



The relationship between the built environment and respiratory health: Evidence from a longitudinal study in Indonesia

Witri Indriyani^{a,b}, Muhammad Halley Yudhistira^{a,c,*}, Prani Sastiono^{a,c}, Djoni Hartono^b

^a Research Cluster on Urban and Transportation Economics, Faculty of Economics and Business, Universitas Indonesia, Indonesia

^b Research Cluster on Energy Modeling and Regional Economic Analysis (RCEMREA), Faculty of Economics and Business, Universitas Indonesia, Indonesia

^c Institute for Economic and Social Research, Faculty of Economics and Business, Universitas Indonesia, Indonesia

ARTICLE INFO

Keywords:

Built environment
Sprawl index
Respiratory infection
Communicable disease
Health
Fixed effect

ABSTRACT

Multiple studies have discussed the relationship between the built environment and non-infectious diseases, but research involving infectious diseases and the built environment is scarce. How the built environment is associated with infectious diseases varies across areas, and previous literature produces mixed results. This study investigated the relationship between the built environment and infectious diseases in Indonesia, which has different settings compared to developed countries. We combined the longitudinal panel data, Indonesian Family Life Survey (IFLS), and land cover data to examine the relationship between the built environment and the likelihood of contracting respiratory infectious diseases. We focused on the sprawl index to measure the built environment. The study confirmed that a sprawling neighbourhood is linked to lower respiratory infection symptoms by employing a fixed effect method. The association is more evident in urban areas and for females. The results also suggested that the linkage works through housing quality, such as housing crowdedness and ventilation, and neighbourhood conditions like neighbourhood transportation modes and air pollution levels. Thus, our results underlined the need to consider the health consequences of the densification policy and determine the direction of landscape planning and policy.

1. Introduction

Where people live could directly affect their health. Some places offer a healthier built environment than other places. Generally, the built environment refers to the human-made physical environment, including land use patterns, transportation systems, physical infrastructure and urban design (Saelens & Handy, 2008). The linkage between the built environment and health outcomes started from ecology theory (Stokols, 1996) and neighbourhood effect theory (Galster, 2012) which find that the built environment could be a health improvement or stressor depending on its characteristics and the human's health-related behaviour.

Whether the built environment is related to health outcomes is an ongoing discussion. On the one hand, a compact built environment might offer higher walkability and lead to active behaviour, thus decreasing the risk of non-communicable diseases like obesity, heart diseases and cancer prevalence (Ewing, Meakins, Hamidi, & Nelson, 2014; Zhao & Kaestner, 2010). On the other hand, a compact built

environment is also associated positively with infectious diseases since compact areas encourage high contact intensity and amplify the risk of disease contagion (Glaeser, 2011). How the built environment is associated with infectious diseases also varies across areas. Recent empirical studies on the built environment and infectious disease show mixed results (Bhadra, Mukherjee, & Sarkar, 2021; Hamidi, Ewing, & Sabouri, 2020; B. Li, Ma, & Zhang, 2021; S. Li, Ma, & Zhang, 2021). While research discovered negative and no significant relationship between a compact built environment and the rate of respiratory infection disease, Covid-19 (Carozzi, Provenzano, & Roth, 2020; Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020), other studies report a positive and significant association between compact built environment and Covid-19 (Bhadra et al., 2021; Emeruwa et al., 2020; B. Li, Ma, & Zhang, 2021; S. Li, Ma, & Zhang, 2021). One explanation for the mixed results was the different characteristics in the backdrop of the studies. In developed countries, even though compactly built environments allow for close physical contact, such areas also benefit more from better health facilities and health education, which enable them to prevent

* Corresponding author.: Institute for Economic and Social Research, Faculty of Economics and Business, Universitas Indonesia. Kampus UI Salemba, Jalan Salemba Raya No.4., Jakarta, 10430, Indonesia.

E-mail address: m.halley@ui.ac.id (M.H. Yudhistira).

<https://doi.org/10.1016/j.ssmph.2022.101193>

Received 25 May 2022; Received in revised form 20 July 2022; Accepted 29 July 2022

Available online 5 August 2022

2352-8273/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

disease infection (Hamidi, Sabouri, & Ewing, 2020). Meanwhile, in developing and less developed countries with limited health facilities, a compact built environment induces a higher contagion due to intense contact. Coupled with overcrowded housing, poor sanitation, and pollution will aggravate the disease transmission and prevalence in developing and less developed countries (Antunes & Waldman, 2001; Cienciewicz & Jaspers, 2007; Mondal & Paul, 2020; Nasim, El-Zein, & Thomas, 2022). Furthermore, the extent to which the built environment is linked to infectious diseases in urban and rural areas could be different. In rural areas, the built environment is less compact which could prevent the villagers from disease transmission. However, the lack of access to health facilities and education increases the disease's prevalence (S. Wang et al., 2015). The higher social capital in rural areas (Ziersch, Baum, Darmawan, Kavanagh, & Bentley, 2009) also has advantages and disadvantages regarding infectious diseases. Higher social capital can accelerate health information exchange. Still, social capital can also increase human contact, such as secondhand cigarette smoking or shared food or drink, potentially accelerating disease transmission (Villalonga-Olives & Kawachi, 2017).

This study investigates the built environment and the likelihood of individuals contracting respiratory infectious diseases in Indonesia, a developing country with a constant rate of respiratory infectious diseases (Ministry of Health, 2018) and experiencing built environment changes (Yudhistira, Indriyani, Pratama, Sofiyandi, & Kurniawan, 2019). Specifically, the study investigates whether the built environment correlates with non-specific acute respiratory infection symptoms, pneumonia symptoms, and tuberculosis (TB). However, measuring the built environment in a small neighbourhood is challenging when the data is limited. Fortunately, technology development provides an alternative approach by utilizing various digital and photographic impressions, such as location tracking by cell phone, satellite imagery, and a street photograph. The analysis employs the combination of social economy and health data from three rounds of the Indonesian Family Life Survey (IFLS) 2000, 2007, and 2014, and land cover data from European Space Agency Climate Change Initiatives (ESA CCI) in the respective years. We measure the built environment with sprawl index using state-of-the-art Geographic Information Systems (GIS) to accommodate an individual's neighbourhood situation at the sub-district level following Burchfield, Overman, Puga, and Turner (2006). The study hypothesizes that sprawled built environments might correlate with the probability of infecting respiratory diseases, and the correlation is statistically significant.

The relationship between the built environment and respiratory infectious diseases could happen due to self-selection. Individuals with certain health statuses self-select into densely populated or sprawled area living conditions. For example, individuals with lower health status might self-select into higher compact living conditions to benefit from better health facilities in densely populated areas. They could also self-select into less dense areas for better environments, such as cleaner air. Hence, the study utilizes a panel fixed-effect method to reduce the bias from time-invariant unobservable due to self-selection. We use the wave-island fixed effect to handle heterogeneity at the island level that might varies across years. Then we restrict our sample to non-migrant individuals to control unobservable factors in sub-districts that might confound the relationship between the built environment and individual health outcomes. Moreover, the longitudinal data, IFLS, also partially deals with individual time-variant confounders. The more refined controls and the data form could accommodate the self-selection problems and make our analysis closer to causality than a simple ordinary least square in cross-section data.

This study is related to the recent literature linking built environments and health outcomes (Bhadra et al., 2021; Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020). Most studies limit themselves by using qualitative explanations in justifying the underlying mechanism of how built environments might associate with health outcomes. We complement these studies by providing a parametric

estimation. The estimation utilizes the IFLS data to identify the channels, i.e., housing crowdedness, housing ventilation, and neighbourhood transportation modes and Village Potential (Podes) data to describe the health facilities at the sub-district level. The study also uses pollution data from the Tropospheric Monitoring Instrument (TROPOMI) onboard European Space Agency Copernicus Sentinel-5 Precursor (S5P) satellite to detect the neighbourhood pollution level. We include three airborne particles, nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and carbon monoxide (CO), to define air pollution. Secondly, the study applies the sprawl index developed by Burchfield et al. (2006) as an approach to the built environment using GIS methods. The sprawl index allows us to capture better the relationship between the built environment and health outcomes than the administrative approach (Drewnowski et al., 2019). Thirdly, we examine a nearly causal relationship between the built environment and an individual's respiratory infectious disease with longitudinal data, while the earlier studies mainly discuss the link to non-communicable diseases or employ cross-sectional data (Ewing et al., 2014; Hamidi, Sabouri, & Ewing, 2020; Zhao & Kaestner, 2010). The longitudinal analysis data could partially absorb time-variant unobservables. Additionally, we enhance the model specification by working on the panel fixed-effect method. The method could reduce the endogeneity by dealing with time-invariant unobservables. Finally, the results of this study will increase the awareness of the implication of the densification policy on respiratory infectious diseases in Indonesia.

The structure of the study is as follows. Section 2 provides the context of the study. Section 3 describes the data and empirical strategy. Section 4 presents the results, and section 5 discusses the findings. Last, section 6 concludes the study.

2. Context: respiratory disease prevalence and built environment in Indonesia

The area's characteristics could affect how the built environment relates to infectious diseases. Indonesia has been facing challenges of limited health facilities, overcrowding housing, and poor sanitation that might influence the relationship between the built environment and respiratory infectious diseases. Statistics of Indonesia 2021 show that 39% of Indonesian households live in indecent housing. They are not meeting the four criteria, i.e., sufficient living space (7.2 m² per capita), having access to clean drinking water, having access to proper sanitation, and building resilience or durable housing.

Indonesia also has issues with health inequality and inadequacy. We note three critical points from the statistics descriptive of Indonesia's health facilities in Table 1. First, the health facilities do not change substantially from 2006 to 2020. Second, the table indicates health facilities' inequality between sub-districts. We still found several sub-districts without Community health care (Puskesmas), whereas the regulation requires at least one Puskesmas as primary health care in each sub-district (Regulation of the Minister of Health of Republic of Indonesia Number 75 the Year 2014 about Community Health Center, 2014). Third, the number of Puskesmas is still inadequate compared to the average sub-district population (Based on Podes, 2007, the average population in each sub-district is around 40 thousand people). Table 1 shows that on average, each sub-district has one to two Puskesmas, serving approximately 40 thousand people, or the ratio is one Puskesmas for more than 20,000 people. The ratio is lower than the government minimum target, 1:16,000 (Regulation of the Minister of Law and Human Rights of Republic of Indonesia Number 34 the Year 2016 about District/City of Human Rights Care, 2016). Indonesia was also the second lowest among Southeast Asia countries, with the ratio of physicians per person, at 0.46 per 1000 people (World Bank, 2019).

The basic health research (RISKESDAS) exhibits that the rate of respiratory infection in Indonesia remains unchanged and tends to increase in more compact neighbourhoods, e.g., urban areas. Table 2 provides the prevalence of Indonesia's respiratory infectious diseases,

Table 1
Number of health facilities at the sub-district level.

	2006			2014			2020		
	mean	min.	max.	mean	min.	max.	mean	min.	max.
Hospital	0.3	0	9	0.3	0	10	0.4	0	16
Community health care (Puskesmas)	1.5	0	11	1.5	0	18	1.5	0	14
Polyclinic and infirmary	2.0	0	71	1.7	0	101	2.0	0	71
Supporting Puskesmas	4.2	0	38	3.7	0	35	3.6	0	32
Doctor's private practice	5.6	0	218	4.7	0	145	5.2	0	125
Pharmacy	1.8	0	88	2.8	0	95	4.1	0	79

Source: Village Potential (Podes), Statistics of Indonesia

Table 2
The prevalence of respiratory infectious diseases in Indonesia.

	2007	2013	2018
Pneumonia			
National	2.1	1.8	4.0
Urban	1.6	1.6	3.8
Rural	2.4	2.0	4.3
Male	1.6	1.9	4.2
Female	2.4	1.7	3.9
Non-specific acute respiratory infections			
National	25.5	25.0	9.3
Urban	23.3	24.1	9.0
Rural	26.9	26.0	9.7
Male	25.6	25.1	9.0
Female	25.5	24.9	9.7
Tuberculosis			
National	0.4	0.4	0.4
Urban	0.4	0.4	0.4
Rural	0.4	0.3	0.4
Male	0.4	0.4	0.5
Female	0.4	0.3	0.4

Source: Indonesia Basic Health Research, Ministry of Health (2008, 2013, 2018).

