Supplementary Material for the Study: Investigating the impact of Socio-Demographic and Urban Built Environments on the Spatial Heterogeneity of Flu Vaccine Locations – A Case Study from New York City, United States

S1 Spatial Lagged Endogenous Regression

The spatial lagged endogenous regression model is a widely used spatial model to account for spatial autocorrelation in the literature (Griffith 2000). More specifically, this model assumes that the number of locations providing flu vaccine in one NTA is influenced by places providing flu vaccine in the surrounding NTAs. The mathematical expression of the spatial lagged model is given in Eq. 1. We implemented the spatial lagged endogenous model using a Python package called PySAL (Rey and Anselin 2010).

$$y_i = \alpha + \lambda \sum_j w_{ij} y_i + \beta_i X_i + \epsilon_i \tag{1}$$

where:

- y_i is the response variable of NTA i, either indicating the number of locations providing flu vaccine in one NTA or the number of locations offering flu vaccine for children.
- w_{ij} is the spatial weight between NTA i and NTA j. This study defines the weight using the Queens Contiguity (Lloyd 2010).
- λ is the estimated spatial coefficient.
- X_i denotes a vector of values of independent variables of NTA i. β is also a vector of coefficients of independent variables.
- ϵ is the error term of regression.

S2 Geographically Weighted Regression (GWR)

The GWR model relaxes the assumption that the Ordinary Least Square (OLS) model holds, which assumes the relationship between dependent and independent variables remains the same across the study area. The GWR calibrates the regression to each NTA. The GWR model in this study can be expressed as in Eq. 2. We implemented the GWR module using a Python package called mgwr (Oshan et al. 2019).

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \epsilon_i a \tag{2}$$

where:

- y_i is the response variable of NTA i, either indicating the number of locations providing flu vaccine in one NTA or the number of locations offering flu vaccine for children.
- $\beta_0(u_i, v_i)$ is the intercept for NTA i. (u_i, v_i) represents the centroid of NTA i.
- $\beta_k(u_i, v_i)$ is the regression coefficient for the kth exploratory variable at NTA i.
- x_{ik} is the kth exploratory variable of NTA i.
- ϵ_i is the error term linked to NTA i.

The bandwidth (the number of neighboring NTA units for local estimation) can significantly influence the performance of the GWR model. We used Akaike Information Criterion (AIC) to choose the optimal value of bandwidth (Akaike 1998), ranging from 10 to the total number of NTAs.

S3 Multiscale Geographically Weighted Regression (MGWR)

The GWR model limits its capability by assigning the same bandwidth for each independent variable. MGWR is a GWR extension that relaxes this bandwidth assumption (Fotheringham, Yang, and Kang 2017). The MGWR model has mathematical expressions similar to

GWR but has a different strategy to compute the bandwidth for different independent variables. The MGWR can be defined as in Eq. 3. We implemented the MGWR model using the Python package mgwr (Oshan et al. 2019). We also used the AIC to choose the optimal value of bandwidth (Akaike 1998), ranging from 10 to the total number of NTAs.

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_{bwk}(u_i, v_i)x_{ik} + \epsilon_i \tag{3}$$

where bw in β_{bwk} indicates the bandwidth used in MGWR for the $k$the explanatory variable.

Reference

Akaike, Hirotogu (1998). "Information Theory and an Extension of the Maximum Likelihood Principle". In: *Selected Papers of Hirotugu Akaike*. Ed. by Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa. New York, NY: Springer New York, pp. 199–213. ISBN: 978-1-4612-1694-0. DOI: 10.1007/978-1-4612-1694-0_15. URL: https://doi.org/10.1007/978-1-4612-1694-0_15.

Fotheringham, A Stewart, Wenbai Yang, and Wei Kang (2017). "Multiscale Geographically Weighted Regression (MGWR)". In: *Annals of the American Association of Geographers* 107.6, pp. 1247–1265. ISSN: 2469-4452. DOI: 10.1080/24694452.2017.1352480. URL: https://www.tandfonline.com/doi/full/10.1080/24694452.2017.1352480.

Griffith, Daniel A (2000). "A linear regression solution to the spatial autocorrelation problem". In: *Journal of Geographical Systems* 2.2, pp. 141–156. ISSN: 1435-5930. DOI: 10.1007/PL00011451. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0033860634&doi=10.1007%2FPL00011451&partnerID=40&md5=a44dc2b7e836a043be0257bc6238712ehttp://link.springer.com/10.1007/PL00011451.

Lloyd, Christopher (2010). *Spatial Data Analysis: An Introduction for GIS Users*. English. Teaching or Research: 15147. Oxford University Press. ISBN: 978-0199554324.

Oshan, Taylor M. et al. (2019). "MGWR: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale". In: *ISPRS International Journal of Geo-Information* 8.6. ISSN: 22209964. DOI: 10.3390/ijgi8060269.

Rey, Sergio J and Luc Anselin (2010). "PySAL: A Python Library of Spatial Analytical Methods". In: *Handbook of Applied Spatial Analysis*. Ed. by Manfred M Fischer and Arthur Getis. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 175–193. ISBN: 978-3-642-03647-7. DOI: 10.1007/978-3-642-03647-7_11. URL: https://doi.org/10.1007/978-3-642-03647-7_11https://link.springer.com/10.1007/978-3-642-03647-7_11.