiles of overall tuberculosis incidence

	Overall TB incidence			
	1st tertile Median incidence = 84/100,000 (N=105)	2nd tertile Median incidence = 120/100,000 (N=105)	3rd tertile Median incidence = 187/100,000 (N=105)	Overall (N=315)

HIV – human immunodeficiency virus, IQR – interquartile range ,km – kilometre, SD – standard deviation, TB - tuberculosis

285

18

Table 3: Adjusted incidence rate ratios and 95% confidence intervals for the association between ward-level indicators and a) total TB incidence and b) MDR-TB case count relative to total TB case count

Indicator	Total TB incidence	MDR-TB case count (relative to total TB case count)
Male proportion of population (%)	1.05 (1.02-1.08)	0.99 (0.94-1.05)
Proportion of population aged 30-59 (%)	1.02 (0.99-1.04)	1.03 (0.98-1.08)
Average number of people per household	1.13 (0.98-1.31)	0.97 (0.76-1.25)
Literacy rate		
1st tertile	1 (reference)	1 (reference)
2nd tertile	1.06 (0.96-1.18)	1.05 (0.89-1.26)
3rd tertile	0.96 (0.86-1.07)	1.01 (0.83-1.25)
Unemployment rate (%)	0.96 (0.92-1.00)	0.95 (0.87-1.03)
HIV prevalence (%)	1.77 (1.54-2.03)	1.08 (0.85-1.38)

286   *HIV – human immunodeficiency virus, MDR-TB – multi-drug resistant tuberculosis, TB - tuberculosis*

Discussion

288 This study characterises the burden of TB in HCMC, with granular, ward-level descriptions of

289 DS- and MDR-TB burden. Both DS- and MDR-TB were heterogeneously distributed

290 throughout HCMC, forming geographic clusters of high incidence, predominantly

291 concentrated in the city's centre. Total TB incidence at the ward level was strongly

292 associated with HIV prevalence and more weakly associated with the proportion of the

293 population that is male.

294 The asymmetric age and sex distributions among TB cases in HCMC we describe are

295 consistent with the findings from the second Vietnamese national TB prevalence survey, in

296 which the prevalence of bacteriologically confirmed TB was 4 times greater in males than

297 females and increased with age[31]. Studies from Vietnam have also demonstrated a greater

298 prevalence of latent TB in men compared to women[32]. The magnitude of this difference in

299 prevalence by sex is smaller in latent TB than for active TB, however, emphasising the

300 importance of sex differences in the risk factors for disease progression. A recent sub-study

301 from the national TB prevalence survey specifically noted the stark differences in the

302 prevalence of smoking (45% of males versus 1% of females)[33] and drinking (44% of males

303 versus 1% of females)[34] in Vietnam as likely contributors to observed differences in the

304 prevalence of active TB by sex[35]. Sex differences, for both latent and active disease,

305 remain incompletely understood but likely reflect the complex interplay between biological,

306 behavioural and environmental factors[36]. We demonstrated a 5% greater TB incidence per

307 percentage increase in the proportion of the population that is male, suggesting sex-specific

308 differences in risk may manifest at the population level.-

20

309   We observed a 5% prevalence of TB-HIV co-infection, approximating previous regionally-

310   representative estimates (14,37). We also observed that TB incidence was substantially

311   greater with each percentage increase in HIV prevalence, emphasising the potential

312   contribution of HIV to the TB epidemic, even in settings with relatively low HIV prevalence.

313   Our incidence estimates for both DS- and MDR-TB, derived from TB notifications, are

314   markedly lower than the WHO's estimates for Vietnam (TB and MDR-TB incidence 173 and

315   9.1 per 100,000, respectively)(14), despite evidence that HCMC has some of the highest TB

316   incidences in the country(15,16). The WHO estimates are derived from multiple data sources

317   including prevalence surveys, case notification data, expert opinion about case detection

318   gaps, and dynamic modelling(38). The differences between our incidence estimates likely

319   reflect a limitation of this study: the 'diagnostic gap' – the difference between the true

320   number of people who fell ill with TB and the number of people who were registered for TB

321   treatment(1). The diagnostic gap is a well-described barrier to TB control in Vietnam and has

322   recently been exacerbated by COVID-19-related health system disruptions, with less than

323   50% of predicted TB cases enrolled on treatment in 2021(1).