Notes: The prevalence of pneumonia and non-specific acute respiratory infections was calculated based on the diagnosis of health workers (doctor, nurse, midwife) and the symptoms. The TB prevalence was estimated from the health workers' diagnosis only. Specifically, in 2018, the TB prevalence was counted if diagnosed by a doctor. The period of pneumonia and non-specific acute respiratory infection symptoms was similar in data between 2007 and 2013, during the last month when surveyed. In 2018, the duration of the pneumonia check was extended up to the last year when surveyed, and the period of non-specific acute respiratory infections was restricted only to the last two weeks when surveyed. The diagnosis of TB in the three surveys was asked during the last year when surveyed.

including non-specific acute respiratory infections, pneumonia, and TB, which are common in Indonesia. We need to read the data carefully since the data in 2018 was collected differently than those in 2007 and 2013. The TB prevalence might still be comparable across the years, but pneumonia and non-specific acute respiratory infections are only comparable in 2007 and 2013. The prevalence of TB nationally was constant and increased slightly in 2018, showing that the health program designed to combat TB had not yet been successful. In the early years, the TB prevalence was higher in rural areas, but the number in urban areas has recently been more evident. Among genders, the TB prevalence was consistently higher for males. In 2021 Indonesia ranked 3rd globally in terms of TB incidence (World Health Organization, 2021).

The prevalence of pneumonia decreased from 2007 to 2013 by more than 15% or around 2% annually. The decrease mainly occurred in rural areas, whereas the decline in urban areas was only 1.8% or about 0.3% annually. It is also noticeable that pneumonia rose substantially (16.6%) for males while the female prevalence dropped drastically (30%) in 2013. The non-specific acute respiratory infection prevalence was steady between the two survey periods, but the prevalence increased in

urban areas. The number was also consistently higher for males, though the gap is small between gender. From the data, we could see the paradox that the prevalence of non-specific acute respiratory infections in the urban area is growing, whereas the national prevalence is declining.

At the same time, population growth and infrastructure development have changed Indonesia's built environment (Pratama, Yudhistira, & Koomen, 2022; Yudhistira et al., 2019). Table 3 present the average sprawl index in a 5 km radius and developed areas in Indonesia. The sprawl index has decreased, indicating that the built areas are more compact nationally. In contrast, the proportion of the developed areas increases by about 2% annually. The data show the changes in the built environment in Indonesia, regardless of the small changes.

Despite Indonesia's socioeconomic and health challenges, the empirical study that linked health, particularly infectious respiratory disease and the built environment in Indonesia is limited. The recent research mainly focused on the socioeconomic and health behaviour determinants of respiratory infectious disease and neglected the role of the built environment (Fathmawati, Rauf, & Indraswari, 2021; Oktaria et al., 2021; Windi et al., 2021). They found that age, parent's occupation, household wealth, place of residence, vitamin D deficiency, and water sources are associated with respiratory infection among children under five years in Indonesia. Our study complements the existing literature by investigating the association between respiratory infectious disease and the built environment in Indonesia and providing potential channels, such as housing, neighbourhood condition and health facilities.

3. Data and identification

3.1. Data

The study employs socioeconomic and health data from IFLS3 (2000), IFLS4 (2007), and IFLS5 (2014). We do not include IFLS1 as IFLS1 does not provide information related to runny nose and difficulty breathing, which is important to define our main interest variable (respiratory infection symptom). We also find incomparable variables in IFLS1 with other waves, such as the breakdown of education level. Also, we exclude the IFLS2 since the data were collected when a severe economic crisis started in Indonesia might affect the socio-health-economic condition, including health. The IFLS data provides longitudinal panel data, covers over 30,000 respondents from 13 out of 34 provinces in Indonesia, and represents 83% of the Indonesian population with detailed sub-district information. The first wave survey was conducted in 1993, followed by the second and third in 1997 and 2000. The total

Table 3
The built environment of Indonesia.

	2000	2007	2014
Sprawl index	88.5	87.4	86.4
Developed areas (%)	7.5	8.4	9.6

Source: ESA CCI (2000, 2007, 2014).

number of households interviewed was about 10,000 households (with around 39,000 individuals) in wave 3; 13,000 (with around 43,000 individuals) in wave 4; and 16,000 households (with around 50,000 individuals) in wave 5.

The IFLS produces detailed information at the individual, household, and community level covering socioeconomics and demographic condition, and health to risk-averse behaviour. Since the self-health assessment data in the IFLS is only available for individuals aged 15+ and the TB information is only available for individuals aged 40+, thus our sample was 15 years old or older interviewed in the three waves (waves 3–5). We use sub-district ID to combine IFLS and sprawl index data. Thus, the individual data is nested in sub-district and time (wave). We combine sprawl index, pollution, and health facilities data similarly. The decision to link the sprawl index with IFLS data at the subdistrict level is due to the data availability, as data on the household location is only available at the subdistrict level. After combining the IFLS data with sprawl index data and taking only observations with no missing values, we get 9758 individuals' and 7021 households' panels and 1070 sub-districts in our samples. We restricted the sample to non-migrant individuals to accommodate the unobservable at the sub-district level, affecting the individual location choice. After the restriction, our observations reduced considerably to around 9279 individuals, 5728 households, and 733 sub-districts. We also observed the community and facilities data in the corresponding waves for further explanation. The community questionnaires were asked of the community leader. In total, we have 170 communities in those three waves. Like the individual data, the household and neighbourhood data are nested in sub-district and time (wave).

Previous research uses various ways when assessing health outcomes, among others conducting a self-reported assessment, medical record, and prescription record analysis (Barbara et al., 2012; Brinkhues et al., 2018). This study employs a self-reported assessment of respiratory infectious diseases from IFLS since formal or medical report data with socioeconomics information in Indonesia is unavailable. For the TB, we use a proxy of self-reported TB medication consumption and medical diagnosis in the related year. The TB information is only available in IFLS4 (2007) and IFLS5 (2014) for respondents aged 40+. Since the IFLS does not provide information on pneumonia and non-specific acute respiratory infectious diagnosis, we adopt the symptom identifications in IFLS following the RISKESDAS report. In the RISKESDAS report, individuals are suspected of having pneumonia symptoms if they have had a high fever, cough, and difficulty breathing during the last year when surveyed. Individuals are suspected of having non-specific acute respiratory infection symptoms when they had a high fever, runny nose, cough, or a sore throat during the previous two weeks when surveyed. Thus, the study utilizes similar criteria of the RISKESDAS to identify pneumonia and the non-specific acute respiratory infection symptoms in the IFLS. Some literature also applied a similar approach (Brinkhues et al., 2018; Windi et al., 2021). We include individual and household characteristics as control variables, such as age, working status, marital status, education level, gender, household members, and the head of the household. Since gender is time-invariant, we exclude gender in the fixed effect estimation.

This study's main variable of interest is the built environment, measured by the sprawl index following Burchfield et al. (2006) at the sub-district level. When constructing the sprawl index, we use land cover maps from ESA CCI in the related year with a spatial resolution of 300mx300 m. The data is available yearly; thus, we can combine it with the IFLS data. We classify the urban and non-urban settlements to define the developed and undeveloped areas within a 5 km radius of the individual grid cells. Forty years timespan (2000–2014) allow a considerable change in the sprawl index that should lead to sufficient variation in the data. The study applies a moderate radius, 5 km, to gain the variability of the sprawl index, especially in the sub-district dominated by undeveloped areas, and to capture the neighbourhood effects, such as transportation modes and pollution. Previous studies also use a similar

radius when measuring the built environment (Surya et al., 2021; Zhang, Ning, Chen, & Yang, 2021). Fig. A2 describes how we calculate the sprawl index.

3.2. Empirical strategy

Our empirical strategy is as follows. First, we measure the sprawl index at the sub-district level where the individuals live when surveyed. Then, we observe the same individuals in three waves. We propose an individual fixed effect and wave fixed effect to account for time-invariant unobservable and time trends. Thus, our main estimation:

$$Y_{1ijt} = \alpha_0 + \alpha_1 SI_{jt} + \alpha_2 X_{1ijt} + \gamma_i + \delta_t + \eta_{st} + u_{ijt} \quad (1)$$

For household channels, we also use panel fixed effect as follows

$$Y_{2hjt} = \beta_0 + \beta_1 SI_{jt} + \beta_2 X_{2hjt} + \gamma_h + \delta_t + \eta_{st} + u_{hjt} \quad (2)$$

For neighbourhood channels, we use ordinary least squares (OLS) due to small observations as follows

$$Y_{3cjt} = \Theta_0 + \Theta_1 SI_{jt} + \Theta_2 X_{3cjt} + u_{cjt} \quad (3)$$

3.3. Dependent variable

The dependent variable of this study is Y_{1ijt} . Y_{1ijt} is an indicator variable of getting the respiratory infectious disease symptoms of individual i in sub-district j at time or wave t . The Y_{1ijt} is a binary variable, assign a value for 1 if an individual has a positive diagnosis or symptoms of respiratory infectious diseases, and 0 otherwise. Y_{2hjt} is an indicator variable of housing crowdedness or housing ventilation adequacy of household h in sub-district j at time or wave t . Housing crowdedness is measured by dividing household members by housing floor size. Ventilation adequacy is a binary variable (1 for good ventilation and 0 otherwise). Y_{3cjt} is an indicator variable of the existence of transportation modes in community c , in sub-district j at time or wave t . Y_{3cjt} consist of three indicators, the existence of public transport, transport mode to the station, bus stop, terminal, and transport mode to market. The existence of public transportation with three or four wheels or motorboat service is defined as 1 if available and 0 otherwise. Similarly, the transport mode was defined as 1 if use two-wheeled motor vehicles, pedicabs, bicycles, and on foot and 0 otherwise.

3.4. Variable of interest and controls

Our variable of interest in this study is SI_{jt} . SI_{jt} is a sprawl index of sub-district j in a certain radius at time or wave t . We use a 5 km radius for our baseline estimate. The sprawl index ranged from 0 to 100, with a higher index indicating a higher sprawl or scatterer neighbourhood. Coefficient α_1 captures the relationship between the sprawl index and the likelihood of having respiratory infectious diseases. Coefficient β_1 describes the relationship between the sprawl index and housing crowdedness or ventilation adequacy. Coefficient Θ_1 accounts for the relationship between the sprawl index and the transportation facility in the community.

By employing the fixed effect, we can analyze how the change in sprawl index correlates with the health outcomes of the same individual over time. This study also employs X_{1ijt} , a vector control variable for individual i in sub-district j at time t that might relate to the diseases such as age, working status, marital status, education level, and households condition following the existing research (Hamidi, Sabouri, & Ewing, 2020; Miandad, Anwar, Ahmed, Rahman, & Khan, 2019; Murad et al., 2019; L. Wang, Chen, Sun, Yang, & Li, 2020). γ_i is an individual fixed effect to control individual time-invariant unobservable factors that could confound the relationship between health outcomes and built environment form, and η_{st} is a wave-island fixed effect or wave-island interaction to absorb heterogeneity at the island level that varies across years. We also include the survey wave fixed effect, δ_t , to account

for any possible changes over time, and u_{ijt} is an error term. Similarly, X_{2hjt} is a vector control variable for household h in sub-district j at time t , i.e., age of the household head, gender of the household head, marital status of the household head, number of household members, and years of schooling of the household head. γ_h is a household fixed effect and u_{hjt} is an error term. X_{3cjt} is a vector control variable for community c in sub-district j at time t . X_{3cjt} consists of road condition (1 if asphalt or cement roads, 0 otherwise) and percentage of households with electricity, and the existence of industry or factory or plant (1 if any, 0 otherwise) and u_{cjt} is an error term.

We use a fixed effect approach in our main estimation to reduce the endogeneity problems associated with individual self-selection. Individuals with certain health statuses might choose their residential locations. For instance, individuals with respiratory diseases might consciously choose to live in cleaner neighbourhoods that are located in sprawled areas. The study also restricts the sample to non-migrants to ensure the non-existence of self-selection across sub-districts. All standard errors in this model are clustered at the sub-district level to deal with the potential serial correlation due to exogenous shock across sub-district that is uncorrelated with sprawl index but explains the likelihood of getting infectious respiratory diseases in all sub-districts.

We also conduct several robustness checks in our model to test the consistency of our baseline results. First, the study applies different radii of the sprawl index. We estimate the model in 1 km and 10 km radii. Second, since we have a binary variable of Y we employ fixed-effect logistic regression. Third, underlining the possibility of endogeneity problems, we test the estimate using a two-stage least square (TSL) with internal instruments adopted from [Baum and Lewbel \(2019\)](#). Last,

we omit 10% of the oldest observations to absorb the possibility of the extreme value of observations.

According to the background and the empirical results of the relationship between the built environment and the likelihood of getting respiratory infectious diseases, we hypothesize that the built environment and the possibility of getting respiratory infectious diseases are correlated and statistically significant.

4. Results

4.1. The relationship between the built environment and respiratory diseases

[Fig. 1](#) illustrates the negative trend of sprawl index and respiratory infectious diseases. The figures suggest that the sprawler areas are associated with a lower rate of respiratory infectious diseases. The point-biserial correlation (r_{pb}) also shows negative correlations between sprawl index and respiratory infectious diseases. [Fig. 1c](#) notes that non-specific respiratory infection has the highest correlation with sprawl index than other mentioned respiratory infectious diseases. Conversely, TB has the lowest correlation with sprawl index among the diseases. The descriptive statistics in [Tables A1 and A2](#) also clarify that the proportion of individuals with infectious respiratory symptoms is higher in a more compact sub-district (the sprawl index is less than the median) than in a sprawling sub-district (the sprawl is higher than the median). The table also indicates the difference in individual characteristics between compact and sprawl sub-district. The more compact sub-district has a lower proportion of married and not working individuals, higher years

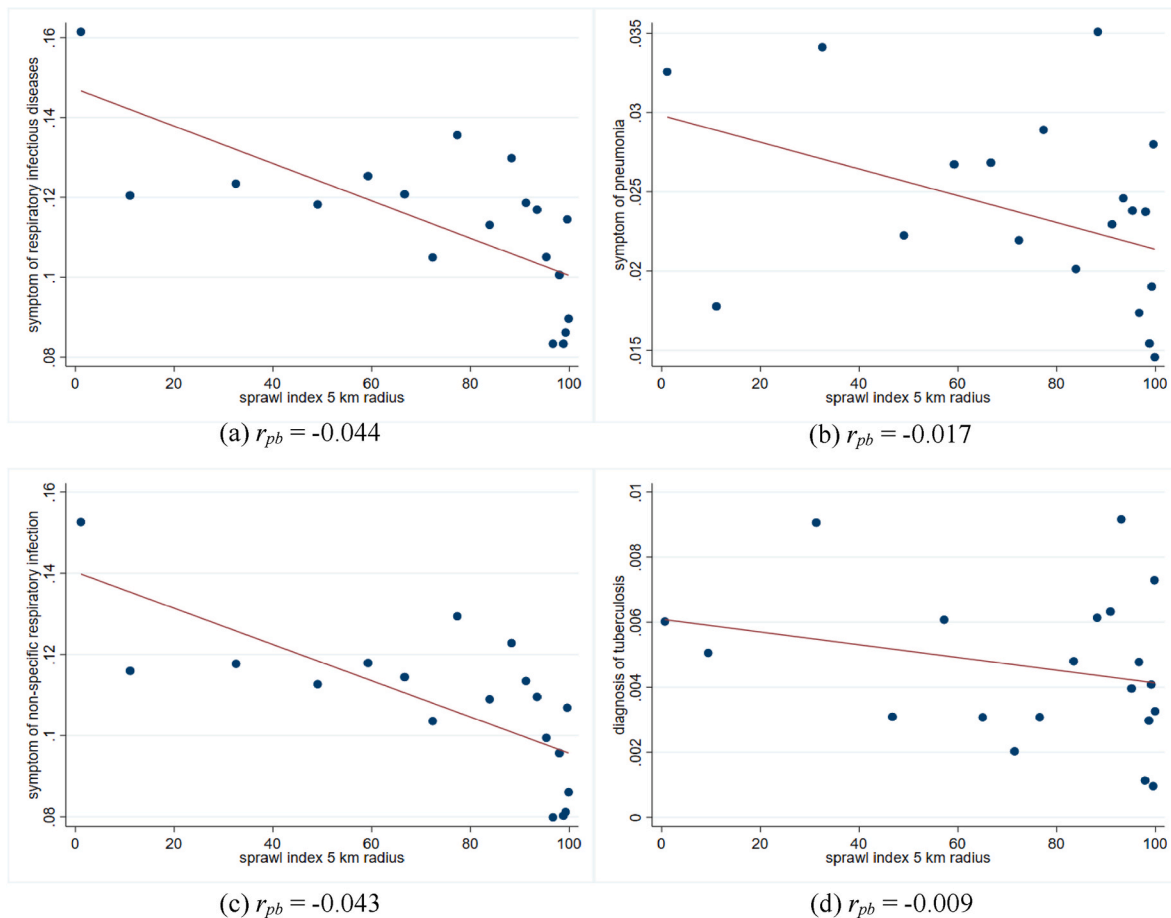


Fig. 1. The relationship between sprawl index and respiratory infectious disease

Notes: [Fig. 1](#) is a binned scatter plot of sprawl index and individuals' respiratory infectious diseases at the sub-district level. The binned scatter group the sprawl index into 20 equally sized bins. r_{pb} is the point biserial correlation coefficient.

of education, and higher household members.

Table 4 presents the results of fixed effect estimation. We start the analysis by identifying the link of sprawl index on the likelihood of having respiratory infectious diseases from a 5 km radius around the neighbourhood with individual characteristics' control in the first column. The result shows that the sprawl index is negatively correlated with the likelihood of respiratory infectious illness symptoms by 0.2% per one percentage point of the sprawl index. The coefficient is statistically significant at the 1% level. Then, we gradually introduce the baseline controls in columns (2)–(4) based on equation (1). The coefficients decrease slightly to 0.17% with lower significance.

The model in columns (1)–(4) might still suffer from the sub-district unobserved heterogeneous variables that might influence individuals' location preferences, for instance, the eviction of residential areas for local infrastructure development. Hence, we restrict our sample to non-migrant individuals in column (5), resulting in a higher coefficient. With the non-migrant sample, one percentage point expansion of the sprawl index corresponds to the lower likelihood of being ill by 0.25%. The coefficient is statistically significant at the 1% level. Ignoring the unobservable at the sub-district level could underestimate the results. We explain that health facilities in each sub-district might confound the relationship. To hold our argument, the study correlates the sprawl index with health facilities, such as the number of hospitals, community health centres, clinics, practitioners, and pharmacies in Fig. A1. The figures depict that denser (less sprawl) areas experience better health facilities, supporting our discussion of a potential confounder between sprawl index and health outcome. Based on the empirical strategy, we will focus on the non-migrant sample with full controls as suggested in column (5) to minimize the potential omitted variable biased.

We also examine the link between the sprawl index and the possibility of having infectious diseases by disease type, non-specific acute respiratory diseases, pneumonia, and tuberculosis in Table 5. The non-

specific acute respiratory infections and pneumonia coefficient are consistent with baseline results when applying the 5 km radius. Columns (1) and (2) report that a one percentage point addition of the sprawl index corresponds to a lower probability of getting non-specific acute respiratory infections and pneumonia by 0.26% and 0.07%. Based on the coefficient estimate, the sprawl index contributes around 7% of the national non-specific acute respiratory infection prevalence and 0.12% of national pneumonia prevalence when simulating the coefficients to the national prevalence.

Meanwhile, the association of the sprawl index with the TB disease in column 3 is insignificant. We argue that the smaller sample size reduces the statistical power. Compared to other infectious diseases, TB's observation numbers are only three-quarters, and the average number of respondents with the disease is only 0.4% of the total sample.

4.2. Sample heterogeneity

Our models suggest that sprawl could reduce the likelihood of getting respiratory infectious diseases. But the results might be heterogeneous between sub-samples, so we checked some sub-samples, the urban-rural sub-sample and the male-female sub-sample.

4.2.1. Urban-rural heterogeneity

We examine the relationship between the sprawl index and the likelihood of contracting respiratory infectious diseases in urban-rural areas. The different characteristics and amenities between urban and rural areas might explain the mixed relationship. Table 6, columns (1)–(2) infer that the link differs between urban and rural areas. The coefficient of the urban sample in column (1) indicates the 0.27% reduction in the probability of being ill for a one percentage point increase in the sprawl index, and the coefficient is statistically significant at a 1% level. In other words, the compact built environment is associated with

Table 4
The relationship between sprawl and respiratory infectious disease, baseline estimates.

	FE: respiratory infection				
	(1)	(2)	(3)	(4)	(5)
Sprawl index 5 km radius	-0.0020*** (0.0007)	-0.0020*** (0.0007)	-0.0012* (0.0007)	-0.0017** (0.0007)	-0.0025*** (0.0009)
Age	-0.0029** (0.0014)	-0.0028** (0.0013)	-0.0029 (0.0029)	-0.0032 (0.0029)	-0.0048 (0.0038)
Age-squared	0.00004** (0.00001)	0.00004** (0.00001)	0.00002 (0.00002)	0.00002 (0.00002)	0.00003 (0.00002)
Years of schooling	0.0022 (0.0020)	0.0023 (0.0021)	0.0028 (0.0021)	0.0028 (0.0021)	0.0044 (0.0029)
Marital = married	-0.0143 (0.0091)	-0.0155* (0.0091)	-0.0135 (0.0091)	-0.0133 (0.0091)	-0.0048 (0.0122)
Working = yes	0.0069 (0.0067)	0.0066 (0.0067)	0.0089 (0.0066)	0.0091 (0.0066)	0.0049 (0.0089)
Number of household members		-0.0020* (0.0011)	-0.0003 (0.0011)	-0.0004 (0.0011)	-0.0015 (0.0017)
Years of schooling of the household head		-0.0002 (0.0009)	0.0001 (0.0009)	0.0001 (0.0009)	-0.0017 (0.0013)
Constant	0.3010*** (0.0685)	0.3040*** (0.0684)	0.2660** (0.1080)	0.3410*** (0.1140)	0.4380*** (0.1540)
Wave fixed effect	No	No	Yes	Yes	Yes
Wave-island fixed effect	No	No	No	Yes	Yes
Sample	All	All	All	All	Non-migrant
Mean of dep variable	0.112	0.112	0.112	0.112	0.111
Adjusted R ²	0.002	0.002	0.007	0.008	0.008
Observation	26442	26442	26442	26442	19748

Notes: The sample in columns (1)–(4) is individuals aged 15+ and in column (5) is non-migrant individuals aged 15+. The study defined non-migrants if the individual did not move to another sub-district across waves. In other words, the individual had the same sub-district address during the three survey waves. All samples were obtained from the three waves of the IFLS survey (2000, 2007, and 2014), with an outcome equal to 1 if the individuals showed respiratory infectious disease symptoms or diagnosis. For comparison, the study provides baseline estimation with cross-section person weight 2000 and longitudinal person weight in Table A3, Appendix. The inverse probability weighting estimates have similar directions and slightly higher coefficients. Standard errors clustered at the sub-district level were reported in parentheses. Asterisks denoted significance: ***p < 0.01, **p < 0.05, *p < 0.10.