324   Interestingly, TB incidence in HCMC was not associated with measures of poverty (literacy

325   rates, unemployment rates and the proportion of homes owning a computer – a proxy for

326   material wealth) despite this being a well-established risk factor(39). The central

327   concentration of the burden of TB in HCMC was instead, in our data, related to factors such

328   as sex distribution and HIV prevalence. This lack of association may reflect the poor

329   representation of poverty and social deprivation by the variables included in our analysis

330   (i.e., literacy rates are high across HCMC, even in poorer and rural areas[23]). It may also be

21

331   that rapid 'equitable', economic growth in Vietnam, coupled with a reduction in TB

332   prevalence over the last 20 years, contributed to a reduction of the concentration of TB

333   among poor households(40).

334   Our spatial analysis demonstrated substantial overlap in geographic clusters of DS- and

335   MDR-TB incidence, raising interesting questions about the relationship between DS- and

336   MDR-TB burden. These findings may be consistent with the hypothesis that drug resistance

337   largely emerges from DS-TB *de novo*, with the distributions of DS-and MDR-TB therefore

338   related. Alternatively, the overlapping distributions may also be consistent with the

339   hypothesis that most MDR-TB is transmitted and that factors associated with the

340   transmission of TB in general are geographically clustered. The lack of association between

341   any demographic and socioeconomic indicators and MDR-TB burden relative to total TB

342   burden we describe potentially supports the latter hypothesis. Ultimately, it is likely that

343   both *de novo* and transmitted resistance contribute to MDR-TB burden. The enrichment of

344   spatial data with genetic data will better demonstrate the relative contributions of each

345   mechanism(41).

346   This study has several limitations. We utilised public sector registry data to identify TB cases

347   and therefore excluded people with undiagnosed TB, potentially biasing our sample

348   selection towards groups who are more likely to present when symptomatic. Furthermore,

349   we had no data on private-sector TB diagnoses – estimated to represent 8% of all TB cases in

350   HCMC(42). Participants in our study were only geolocated through their home addresses,

351   however, several studies have demonstrated the importance of transmission outside of the

352   home with the emergence of genetic data demonstrating geographically unrelated, cryptic

22

353 transmission networks mediated by mobility-linked locations in high-burden settings(43,44).

354 Future work on the transmission of TB in HCMC will benefit from whole genome sequencing-

355 derived genetic data being generated by a parallel, related study. The degree to which our

356 findings are relevant to other settings is uncertain, but it is likely the dynamics in HCMC are

357 not markedly different from other major cities with similar economic metrics in South-East

358 Asia where nearly half the world's TB patients reside(1).

359 Our study nevertheless characterises the demographic profile of people with DS- and MDR-

360 TB in HCMC and maps parts of the city most impacted. Our findings provide a starting point

361 for deeper research into TB acquisition and transmission dynamics and spatially-informed TB

362 control interventions in HCMC, Vietnam and the greater South-East Asia region.

Acknowledgements

Funding

23

References

1. World Health Organisation. *Global tuberculosis report 2022*.

   https://www.who.int/publications-detail-redirect/9789240061729.

2. World Health Organisation. WHO consolidated guidelines on tuberculosis. Module 4:

   Treatment. Drug-resistant tuberculosis treatment. https://www.who.int/publications-

   detail-redirect/9789240007048.

3. Jenkins, H. E. *et al*. Assessing spatial heterogeneity of multidrug-resistant tuberculosis in a

   high-burden country. *Eur. Respir. J.* **42**, 1291–1301 (2013).

4. Oliveira, O. *et al*. Using Bayesian spatial models to map and to identify geographical

   hotspots of multidrug-resistant tuberculosis in Portugal between 2000 and 2016. *Sci.*

   Rep. **10**, 16646 (2020).

5. Aturinde, A., Farnaghi, M., Pilesjö, P. & Mansourian, A. Spatial analysis of HIV-TB co-

   clustering in Uganda. *BMC Infect. Dis.* **19**, 612 (2019).

6. Ding, P., Li, X., Jia, Z. & Lu, Z. Multidrug-resistant tuberculosis (MDR-TB) disease burden in

   China: a systematic review and spatio-temporal analysis. *BMC Infect. Dis.* **17**, 57 (2017).

7. Alene, K. A., Viney, K., McBryde, E. S. & Clements, A. C. A. Spatial patterns of multidrug

   resistant tuberculosis and relationships to socio-economic, demographic and household

   factors in northwest Ethiopia. *PLOS ONE* **12**, e0171800 (2017).

392   8. Mashamba, M. A., Tanser, F., Afagbedzi, S. & Beke, A. Multi-drug-resistant tuberculosis

393   clusters in Mpumalanga province, South Africa, 2013–2016: A spatial analysis. *Trop.*

394   *Med. Int. Health* **27**, 185–191 (2022).