Table 5
The relationship between sprawl and respiratory infectious disease by disease's type.

	non-specific acute respiratory infection	pneumonia	tuberculosis
	(1)	(2)	(3)
Sprawl index 5 km radius	-0.0026*** (0.0008)	-0.0007* (0.0004)	0.00003 (0.0003)
Age	-0.0049 (0.0039)	-0.0001 (0.0023)	-0.0004 (0.0027)
Age-squared	0.00003 (0.00002)	0.00002 (0.00001)	0.000003 (0.000008)
Years of schooling	0.0038 (0.0028)	0.0006 (0.0014)	0.0004 (0.0010)
Marital = married	-0.0041 (0.0121)	0.0026 (0.0056)	-0.0003 (0.0033)
Working = yes	0.0059 (0.0087)	-0.0027 (0.0041)	0.0009 (0.0029)
Number of household members	-0.0021 (0.0016)	0.00001 (0.0008)	0.0011 (0.0007)
Years of schooling of the household head	-0.0015 (0.0013)	0.0003 (0.0006)	0.0001 (0.0003)
Constant	0.4410*** (0.1540)	0.0481 (0.0805)	-0.0020 (0.1220)
Wave fixed effect	Yes	Yes	Yes
Wave-island fixed effect	Yes	Yes	Yes
Sample	Non-migrant	Non-migrant	Non-migrant
Mean of dep variable	0.106	0.024	0.004
Adjusted R ²	0.007	0.005	0.004
Observation	19745	19745	14643

Notes: The sample is non-migrant individuals aged 15+ from three waves of the IFLS survey (2000, 2007, or 2014), with an outcome equal to 1 if the individual showed non-specific acute respiratory infection or pneumonia symptoms or consumed TB medication or diagnoses with TB. The TB information is only available in IFLS4 and IFLS5 and was asked for individuals aged 40+. The sprawl index was calculated with a 5 km radius. We also control for health facilities (e.g. number of hospitals, number of community health care (Puskesmas), and number of Supporting Puskesmas) to account for access to health facilities that might be associated with the sprawl index. The health facilities data is obtained from Potential Village data in 1999, 2006, and 2014 (the closest year of IFLS) since the IFLS data do not provide detailed information on health facilities at the sub-district level. The result is similar to the model without health facilities control estimation, suggesting that this problem might be less prevalent in our model. Standard errors clustered at the sub-district level were reported in parentheses. Asterisks denote significance: ***p < 0.01, **p < 0.05, *p < 0.10.

respiratory disease symptoms in urban areas. In comparison, the rural areas' coefficient in column (2) is positive but insignificant. The results support our earlier argument that the relationship between built environment forms and health outcomes in more compact areas, like urban areas, is more significant.

4.2.2. Gender heterogeneity

Additionally, we estimate the relationship between gender in Table 6, columns (3)–(4), to account for the variation of biological characteristics between males and females that might affect the probability of getting respiratory diseases (Groeneveld et al., 2020). Our results show that the sprawl and health outcomes linkage between gender is similar but seem more susceptible for females as the coefficient for the female category in column (4) is higher than the coefficient for the male in column (3). In column (4), a one percentage point of sprawl index is related to the lower probability of contracting the diseases by 0.34% for females. The association is insignificant for the male sub-sample.

4.3. Channels

This section provides two general potential mechanisms to what extent sprawl index is associated with respiratory infectious diseases. We test the land rent theory by Mills (1972) and Muth (1969) and

Table 6
Sample heterogeneity: areas and gender.

	Areas		Gender	
	urban	rural	male	female
	(1)	(2)	(3)	(4)
Sprawl index 5 km radius	-0.0027*** (0.0010)	0.0028 (0.0029)	-0.0013 (0.0012)	-0.0034*** (0.0012)
Age	-0.0061 (0.0055)	-0.0055 (0.0062)	-0.0048 (0.0061)	-0.0047 (0.0052)
Age-squared	0.00002 (0.00003)	0.00005* (0.00003)	0.00002 (0.00003)	0.00003 (0.00003)
Years of schooling	0.0059* (0.0035)	0.0029 (0.0054)	-0.0015 (0.0044)	0.0109*** (0.0035)
Marital = married	0.0027 (0.0152)	-0.0061 (0.0263)	-0.0061 (0.0188)	0.0007 (0.0145)
Working = yes	0.0060 (0.0124)	-0.0039 (0.0141)	-0.0047 (0.0181)	0.0071 (0.0094)
Number of household members	-0.0014 (0.0021)	-0.0020 (0.0036)	-0.0036 (0.0026)	-0.0001 (0.0020)
Years of schooling of the household head	-0.0021 (0.0017)	-0.0005 (0.0025)	-0.0044* (0.0024)	0.0001 (0.0016)
Constant	0.1950 (0.2470)	-0.0146 (0.3330)	0.2900 (0.2270)	0.6340*** (0.2310)
Wave fixed effect	Yes	Yes	Yes	Yes
Wave-island fixed effect	Yes	Yes	Yes	Yes
Sample	Non-migrant	Non-migrant	Non-migrant	Non-migrant
Mean of dep variable	0.118	0.100	0.112	0.111
Adjusted R ²	0.008	0.014	0.009	0.010
Observation	12331	7417	8599	11149

Notes: The sample is non-migrant individuals aged 15+ from three waves of the IFLS survey (2000, 2007, and 2014), with an outcome equal to 1 if the individual showed respiratory infectious disease symptoms or diagnosis. Standard errors clustered at the sub-district level were reported in parentheses. Asterisks denoted significance: ***p < 0.01, **p < 0.05, *p < 0.10.

transportation modes choice by McFadden (1974) to explain the mediators.

4.3.1. Household effect: housing crowdedness and housing ventilation

The land rent in dense areas is higher than in less dense areas since the land is economically more productive (Mills, 1972; Muth, 1969). As a result, a moderate income is only able to afford smaller and lower-quality housing in dense areas. The higher housing crowdedness could speed up the transmission of infectious diseases (Hu, Roberts, Azevedo, & Milner, 2021). Poor housing ventilation also contributes to disease transmission (Luongo et al., 2016). Reversely, people could choose a bigger and better housing quality by living further away from dense areas and having a lower risk of contracting diseases.

To support the argument of housing-related changes, we assess the linkage of the sprawl index on housing crowdedness and housing ventilation in different areas (urban vs rural) following Equation (2). We use the housing conditions as a potential channel to explain whether the built environment is associated with illness due to respiratory infectious diseases by area in Table 7, columns (1)–(4). Table 7 shows some evidence of the link between the sprawl index and household conditions. The results present that the sprawl index is negatively correlated with the housing crowdedness in urban and rural areas (columns 1 and 2) but insignificant. The coefficients indicate that the sprawled neighbourhood in urban areas is associated with lower housing crowdedness, but the ventilation adequation is better in rural areas than in urban areas with a higher sprawl index (columns 2 and 4). The mechanism results partly support our baseline estimates.

4.3.2. Neighbourhood effect: neighbourhood transportation modes and air pollution

We also explain the channel of sprawl index and the likelihood of

Table 7
Channels: Household effects.

	Household effect			
	Urban		Rural	
	crowded	ventilation	crowded	ventilation
	(1)	(2)	(3)	(4)
Sprawl index 5 km radius	-0.0004 (0.0009)	0.0004 (0.0065)	-0.0006 (0.0022)	0.0269** (0.0137)
Age of the household head	-0.0018 (0.0027)	0.0224* (0.0135)	-0.0029 (0.0021)	0.0037 (0.0170)
Age of the household head-squared	0.00002 (0.00003)	-0.0002 (0.0001)	0.00002 (0.00002)	-0.00004 (0.0002)
Gender of the household head = male	0.0134 (0.0167)	0.1210 (0.0939)	-0.0346*** (0.0132)	0.1770 (0.1540)
Marital status of the household head = married	0.0016 (0.0178)	-0.1110 (0.1080)	0.0373** (0.0155)	-0.2070 (0.1490)
Number of household members	0.0053** (0.0022)	0.0036 (0.0086)	0.0209*** (0.0038)	0.0096 (0.0098)
Years of schooling of the household head	0.0006 (0.0017)	0.0063 (0.0068)	-0.0025 (0.0016)	-0.0002 (0.0111)
Constant	0.0884 (0.0953)	0.3520 (0.5960)	0.1310 (0.1880)	-1.8200 (1.3340)
Wave fixed effect	Yes	Yes	Yes	Yes
Wave-island fixed effect	Yes	Yes	Yes	Yes
Sample	Household	Household	Household	Household
Mean of dep variable	0.103	0.801	0.089	0.812
Adjusted R ²	0.055	0.056	0.449	0.204
Observation	3631	3633	2481	2484

Notes: The household sample is the household where our individual sample is located. Standard errors clustered at the sub-district level were reported in parentheses. Asterisks denoted significance: ***p < 0.01, **p < 0.05, *p < 0.10.

having respiratory infectious diseases via transport mode choice following Equation (3). The rationale is residents in the densely built environment have several options of mass public transport or high occupancy vehicles, allowing intense physical contact among the passengers. Since a respiratory virus or pathogen could transmit from humans to humans by direct body contact or droplet (Xu, Tian, Sabel, & Xu, 2019), public mass transport or high occupancy vehicles increases the risk of disease contamination.

Unfortunately, the individual transport modes choice data is unavailable in the IFLS; thus, we used some approaches at the community level or equal to the village. From community and facility data, we could identify the villagers' transportation modes for daily activities. Table 8

Table 8
Channel: Neighbourhood effect.

	Neighbourhood effect					
	Urban			Rural		
	public transport	transport mode to the station, bus stop, terminal	transport mode to market	public transport	transport mode to the station, bus stop, terminal	transport mode to market
	(1)	(2)	(3)	(4)	(5)	(6)
Sprawl index 5 km radius	-0.0014*** (0.0005)	0.0029*** (0.0006)	0.0019*** (0.0007)	-0.0046** (0.0021)	0.0010 (0.0021)	-0.0015 (0.0039)
Road = asphalt or cement	0.4300*** (0.1500)	-0.1570 (0.1410)	0.0019 (0.1570)	0.2450** (0.1040)	-0.1780 (0.1200)	-0.2110* (0.1080)
Electricity (%)	0.0016 (0.0037)	0.0019 (0.0034)	-0.0059 (0.0047)	-0.0043** (0.0019)	0.0059*** (0.0018)	0.0044*** (0.0016)
Industry/factory/plant = yes	0.0310 (0.0425)	-0.0451 (0.0447)	0.0217 (0.0487)	0.0399 (0.0675)	-0.1300* (0.0739)	-0.0068 (0.0671)
Constant	0.3240 (0.3750)	0.0745 (0.3590)	0.6920 (0.4740)	1.1830*** (0.2470)	0.0151 (0.2510)	0.2350 (0.4020)
Sample	Community	Community	Community	Community	Community	Community
Mean of dep variable	0.847	0.237	0.246	0.613	0.393	0.307
Adjusted R ²	0.050	0.051	0.030	0.032	0.031	0.017
Observation	452	452	452	163	163	163

Notes: The community sample is obtained from the community and facilities data of IFLS. The questionnaires were asked of the community leader. Standard errors clustered at the sub-district level were reported in parentheses. Asterisks denoted significance: ***p < 0.01, **p < 0.05, *p < 0.10.

columns (1)–(6) illustrate the association between the sprawl index and neighbourhood situations. Columns (1) and (4) show the link between the sprawl index and the availability of public transport with three or four wheels or motorboat service. The results suggest the higher sprawl index in urban areas correlates with less public (mass) transport with three or four-wheel or motorboat service. We also categorize transportation modes into single and high-occupancy vehicles. We assume that high occupancy vehicles have a greater risk of infection than single-occupancy vehicles as they could hold two or more passengers. To be more specific, we estimate the relationship between the sprawl index and the type of transportation modes to reach the nearest bus stop, terminal, station, or pier in columns (2) and (5) and the nearest market in columns (3) and (6). We define two categories of transportation modes, 1 for single-occupancy vehicles and 0 for high occupancy vehicles, privately or publicly owned. In urban areas, all coefficients are positive and significant in 5 km, implying that the sprawl index is positively related to the use of single-occupancy vehicles, such as two-wheeled motor vehicles, pedicabs, bicycles, and on foot to reach sub-district centres. The findings support our baseline estimates where the relationship is more evident in urban areas.

In addition, previous studies show a linkage between air pollution and the transmission and severity of respiratory infectious diseases. They suggest that the interaction between pollution concentration and respiratory viruses could negatively affect the respiratory system (Bereitschaft & Debbage, 2013; Domingo & Rovira, 2020). As planned compactly built environments might have lower pollution levels than unplanned compactly built environments (Hankey & Marshall, 2017; F. Li & Zhou, 2019; She et al., 2017), the impact of pollution on diseases could be different. Hence, we evaluate the association between the sprawl index and outdoor pollution at the sub-district level as a potential mediator.

In Fig. A3 – A5, we provide the scatter plot and Pearson correlation coefficients of sprawl index and three airborne particulates, i.e., NO2, SO2, and CO. Due to data unavailability in the same study year, we exploit the data at the sub-district level in the closest year, 2018. The air particulates are calculated in mol per m² unit. Unfortunately, we could not directly define the urban-rural sub-sample at the sub-district since the categorization is conducted at the village level. So, we match the pollution data at the sub-district level with the IFLS sample and classify it into an urban-rural subsample. Fig. A3 shows that the sprawl index generally has a negative trend with the three air particulates, which implies that the areas with higher sprawl indexes are linked to lower

airborne particulates or higher air quality. The patterns are relatively similar between urban and rural areas (Fig. A4 – A.5). The NO2 seems to have the highest negative linkage with the sprawl index. The findings could explain how the sprawl index is correlated with the probability of having infectious respiratory symptoms.

4.4. Robustness check: different radii, method alternatives, observations omission

We perform a robustness check to see whether our results are affected in any way by identifying the relationship between the sprawl index and the likelihood of contracting respiratory infectious diseases in multiple aspects, which indicates that our results consistently maintain the negative association. Table 9, columns (1) and (2) present the estimation within a 1 km and 10 km radius. The coefficients' direction is similar to the baseline results, with a lower coefficient and statistical power in a 1 km radius and a higher coefficient in a 10 km radius than the baseline estimates.

We also test our estimation with a fixed-effect logistic model. The result in column (3) shows that the logistic fixed effect approach has a negative and significant coefficient comparable with our baseline estimates, despite the lower marginal effect value. The lower effect might be due to the unchanged outcome across waves causing considerable observation omission in panel logistic estimation, thus affecting the marginal effect value. Considering the potential endogeneity issue, we also estimate our model using the Lewbel internal instrument following Baum and Lewbel (2019). The result in column (4) gives similar

evidence with a smaller coefficient than the baseline estimate. It makes sense as the generated instrument yields higher standard errors affecting the coefficient's magnitude (Baum & Lewbel, 2019). Additionally, we observe the model by dropping the eldest age in each wave by 10% in column (5). The results are comparable to our baseline estimates that indicate the sprawler area has a lower rate of respiratory infectious diseases.

5. Discussion

Several studies identify the empirical evidence of how built environments are linked to health outcomes (Ewing et al., 2014; Zhao & Kaestner, 2010). But the study investigating the relationship between the built environments and the likelihood of contracting respiratory infectious diseases are scarce, particularly in Indonesia, with considerably built environment changes and persistent infectious diseases. The prior research on respiratory infectious diseases in Indonesia mainly focused on the socioeconomic and health behaviour determinants of respiratory infectious disease and neglected the role of the neighbourhood. The previous study is also limited in providing the mediator on how the built environment could correlate with health outcomes and identifying the finer built environment measurement. Thus, this study resolves those limitations by investigating the quantitative relationship between the built environment and respiratory infectious diseases and measuring quantitative calculations of potential channels on how the built environment relates to respiratory infectious diseases.

Our research estimates the built environments via a sprawl index

Table 9
The relationship between sprawl and respiratory infectious disease on various radii, methods, and sample.

Dependent variable: respiratory infectious diseases	Radii		Method alternative		Dropping Sample
	1 km	10 km	panel logistic FE	Lewbels internal instrument	drop 10% the eldest
	(1)	(2)	(3)	(4)	(5)
Sprawl index 1 km radius	-0.0009 (0.0006)				
Sprawl index 5 km radius			-0.0302*** (0.0105)	-0.0006*** (0.0002)	-0.0028*** (0.0009)
Sprawl index 10 km radius		-0.0042*** (0.0011)			
Marginal effect		-0.0011			
Age	-0.0045 (0.0038)	-0.0050 (0.0038)	-0.0558 (0.0402)	-0.0016 (0.0012)	-0.0044 (0.0044)
Age-squared	0.00003 (0.00002)	0.00003 (0.00002)	0.0003 (0.0002)	0.00001 (0.00001)	0.00001 (0.00002)
Years of schooling	0.0044 (0.0029)	0.0044 (0.0029)	0.0565 (0.0349)	-0.0016* (0.0009)	0.0048 (0.0029)
Gender = male				0.0044 (0.0054)	
Marital = married	-0.0041 (0.0121)	-0.0047 (0.0122)	-0.0661 (0.1310)	-0.0117* (0.0066)	-0.0114 (0.0132)
Working = yes	0.0046 (0.0089)	0.0049 (0.0089)	0.0380 (0.1010)	0.0005 (0.0064)	0.0059 (0.0094)
Number of household members	-0.0015 (0.0017)	-0.0015 (0.0017)	-0.0208 (0.0208)	0.0006 (0.0009)	-0.0015 (0.0016)
Years of schooling of the household head	-0.0017 (0.0013)	-0.0016 (0.0013)	-0.0160 (0.0159)	-0.0005 (0.0007)	-0.0019 (0.0014)
Constant	0.2960** (0.1410)	0.5940*** (0.1690)		0.2520*** (0.0352)	0.4540*** (0.1670)
Wave fixed effect	Yes	Yes	Yes	Yes	Yes
Wave-island fixed effect	Yes	Yes	Yes	No	Yes
Sample	Non-migrant	Non-migrant	Non-migrant	Non-migrant	Non-migrant
Mean of dep variable	0.111	0.111		0.111	0.112
Adjusted R ²	0.007	0.008		0.009	0.008
Observation	19748	19748	3795	19748	17977

Notes: The sample is non-migrant individuals aged 15+ from three waves of the IFLS survey (2000, 2007, and 2014), with an outcome equal to 1 if the individual showed non-specific acute respiratory infection or pneumonia symptoms or consumed TB medication or diagnosed with TB. Standard errors clustered at the sub-district level were reported in parentheses. A. Asterisks denote significance: ***p < 0.01, **p < 0.05, *p < 0.10.

calculated from land cover satellite data. We apply a fixed effect method to accommodate time-invariant unobservable factors and use the IFLS longitudinal data to deal with time-varying confounders. Controlling time-invariant and time-variant unobservable variables makes our results closer to causal identification than simple OLS on cross-sectional data. The results find that a 1% increase in sprawl index generally is related to a lower likelihood of contracting respiratory infectious diseases, about 0.25%. The magnitudes vary across areas and gender, with urban areas and female sub-samples showing more evidence than rural areas and male sub-samples. The higher relationship in urban areas could be caused by the compactness level between urban and rural areas, with urban areas being more compact than rural areas. Also, the differences in commuting duration and the number of modes of transportation between gender might explain the higher association between sprawl and infectious diseases in the female sample. Commuter survey data of Jakarta Metropolitan Areas in 2014 shows that the average commuting duration between males and females aged 15+ is similar, 62 min for males and 60 min for females. When estimating the commuting's average number of transportation modes, female commuters have a higher average number of transportation modes than male commuters, almost 3 for females and 2 for males (Statistics of Indonesia, 2014). In disease transmission, the number of transportation modes taken might explain why females have a higher prevalence of respiratory infectious diseases. The findings also correspond to Groeneveld et al. (2021), revealing that females have higher respiratory symptoms.

Despite the lower coefficient, the result is consistent with previous research, which found that a compact built environment is positively related to a higher rate of respiratory infectious diseases (B. Li, Ma, & Zhang, 2021; Yip, Huang, & Liang, 2021). The study provides several robustness tests to check the result's reliability by applying different radii of the sprawl index, employing alternative methods, and omitting the eldest 10% of observations. The results suggest that the relationship is more robust with a wider radius and lower age range. The study also reveals that compact neighbourhoods benefit from better health facilities but are insufficient compared to the population. The current Puskesmas-population ratio (1:20000) is under the government minimum standard (1:16000), and among Southeast Asia countries, Indonesia has the second lowest ratio of physicians and population (World Bank, 2019).

However, the study does not suggest that the sprawling-built environment is better than the compactly built environment in the context of contamination of infectious diseases. Instead, densification policies should consider health effects. Our results could be subjects of public health intervention. The built environment, population characteristics, neighbourhood infrastructure, individual characteristics, and socioeconomic status must be simultaneously considered when deploying non-pharmacological public health intervention. Non-pharmaceutical interventions, such as improving housing quality, particularly sufficient housing ventilation and capacity. Promoting better neighbouring quality by changing the population behaviour that produces fewer pollutants, such as using electricity wisely and using or riding eco-friendly transportation, particularly in compact neighbourhoods, could prevent people from respiratory infectious diseases. Also, the densification policy should be accompanied by adequate and equal health facilities to prevent and control diseases. The government should consistently implement a minimum housing floor size policy, particularly in dense areas. So far, the government of Indonesia has issued the regulation related minimum floor size policy, but unfortunately, the implementation is still weak. The government could improve public transportation policies that support cleaner air and fewer physical contact.

This study also has limitations. Firstly, the study might contain measurement errors when identifying the relationship between sprawl

index and TB prevalence. TB could be systematically undiagnosed due to limited access to health facilities which might be associated with the sprawl index. Secondly, we could not provide individual transportation mode choices, impacting individuals' respiratory diseases at the sub-district level since the data is available at the district level. The cross-district movement behaviour might differ from sub-district movement behaviour; thus, including the mobility across districts could give misleading results. Thirdly, self-reported health data used in this study might be less accurate than medical or prescription reports. Those limitations are beyond our study, and we left them for future research improvement.

6. Conclusion

The study confirms that a sprawling neighbourhood is linked to lower respiratory infection symptoms. The link is more evident in urban areas and for females. The study also suggested that the linkage works through housing quality, such as housing crowdedness and ventilation, and neighbourhood conditions like neighbourhood transportation modes and air pollution levels. Sprawl neighbourhood is associated with less crowded housing and better housing ventilation, but the association is insignificant. Sprawl neighbourhoods are also related to higher air quality and single-occupancy vehicles used. Those sprawl neighbourhood characteristics could prevent people from contracting respiratory infectious diseases. Even though compact neighbourhood benefits from better health facilities, the quantity is still limited compared to the population. Thus, our results underline the need to consider the health consequences of the densification policy and determine the direction of landscape planning and policy.

Ethical statement

The authors used publicly available deidentified data for this analysis, thus this study has no ethical issues.

Funding and acknowledgements

Indriyani acknowledges the Indonesia Endowment Fund for Education/Lembaga Pengelola Dana Pendidikan (LPDP), Ministry of Finance, the Republic of Indonesia, for providing financial support for this study. The funding source had no role in the study design, data collection, data analysis, data interpretation or preparation of the manuscript.

Author statement

Witri Indriyani: conceptualization and design, methodology, analysis and interpretation, writing-original draft.

Muhammad Halley Yudhistira: conceptualization and design, methodology, analysis and interpretation, writing-original draft.

Prani Sastiono: conceptualization and design, methodology, analysis and interpretation.

Djoni Hartono: conceptualization and design, methodology, analysis and interpretation.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Appendix A. Variable definitions

Respiratory disease status: Equal to 1 if the individual has symptoms of one or more infectious respiratory diseases, i.e., non-specific acute respiratory infection, pneumonia, or tuberculosis. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Pneumonia symptom: individual respondent who has a fever, cough, and breathing difficulty during the last two weeks when surveyed. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Non-specific acute respiratory diseases symptom: individual respondent who has a fever, runny nose, and cough during the last two weeks when surveyed. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Tuberculosis diagnosis: individual respondents who consume medication for tuberculosis diseases. Source: Indonesia Family Life Survey (2007, 2014).