395   9. Sy, K. T. L. *et al.* Spatial heterogeneity of extensively drug resistant-tuberculosis in Western

396   Cape Province, South Africa. *Sci. Rep.* **12**, 10844 (2022).

397   10. Chirenda, J. *et al.* Spatial distribution of Mycobacterium Tuberculosis in metropolitan

398   Harare, Zimbabwe. *PLoS ONE* **15**, e0231637 (2020).

399   11. Lima, S. V. M. A. *et al.* Spatial and temporal analysis of tuberculosis in an area of social

400   inequality in Northeast Brazil. *BMC Public Health* **19**, 873 (2019).

401   12. Zaidi, S. M. A. *et al.* A spatial analysis of TB cases and abnormal X-rays detected through

402   active case-finding in Karachi, Pakistan. *Sci. Rep.* **13**, 1336 (2023).

403   13. Trauer, J. M. *et al.* The Importance of Heterogeneity to the Epidemiology of Tuberculosis.

404   *Clin. Infect. Dis.* **69**, 159–166 (2019).

405   14. World Health Organisation. Tuberculosis profile: Viet Nam.

406   https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country

407   %22&lan=%22EN%22&iso2=%22VN%22.

408   15. Bonell, A. *et al.* Does sunlight drive seasonality of TB in Vietnam? A retrospective

409   environmental ecological study of tuberculosis seasonality in Vietnam from 2010 to

410   2015. *BMC Infect. Dis.* **20**, 184 (2020).

411   16. Dao, T. P. *et al.* A geospatial platform to support visualization, analysis, and prediction of

412   tuberculosis notification in space and time. *Front. Public Health* **10**, 973362 (2022).

413   17. Srinivasan V, Ha VTN, Vinh DN, Thai PVK, Ha DTM, Lan NH, et al. Sources of Multidrug

414   Resistance in Patients With Previous Isoniazid-Resistant Tuberculosis Identified Using

415   Whole Genome Sequencing: A Longitudinal Cohort Study. Clinical Infectious Diseases.

416   2020 Nov 15;71(10):e532–9.

417   18. Theron, G. *et al.* Data for action: collection and use of local data to end tuberculosis. *The*

418   *Lancet* **386**, 2324–2333 (2015).

419   19. Shrestha, S., Chatterjee, S., Rao, K. D. & Dowdy, D. W. Potential impact of spatially

420   targeted adult tuberculosis vaccine in Gujarat, India. *J. R. Soc. Interface* **13**, 20151016

421   (2016).

422   20. Yuen, C. M. *et al.* Turning off the tap: stopping tuberculosis transmission through active

423   case-finding and prompt effective treatment. *The Lancet* **386**, 2334–2343 (2015).

424   21. Cudahy, P. G. T. *et al.* Spatially targeted screening to reduce tuberculosis transmission in

425   high-incidence settings. *Lancet Infect. Dis.* **19**, e89–e95 (2019).

426   22. Hoa NB, Khanh PH, Chinh NV, Hennig CM. Prescription patterns and treatment outcomes

427   of MDR-TB patients treated within and outside the National Tuberculosis Programme in

428   Pham Ngoc Thach hospital, Viet Nam. Tropical Medicine & International Health.

429   2014;19(9):1076–81.

430   23. Ho Chi Minh City Statistics Office. Statistical Yearbook of Ho Chi Minh City. 2021.

431   24. Thai L, Nhat L, Shah N, Lyss S, Ackers M. Sensitivity, completeness and agreement of the

432   tuberculosis electronic system in Ho Chi Minh City, Vietnam. Public Health Action. 2017;

433   7(4):294-298

434   25. Cambon, J., Hernangómez, D., Belanger, C. & Possenriede, D. tidygeocoder: An R package

435   for geocoding. *J. Open Source Softw.* **6**, 3544 (2021).

436   26. GADM. Vietnam Administrative Areas. https://gadm.org/download_country.html.

437   27. Ord JK, Getis A. Local Spatial Autocorrelation Statistics: Distributional Issues and an

438   Application. Geographical Analysis. 1995;27(4):286–306.

439   28. Anselin L. Local Indicators of Spatial Association—LISA. Geographical Analysis.

440   1995;27(2):93–115.

441   29. R Foundation for Statistical Computing. R: A language and environment for statistical

442   computing. (2022).

443   30. Environmental Systems Research Institute. ArcGIS Online. Redlands, California.

444   31. Nguyen, H. V. *et al.* The second national tuberculosis prevalence survey in Vietnam. *PLoS*

445   *ONE* **15**, e0232142 (2020).

446   32. Marks G, Nhung N, Nguyen T, Hoa N, Khoa T, Son N, Phuong N, Tin D, Ho J, Fox G.