Sub-district sprawl index: share of the undeveloped area around the urban settlement in a sub-district area (%). The higher the index, the more compact the areas. Measured from the land-cover map in 2000, 2007, 2014, and 2018 from the European Space Agency (ESA) Climate Change Initiative (CCI) and the digital administrative map of Indonesia 2017 from Statistics of Indonesia. Source: ESA CCI (2000, 2007, 2014, 2018); Statistics of Indonesia, BPS (2017).

Age: Individual respondents' age when surveyed. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Years of schooling: Individual total years of formal education. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Marital status: Equal to 1 if the individual respondent is married or living together, 0 otherwise. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Working status: Equal to 1 if the individual respondent is working, 0 otherwise. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Number of household members: number of individuals who reside in the same household. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Years of schooling of the household head: total formal education years of the head of household. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Housing crowdedness: the number of household members divided by floor size area. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Public transport: the availability of public three-wheeled, four-wheeled, or motorboat services in the village. Equal to 1 if available, 0 otherwise. Source: Indonesia Family Life Survey (2000, 2007, 2014).

Transport mode to the station, bus stop, terminal: The type of transportation mode used to reach the nearest station, bus stop, terminal, or pier. Equal to 1 if using two-wheeled motor vehicles, pedicabs, bicycles, and foot, and 0 otherwise.

Transport mode to market: The type of transportation mode used to reach the nearest market. Equal to 1 if using two-wheeled motor vehicles, pedicabs, bicycles, and foot, and 0 otherwise.

Table A.1

Descriptive statistics for all samples

	All			≤ median of sprawl index 5 km radius			> median of sprawl index 5 km radius		
	obs.	mean/%	s.d.	obs.	mean/%	s.d.	obs.	mean/%	s.d.
Dependent variables									
Respiratory infectious diseases									
- Not showing symptom	23490	88.84%	–	11582	87.60%	–	11908	90.08%	–
- Showing symptom	2952	11.16%	–	1640	12.40%	–	1312	9.92%	–
Pneumonia									
- Not showing symptom	25815	97.66%	–	12869	97.36%	–	12946	97.96%	–
- Showing symptom	619	2.34%	–	349	2.64%	–	270	2.04%	–
Acute respiratory infection = yes									
- Not showing symptom	23624	89.37%	–	11654	88.17%	–	11970	90.57%	–
- Showing symptom	2810	10.63%	–	1564	11.83%	–	1246	9.43%	–
Tuberculosis									
- Not showing symptom	17916	99.54%	–	9081	99.51%	–	8835	99.58%	–
- Showing symptom	82	0.46%	–	45	0.49%	–	37	0.42%	–
Independent variables									
Sprawl radius 5 km	26442	75.08	30.00	13222	53.12	28.75	13220	97.04	3.02
Individual characteristics									
Marital status = yes									
- Unmarried	5906	22.34%	–	3047	23.04%	–	2859	21.63%	–
- Married	20536	77.66%	–	10175	76.96%	–	10361	78.37%	–
Working status = yes									
- Not working	7231	27.35%	–	4033	30.50%	–	3198	24.19%	–
- Working	19211	72.65%	–	9189	69.50%	–	10022	75.81%	–
Age	26442	42.06	14.61	13222	41.14	14.28	13220	42.98	14.88
Years of schooling	26442	7.81	4.68	13222	8.59	4.59	13220	7.03	4.64
Household characteristics									
Number of household members	26442	5.85	2.83	13222	5.98	2.99	13220	5.71	2.65
Years of schooling of the household head	26442	7.39	4.84	13222	8.17	4.80	13220	6.61	4.76

Table A.2
Descriptive statistics for the non-migrant sample

	All			≤ median of sprawl index 5 km radius			> median of sprawl index 5 km radius		
	obs.	mean/%	s.d.	obs.	mean/%	s.d.	obs.	mean/%	s.d.
Dependent variables									
Respiratory infectious diseases									
- Not showing symptom	17552	88.88%	–	8644	87.45%	–	8908	90.32%	–
- Showing symptom	2196	11.12%	–	1241	12.55%	–	955	9.68%	–
Pneumonia									
- Not showing symptom	19277	97.63%	–	9616	97.29%	–	9661	97.97%	–
- Showing symptom	468	2.37%	–	268	2.71%	–	200	2.03%	–
Acute respiratory infection = yes									
- Not showing symptom	17649	89.38%	–	8702	88.04%	–	8947	90.73%	–
- Showing symptom	2096	10.62%	–	1182	11.96%	–	914	9.27%	–
Tuberculosis									
- Not showing symptom	14581	99.58%	–	7326	99.51%	–	7255	99.64%	–
- Showing symptom	62	0.42%	–	36	0.49%	–	26	0.36%	–
Independent variables									
Sprawl radius 5 km	19748	76.03	29.88	9885	54.69	29.41	9863	97.42	2.69
Individual characteristics									
Marital status = yes									
- Unmarried	4028	20.40%	–	2107	21.32%	–	1921	19.48%	–
- Married	15720	79.60%	–	7778	78.68%	–	7942	80.52%	–
Working status = yes									
- Not working	4866	24.64%	–	2803	28.36%	–	2063	20.92%	–
- Working	14882	75.36%	–	7082	71.64%	–	7800	79.08%	–
Age	19748	43.98	13.67	9885	43.12	13.44	9863	44.84	13.84
Years of schooling	19748	7.49	4.68	9885	8.31	4.61	9863	6.67	4.60
Household characteristics									
Number of household members	19748	6.10	2.80	9885	6.28	2.96	9863	5.91	2.63
Years of schooling of the household head	19748	7.11	4.79	9885	7.90	4.75	9863	6.33	4.70

Table A.3
The relationship between sprawl and respiratory infections, weighted baseline estimates

	All sample		Non-migrant sample							
	respiratory infection		respiratory infection		non-specific acute respiratory infection		pneumonia		TB	
	cross-section person weight	longitudinal person weight	cross-section person weight	longitudinal person weight	cross-section person weight	longitudinal person weight	cross-section person weight	longitudinal person weight	cross-section person weight	longitudinal person weight
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sprawl index 5 km radius	−0.0024*** (0.0008)	−0.0024** (0.0010)	−0.0034*** (0.0009)	−0.0034*** (0.0010)	−0.0034*** (0.0009)	−0.0033*** (0.0010)	−0.0009** (0.0004)	−0.0009** (0.0004)	−0.0001 (0.0003)	−0.0002 (0.0004)
Age	−0.0030 (0.0032)	−0.0012 (0.0045)	−0.0043 (0.0041)	−0.0039 (0.0054)	−0.0043 (0.0041)	−0.0035 (0.0055)	−0.00001 (0.0025)	−0.0006 (0.0034)	−0.0009 (0.0028)	−0.0023 (0.0028)
Age-squared	0.00002 (0.00002)	−0.00001 (0.00002)	0.00002 (0.00002)	0.00001 (0.00003)	0.00003 (0.00002)	0.00001 (0.00003)	0.00002 (0.00001)	0.00002 (0.00001)	−0.000001 (0.00001)	0.000003 (0.00001)
Years of schooling	0.0025 (0.0022)	−0.0002 (0.0027)	0.0044 (0.0029)	0.0013 (0.0039)	0.0042 (0.0028)	0.0010 (0.0038)	0.0005 (0.0014)	−0.0002 (0.0017)	−0.0001 (0.0010)	0.0004 (0.0011)
Marital = married	−0.0123 (0.0097)	−0.0226** (0.0105)	−0.0069 (0.0128)	−0.0178 (0.0138)	−0.0047 (0.0128)	−0.0156 (0.0136)	0.0006 (0.0055)	0.0023 (0.0057)	−0.0045 (0.0045)	−0.0045 (0.0048)
Working = yes	0.0089 (0.0082)	0.0051 (0.0094)	0.0092 (0.0105)	0.0108 (0.0113)	0.0093 (0.0101)	0.0111 (0.0111)	−0.0037 (0.0043)	−0.0057 (0.0048)	0.0009 (0.0042)	−0.0024 (0.0050)
Number of household members	−0.0009 (0.0011)	−0.0003 (0.0014)	−0.0022 (0.0017)	−0.0017 (0.0020)	−0.0028* (0.0016)	−0.0025 (0.0020)	−0.0001 (0.0009)	0.00001 (0.0008)	0.0005 (0.0006)	0.0006 (0.0007)
Years of schooling of the household head	0.00005 (0.0010)	−0.0006 (0.0012)	−0.0019 (0.0013)	−0.0026 (0.0017)	−0.0019 (0.0013)	−0.0026 (0.0017)	0.0003 (0.0006)	0.0002 (0.0006)	0.0003 (0.0003)	0.0004 (0.0004)
Constant	0.4420*** (0.1310)	0.4330** (0.1720)	0.5390*** (0.1650)	0.5710*** (0.2060)	0.5300*** (0.1640)	0.5470*** (0.2070)	0.0640 (0.0974)	0.0893 (0.1230)	0.0524 (0.1310)	0.1070 (0.1270)
Wave fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave-island fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dep variable	0.108	0.111	0.106	0.110	0.102	0.105	0.023	0.022	0.004	0.004
Adjusted R ²	0.009	0.011	0.009	0.011	0.008	0.009	0.006	0.005	0.003	0.004
Observation	26436	20769	19743	16148	19740	16145	19740	16145	14640	11444

Notes: The estimated use inverse probability weighting and use cross-section person weight year 2000 for all years (waves) and longitudinal person weight. Cross-section person weight in IFLS3 is constructed so that the estimates will represent the Indonesian population living in the 13 IFLS provinces at the time of IFLS3 in

2000. The longitudinal person weight is constructed so that the IFLS3 panel sample is representative of the Indonesian population living in the 13 IFLS provinces in 1993. Standard errors clustered at the sub-district level were reported in parentheses. Asterisks denoted significance: ***p < 0.01, **p < 0.05, *p < 0.10.

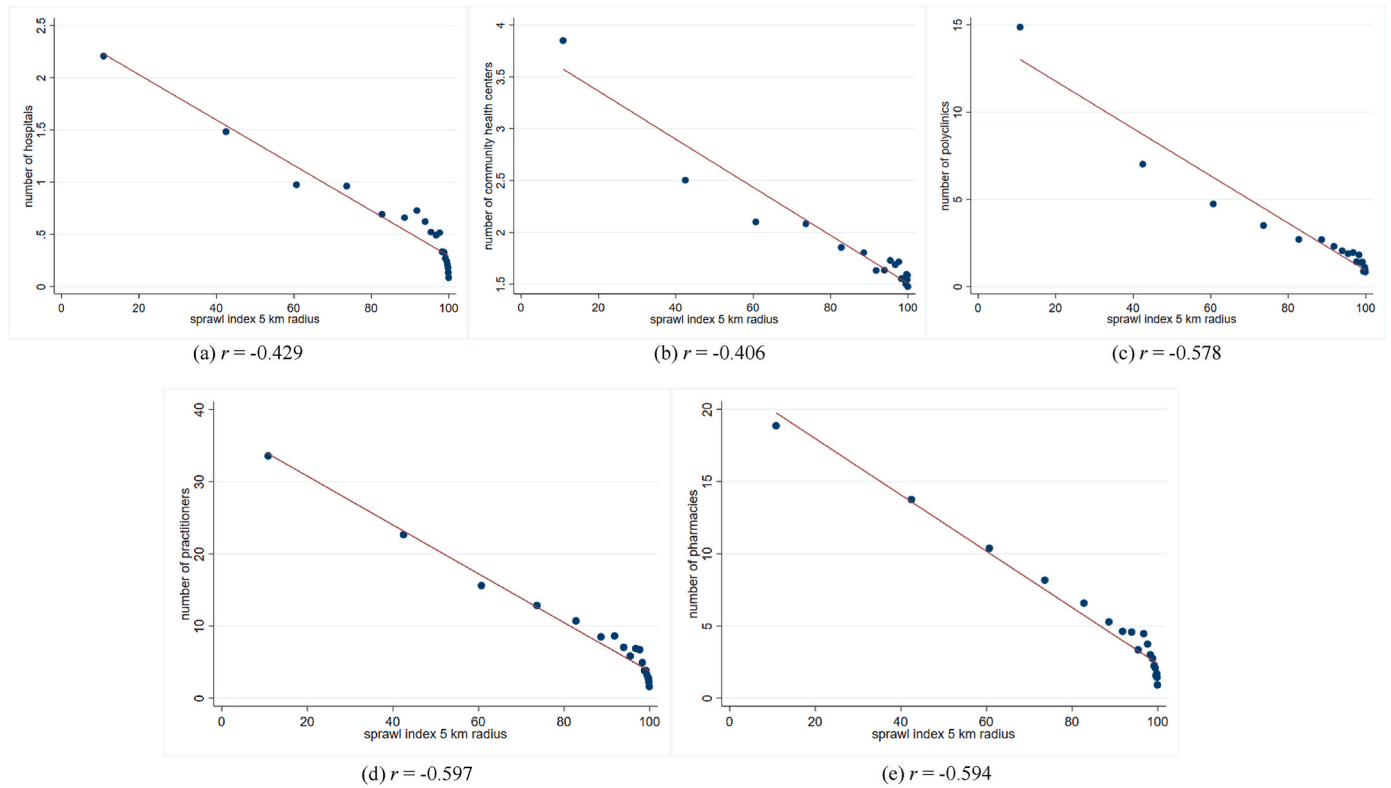


Fig. A1. The association between urban sprawl and health facilities

Notes: Fig. A1 is a binned scatter plot of sprawl index and health facilities at the sub-district level. The binned scatter group the sprawl index into 20 equally sized bins. *r* is the Pearson correlation coefficient. Source of health facilities data: National Village Survey 2014, Statistics of Indonesia.

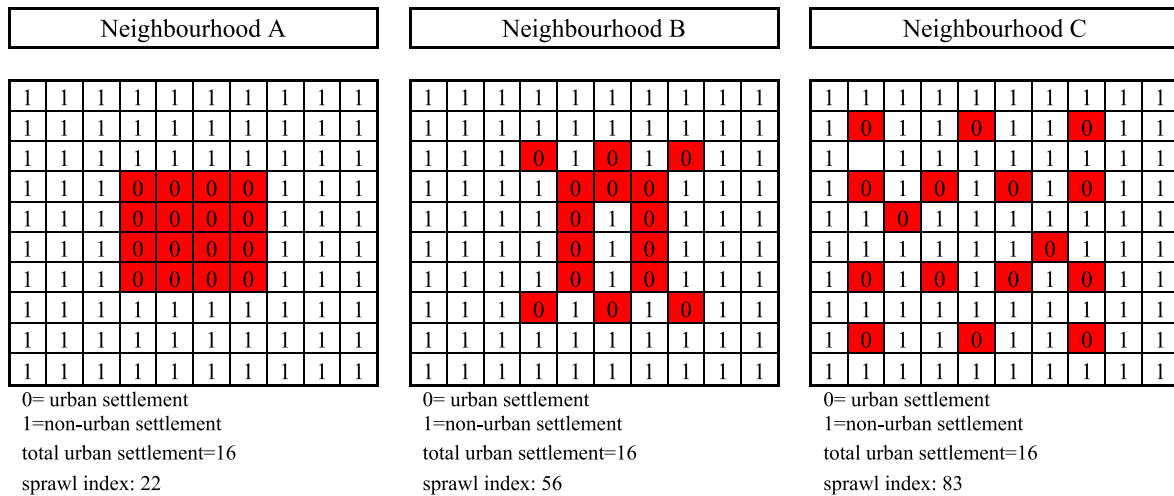


Fig. A2. Illustration of built environment form.

Fig. A2 shows the example pattern of the neighbourhoods and their sprawl index calculation. One big square is a sub-district area consisting of 100 grid cells, and each neighbourhood has 16 grid cells of urban settlements. Calculating the pattern or density by simply averaging the total urban settlement with total areas (grid cells) would result in the same value (16). Yet, they had a different form, where neighbourhood A had a more compact shape than neighbourhoods B and C. Thus, simple density calculation might give misleading information. Meanwhile, we got different values if calculating the sprawl index suggested by Burchfield et al. (2006). The measurement was as follows: For each urban settlement (0 code), we summed up the non-urban settlements or undeveloped areas in a certain radius, for example, one block adjacent to the urban cell, and averaged the value to get the sprawl index. Sprawl index calculation showed that neighbourhood A has a lower value (22) than neighbourhood B (56) and C (83). It means that neighbourhood A is less sprawled or more compact than B and C. Therefore, we constructed the sprawl index to identify the built environment form instead of a simple density.

We employed a land cover map from ESA CCI in the same year with the IFLS data. The map has a high spatial resolution of 300 m × 300 m, allowing us to construct at the sub-district level. The map is projected with Indonesia’s digital map to get Indonesia’s land cover map. Then we reclassified urban and non-urban settlements around the neighbourhood and defined the preferred radius. Based on urban settlement information in a certain

radius, we could calculate the percentage of developed and undeveloped around the urban settlement grid cells. The sprawl index is obtained by averaging the ratio of the undeveloped areas around all urban settlement grid cells in one sub-district.

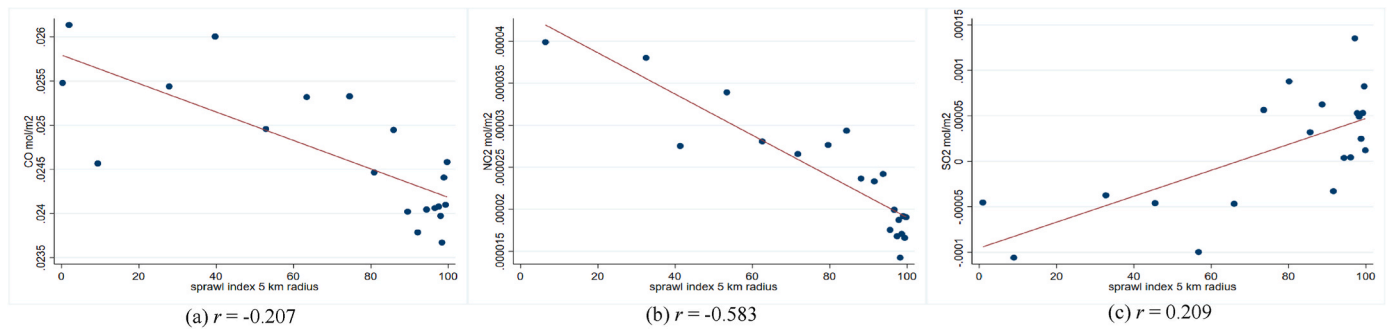


Fig. A3. The association between urban sprawl and air pollution, all samples

Notes: Fig. A3 is a binned scatter plot of sprawl index and pollution at the sub-district level. The binned scatter group the sprawl index into 20 equally sized bins. r is the Pearson correlation coefficient. Source: TROPOMI Sentinel-5 Precursor, 2018.

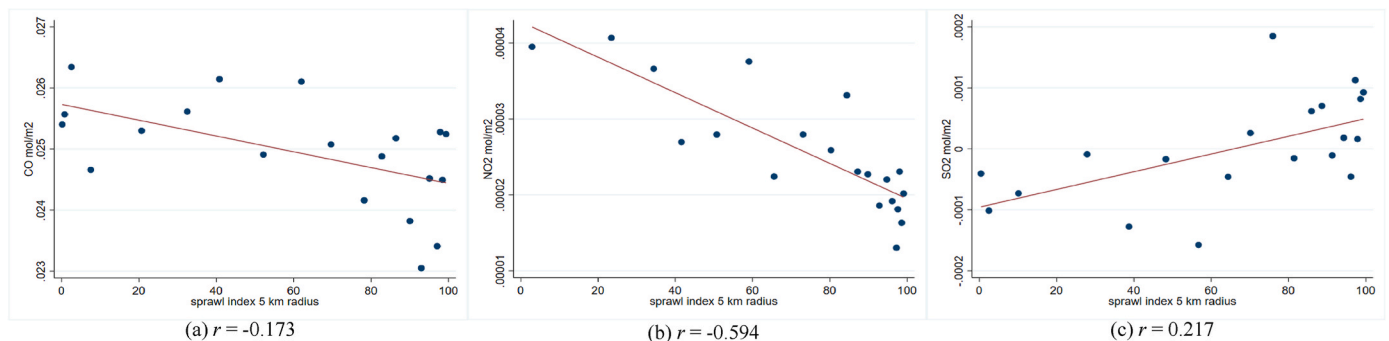


Fig. A4. The association between urban sprawl and air pollution, urban sample

Notes: Fig. A4 is a binned scatter plot of sprawl index and pollution at the sub-district level. The binned scatter group the sprawl index into 20 equally sized bins. r is the Pearson correlation coefficient. Source: TROPOMI Sentinel-5 Precursor, 2018.

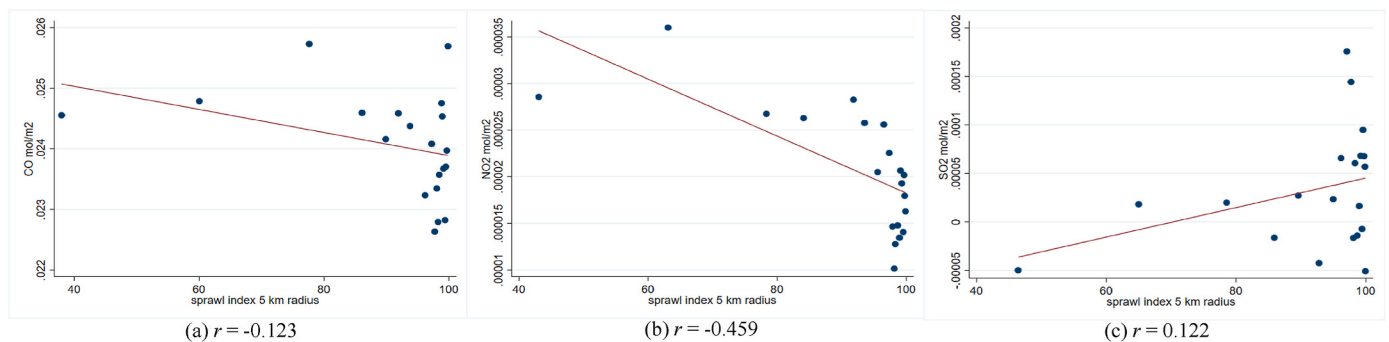


Fig. A5. The association between urban sprawl and air pollution, rural sample

Notes: Fig. A5 is a binned scatter plot of sprawl index and pollution at the sub-district level. The binned scatter group the sprawl index into 20 equally sized bins. r is the Pearson correlation coefficient. Source: TROPOMI Sentinel-5 Precursor, 2018.