447   Prevalance of latent tuberculosis infection among adults in the general population of Ca

448   Mau, Viet Nam. Int J Tuberc Lung Dis. 2018 Mar 1;22(3):246-251.

449   33. Van Minh, H. *et al.* Prevalence of tobacco smoking in Vietnam: findings from the Global

450   Adult Tobacco Survey 2015. *Int. J. Public Health* **62**, 121–129 (2017).

451 34. Tran, Q. B. *et al.* Risk factors for Non-Communicable Diseases among adults in Vietnam:

452   Findings from the Vietnam STEPS Survey 2015. *J. Glob. Health Sci.* **2**, (2020).

453 35. Nguyen, H. V. *et al.* Influence of Sex and Sex-Based Disparities on Prevalent Tuberculosis,

454   Vietnam, 2017–2018. *Emerg. Infect. Dis.* **29**, 967–976 (2023).

455 36. Long N. Gender specific epidemiology of tuberculosis in Vietnam. Doctoral thesis,

456 Department of Public Health Sciences, Karolinska Institute, Sweden, 2000. Available from:

457 https://openarchive.ki.se/xmlui/handle/10616/44375

458 37. Trinh, Q. M. *et al.* Tuberculosis and HIV co-infection in Vietnam. *Int. J. Infect. Dis.* **46**, 56–

459   60 (2016).

460 38. Glaziou P; Arinaminpathy N; Dodd P.J.; Dean A; Floyd K. *Methods used by WHO to*

461   *estimate the global burden of TB disease.*

462   https://www.who.int/publications/m/item/methods-used-by-who-to-estimate-the-

463   global-burden-of-tb-disease-2022 (2023).

464 39. Wingfield, T. *et al.* Beyond pills and tests: addressing the social determinants of

465   tuberculosis. *Clin. Med.* **16**, s79–s91 (2016).

466 40. Foster, N. *et al.* Social determinants of the changing tuberculosis prevalence in Việt Nam:

467   Analysis of population-level cross-sectional studies. *PLOS Med.* **19**, e1003935 (2022).

468 41. Liebenberg, D., Gordhan, B. G. & Kana, B. D. Drug resistant tuberculosis: Implications for

469   transmission, diagnosis, and disease management. *Front. Cell. Infect. Microbiol.* **12**,

470   (2022).

471 42. Hoa NB, Cobelens FG, Sy DN, Nhung NV, Borgdorff MW, Tiemersma EW. Diagnosis and

472 Treatment of Tuberculosis in the Private Sector, Vietnam. Emerg Infect Dis. 2011;17(3):562-

473 566.

474 43. Wang, W. *et al*. Role of casual contacts in the recent transmission of tuberculosis in

475 settings with high disease burden. *Clin. Microbiol. Infect*. **20**, 1140–1145 (2014).

476 44. Brown, T. S., Robinson, D. A., Buckee, C. O. & Mathema, B. Connecting the dots:

477 understanding how human mobility shapes TB epidemics. *Trends Microbiol*. **30**, 1036–

478 1044 (2022).

479

Supporting information

Figure S1: Ho Chi Minh City administrative boundaries, subdivided by district. 1 to 12 – Districts 1 to 12, BT - Bình Tân, BTH - Bình Thạnh, , GV - Gò Vấp , TB - Tân Bình, TD - Thủ Đức, TP - Tân Phú (all urban). BC - Bình Chánh, CC - Củ Chi, CG - Cần Giờ, HM - Hóc Môn, NB - Nhà Bè (all rural).

Figure S2a: Scatter plots demonstrating univariate associations between ward-level demographic and socioeconomic indicators and the natural logarithm of total TB incidence. Plots include robust line of best fit and Spearman correlation coefficient with corresponding p-value. A – the proportion of the population that is male, B – the proportion of the population aged 30 to 59, C – average number of people per household, D – unemployment rate, E – HIV prevalence. HIV – human immunodeficiency virus, TB – tuberculosis.

Figure S2b: Scatter plots demonstrating univariate associations between ward-level demographic and socioeconomic indicators and the natural logarithm of proportion of all cases that are MDR-TB. Plots include robust line of best fit and Spearman correlation coefficient with corresponding p-value. A – the male proportion of the population, B – the proportion of the population aged 30 to 59, C – population density, D – average number of people per household, E – literacy rate, F – unemployment rate, G – the proportions of households that own a computer, H - HIV prevalence. HIV – human immunodeficiency virus, TB – tuberculosis

31

500   **Figure S3: Scatter plots of fitted versus observed values for 4 negative binomial regression**

501   **models.** Plots include line of best fit.

502   **Table S1: Adjusted incidence rate ratios and 95% confidence**

503   **intervals for the association between ward-level indicators and a)**

504   **total TB incidence and b) MDR-TB case count relative to total TB**

505   **case count, derived from sensitivity analysis.**

506

507

32