References

- Antunes, J. L. F., & Waldman, E. A. (2001). The impact of AIDS, immigration and housing overcrowding on tuberculosis deaths in Sao Paulo, Brazil, 1994-1998. *Social Science & Medicine*, 52(7), 1071-1080. [https://doi.org/10.1016/S0277-9536\(00\)00214-8](https://doi.org/10.1016/S0277-9536(00)00214-8)
- Barbara, A. M., Loeb, M., Dolovich, L., Brazil, K., & Russell, M. (2012). Agreement between self-report and medical records on signs and symptoms of respiratory illness. *Primary Care Respiratory Journal*, 21(2), 145-152. <https://doi.org/10.4104/pcrj.2011.00098>
- Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *STATA Journal*, 19(4), 757-767. <https://doi.org/10.1177/1536867X19893614>
- Bereitschaft, B., & Debbage, K. (2013). Urban form, air pollution, and CO2 emissions in large U.S. Metropolitan areas. *The Professional Geographer*, 65(4), 612-635. <https://doi.org/10.1080/00330124.2013.799991>
- Bhadra, A., Mukherjee, A., & Sarkar, K. (2021). Impact of population density on Covid-19 infected and mortality rate in India. *Modeling Earth Systems and Environment*, 7(1), 623-629. <https://doi.org/10.1007/s40808-020-00984-7>

- Brinkhues, S., Schram, M. T., Hoebe, C. J. P. A., Kretzschmar, M. E. E., Koster, A., Dagnelie, P. C., et al. (2018). Social network in relation to self-reported symptomatic infections in individuals aged 40-70 - the Maastricht study-. *BMC Infectious Diseases*, 18(300). <https://doi.org/10.1186/s12879-018-3197-3>
- Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of sprawl: A portrait from space. *Quarterly Journal of Economics*, 121(2), 587-633. <https://doi.org/10.1162/qjec.2006.121.2.587>
- Carozzi, F., Provenzano, S., & Roth, S. (2020). *Urban density and covid-19. CEP Discussion Paper* (No. 1711).
- Giencewicki, J., & Jaspers, I. (2007). Air pollution and respiratory viral infection. *Inhalation Toxicology*, 19(14), 1135-1146.
- Domingo, J. L., & Rovira, J. (2020). Effects of air pollutants on the transmission and severity of respiratory viral infections. *Environmental Research*, 187, 1-7. <https://doi.org/10.1016/j.envres.2020.109650>
- Drewnowski, A., Arterburn, D., Zane, J., Aggarwal, A., Gupta, S., Hurvitz, P. M., et al. (2019). The moving to health (M2H) approach to natural experiment research: A paradigm shift for studies on built environment and health. *SSM - Population Health*, 7. <https://doi.org/10.1016/j.ssmph.2018.100345>

- Emeruwa, U. N., Ona, S., Shaman, J. L., Turitz, A., Wright, J. D., Gyamfi-Bannerman, C., et al. (2020). Associations between built environment, neighborhood socioeconomic status, and SARS-CoV-2 infection among pregnant women in New York city. *JAMA*, 324(4), 390–392. <https://doi.org/10.1001/jama.2020.11370>
- Ewing, R., Meakins, G., Hamidi, S., & Nelson, A. C. (2014). Relationship between urban sprawl and physical activity, obesity, and morbidity - update and refinement. *Health & Place*, 26, 118–126. <https://doi.org/10.1016/j.healthplace.2013.12.008>
- Fathmawati, F., Rauf, S., & Indraswari, B. W. (2021). Factors related with the incidence of acute respiratory infections in toddlers in Sleman, Yogyakarta, Indonesia: Evidence from the Sleman health and demographic Surveillance system. *PLoS One*, 16(9 September), 1–13. <https://doi.org/10.1371/journal.pone.0257881>
- Galster, G. C. (2012). The mechanism(s) of neighborhood effects: Theory, evidence, and policy implications. In D. M. M. van Ham, D. Manley, N. Bailey, & L. Simpson (Eds.), *Neighbourhood effect research: New perspectives* (pp. 23–56). Springer.
- Glaeser, E. (2011). Cities, productivity, and quality of Life. *Science*, 333, 592–594.
- Groeneveld, J. M., Ballering, A. V., van Boven, K., Akkermans, R. P., Olde Hartman, T. C., & Uijen, A. A. (2020). Sex differences in incidence of respiratory symptoms and management by general practitioners. *Family Practice*, 37(5), 631–636. <https://doi.org/10.1093/FAMPRA/CMAA040>
- Hamidi, S., Ewing, R., & Sabouri, S. (2020). Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the United States. *Health & Place*, 64, 1–8. <https://doi.org/10.1016/j.healthplace.2020.102378>
- Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *Journal of the American Planning Association*, 86(4), 495–509. <https://doi.org/10.1080/01944363.2020.1777891>
- Hankey, S., & Marshall, J. D. (2017). Urban form, air pollution, and health. *Current Environmental Health Reports*, 4, 491–503. <https://doi.org/10.1007/s40572-017-0167-7>
- Hu, M., Roberts, J. D., Azevedo, G. P., & Milner, D. (2021). The role of built and social environmental factors in covid-19 transmission: A look at America's capital city. *Sustainable Cities and Society*, 65, 1–14. <https://doi.org/10.1016/j.scs.2020.102580>
- Li, S., Ma, S., & Zhang, J. (2021). Association of built environment attributes with the spread of COVID-19 at its initial stage in China. *Sustainable Cities and Society*, 67, 1–14. <https://doi.org/10.1016/j.scs.2021.102752>
- Li, B., Peng, Y., He, H., Wang, M., & Feng, T. (2021). Built environment and early infection of COVID-19 in urban districts: A case study of Huangzhou. *Sustainable Cities and Society*, 66, 1–10. <https://doi.org/10.1016/j.scs.2020.102685>
- Li, F., & Zhou, T. (2019). Effects of urban form on air quality in China: An analysis based on the spatial autoregressive model. *Cities*, 89, 130–140. <https://doi.org/10.1016/j.cities.2019.01.025>
- Luongo, J. C., Fennelly, K. P., Keen, J. A., Zhai, Z. J., Jones, B. W., & Miller, S. L. (2016). Role of mechanical ventilation in the airborne transmission of infectious agents in buildings. *Indoor Air*, 26, 666–678. <https://doi.org/10.1111/ina.12267>
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(4), 303–328. [https://doi.org/10.1016/0047-2727\(74\)90003-6](https://doi.org/10.1016/0047-2727(74)90003-6)
- Miandad, M., Anwar, M. M., Ahmed, S., Rahman, G., & Khan, M. A. (2019). Assessment of risk factors associated with spread of tuberculosis in Gujrat city Pakistan. *Journal of Geography*, 39, 41–60. <https://doi.org/10.26650/JGEOG2019-0023>
- Mills, E. S. (1972). *Urban economics*. Foresman and Company.
- Ministry of Health. (2018). Riset Kesehatan Dasar 2018. In *Ministry of health, Republic of Indonesia*, 53. <https://doi.org/10.1017/CBO9781107415324.004>
- Mondal, D., & Paul, P. (2020). Effects of indoor pollution on acute respiratory infections among under-five children in India: Evidence from a nationally representative population-based study. *PLoS One*, 15(8 August), 1–13. <https://doi.org/10.1371/journal.pone.0237611>
- Murad, C., Dunne, E. M., Sudigdoadi, S., Fadlyana, E., Tarigan, R., Pell, C. L., et al. (2019). Pneumococcal carriage, density, and co-colonization dynamics: A longitudinal study in Indonesian infants. *International Journal of Infectious Diseases*, 86, 73–81. <https://doi.org/10.1016/j.ijid.2019.06.024>
- Muth, R. F. (1969). *Cities and housing*. University of Chicago Press.
- Nasim, N., El-Zein, A., & Thomas, J. (2022). A review of rural and peri-urban sanitation infrastructure in South-East Asia and the Western Pacific: Highlighting regional inequalities and limited data. *International Journal of Hygiene and Environmental Health*, 244. <https://doi.org/10.1016/j.ijheh.2022.113992>
- Oktaria, V., Danchin, M., Triasih, R., Soenarto, Y., Bines, J. E., Ponsonby, A. L., et al. (2021). The incidence of acute respiratory infection in Indonesian infants and association with vitamin D deficiency. *PLoS One*, 16(3 March), 1–18. <https://doi.org/10.1371/journal.pone.0248722>
- Pratama, A. P., Yudhistira, M. H., & Koomen, E. (2022). Highway expansion and urban sprawl in the Jakarta metropolitan area. *Land Use Policy*, 112. <https://doi.org/10.1016/j.landusepol.2021.105856>
- Regulation of the minister of Law and human Rights of republic of Indonesia number 34 the Year 2016 about district/City of human Rights care.(2016).
- Regulation of the minister of health of republic of Indonesia number 75 the Year 2014 about community health center.(2014).
- Saelens, B. E., & Handy, S. L. (2008). Built environment correlates of walking: A review. *Medicine&Science in Sports&Exercise*, 40(7). <https://doi.org/10.1249/MSS.0b013e31817c67a4>.Built
- She, Q., Peng, X., Xu, Q., Long, L., Wei, N., Liu, M., et al. (2017). Air quality and its response to satellite-derived urban form in the Yangtze River Delta, China. *Ecological Indicators*, 75, 297–306. <https://doi.org/10.1016/j.ecolind.2016.12.045>
- Statistics of Indonesia. (2014). Hasil Survei Komuter Jabodetabek 2014. In *Statistik Komuter Jabodetabek*.
- Stokols, D. (1996). Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion*, 10(4), 282–298. <https://doi.org/10.4278/0890-1171-10.4.282>
- Surya, B., Salim, A., Hernita, H., Suriani, S., Menne, F., & Rasyidi, E. S. (2021). Land use change, urban agglomeration, and urban sprawl: A Sustainable development perspective of Makassar city, Indonesia. *Land*, 10(556), 1–31. <https://doi.org/10.3390/land10060556>
- Villalonga-Olives, E., & Kawachi, I. (2017). The dark side of social capital: A systematic review of the negative health effects of social capital. *Social Science & Medicine*, 194, 105–127. <https://doi.org/10.1016/j.socscimed.2017.10.020>
- Wang, L., Chen, R., Sun, W., Yang, X., & Li, X. (2020). Impact of high-density urban built environment on chronic obstructive pulmonary disease: A case study of Jing'an district, Shanghai. *International Journal of Environmental Research and Public Health*, 17(252), 1–15. <https://doi.org/10.3390/ijerph17010252>
- Wang, S., Kou, C., Liu, Y., Li, B., Tao, Y., D'Arcy, C., et al. (2015). Rural-urban differences in the prevalence of chronic disease in northeast China. *Asia-Pacific Journal of Public Health*, 27(4), 394–406. <https://doi.org/10.1177/1010539514551200>
- Windi, R., Efendi, F., Qona'ah, A., Adnani, Q. E. S., Ramadhan, K., & Almutairi, W. M. (2021). Determinants of acute respiratory infection among under-five Years in Indonesia. *Journal of Pediatric Nursing*, 60. <https://doi.org/10.1016/j.pedn.2021.03.010>
- World Bank. (2019). Physicians (per 1,000 people). *World health organization's global health workforce statistics, OECD, Supplemented by country*.
- World Health Organization. (2021). World Health Organization. In *Global tuberculosis report 2021*.
- Xu, B., Tian, H., Sabel, C. E., & Xu, B. (2019). Impacts of road traffic network and socioeconomic factors on the diffusion of 2009 pandemic influenza A (H1N1) in mainland China. *International Journal of Environmental Research and Public Health*, 16(7), 1–14. <https://doi.org/10.3390/ijerph16071223>
- Yip, T. L., Huang, Y., & Liang, C. (2021). Built environment and the metropolitan pandemic: Analysis of the COVID-19 spread in Hong Kong. *Building and Environment*, 188, 1–17. <https://doi.org/10.1016/j.buildenv.2020.107471>
- Yudhistira, M. H., Indriyani, W., Pratama, A. P., Sofiyandi, Y., & Kurniawan, Y. R. (2019). Transportation network and changes in urban structure: Evidence from the Jakarta metropolitan area. *Research in Transportation Economics*, 74, 52–63. <https://doi.org/10.1016/j.retrec.2018.12.003>
- Zhang, Y., Ning, G., Chen, S., & Yang, Y. (2021). Impact of rapid urban sprawl on the local Meteorological observational environment based on remote sensing images and GIS technology. *Remote Sensing*, 13(2624), 1–17. <https://doi.org/10.3390/rs13132624>
- Zhao, Z., & Kaestner, R. (2010). Effects of urban sprawl on obesity. *Journal of Health Economics*, 29(6), 779–787. <https://doi.org/10.1016/j.jhealeco.2010.07.006>
- Ziersch, A. M., Baum, F., Darmawan, I. G. N., Kavanagh, A. M., & Bentley, R. J. (2009). Social capital and health in rural and urban communities in South Australia. *Australian & New Zealand Journal of Public Health*, 33, 7–16. <https://doi.org/10.1111/j.1753-6405.2009.00332.x>