

Leveraging Open Data to Reconstruct the Singapore Housing Index and Other Building-level Markers of Socioeconomic Status for Health Services Research.

Daniel Yan Zheng Lim (✉ limydz@gmail.com)

Health Services Research Unit, Medical Board, Singapore General Hospital <https://orcid.org/0000-0002-9715-6970>

Ting Hway Wong

Singapore General Hospital

Mengling Feng

National University Singapore Saw Swee Hock School of Public Health

Marcus Eng Hock Ong

Singapore General Hospital

Andrew Fu Wah Ho

Singapore General Hospital

Research

Keywords: Socioeconomic status, social class, population characteristics, socioeconomic determinants of health, health inequality, Singapore, population health, health services research

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-428309/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Socioeconomic status (SES) is an important determinant of health, and SES data is an important confounder to control for in epidemiology and health services research. Individual level SES measures are cumbersome to collect and susceptible to biases, while area level SES measures may have insufficient granularity. The 'Singapore Housing Index' (SHI) is a validated, building level SES measure that bridges individual and area level measures. However, determination of the SHI has previously required periodic data purchase and manual parsing. In this study, we describe a means of SHI determination for public housing buildings with open government data, and validate this against the previous SHI determination method.

Methods

Government open data sources (e.g. data.gov.sg, Singapore Land Authority OneMAP API, Urban Redevelopment Authority API) were queried using custom Python scripts. Data on residential public housing block address and composition from the HDB Property Information dataset (data.gov.sg) was matched to postal code and geographical coordinates via OneMAP API calls. The SHI was calculated from open data, and compared to the original SHI dataset that was curated from non-open data sources in 2018.

Results

10077 unique residential buildings were identified from open data. OneMAP API calls generated valid geographical coordinates for all (100%) buildings, and valid postal code for 10012 (99.36%) buildings. There was an overlap of 10011 buildings between the open dataset and the original SHI dataset. Intraclass correlation coefficient was 0.999 for the two sources of SHI, indicating almost perfect agreement. A Bland-Altman plot analysis identified a small number of outliers, and this revealed 5 properties that had an incorrect SHI assigned by the original dataset. Information on recently transacted property prices was also obtained for 8599 (85.3%) of buildings.

Conclusion

SHI, a useful tool for health services research, can be accurately reconstructed using open datasets. This method provides more updated data at lower cost. This application highlights the potential for leveraging open data to enable healthcare research

Introduction

Socioeconomic status (SES) is a well-established determinant of health and is relevant to medical research and public health policy [1]. SES data is of high value to policy makers and researchers. In particular, they are frequently used to control for confounding in health services and epidemiology

research. However, obtaining individual-level SES measures (e.g. education, income and occupation) is challenging. The collection and handling of this sensitive data is limited by an increasing awareness of confidentiality issues, and its associated regulatory burden. These difficulties often mean that these data need to be collected repetitively for separate studies. Direct data collection of SES markers can also be vulnerable to biases such as recall bias and social desirability bias.

Area-level SES measures, such as census tract and neighbourhood-level studies, have thus emerged as alternatives to individual-level measures [2]. However, area measures that cover a large heterogeneous subpopulation are vulnerable to the ecological fallacy (wrongly classifying individuals' characteristics when using group level data). Therefore, a building-level marker of SES is a sensible compromise to bridge the limitations of area-level and individual-level SES data [3].

The Singapore Housing Index (SHI), also known as "room-index", is a Singapore-contextualised building-level, asset-based SES measure, first described by Wong et al. [4], and subsequently applied to a range of health services research. In brief, it classifies each building by the weighted average number of rooms of each unit and takes a value from 1 to 7 (larger values indicating higher SES). Housing property value provides a good surrogate measure of SES in Singapore, where the majority of the population dwell in public housing under a tiered subsidy scheme. The SHI applies to individuals who stay in the community (i.e. not in welfare homes for the destitute, or residential care homes for the aged). The utility of the SHI in local health services research has been demonstrated - both as a primary exposure of interest, and as a confounder to control for in clinical outcomes and healthcare utilization studies [4–10]. Examples included disease-specific cohorts (head and neck cancer, breast cancer), demographic subgroups (elderly), and both hospitalised and out-of-hospital settings. However, previous use of the SHI has required purchase of data from government agencies (e.g. Singapore Land Authority) both to construct and to update, due to urban renewal and construction. The exact methodology has varied with the availability of paid and open sources of data at the time of analysis, and has been an expensive undertaking, with recurrent expenditure for updated data. This poses a significant barrier to using this SES measure for research.

We hypothesized that new open data sources would provide an alternative means of obtaining SES data. Open data is defined as structured data that is machine-readable, freely shared, used and built on without restrictions [11]. It may be provided by government agencies or private bodies, and resides in the public domain. In Singapore, the majority of government open data is hosted at the National Open Data Portal (data.gov.sg) [12], which is administered by Smart Nation Singapore and has been live since 2018. Other government agencies may also provide open data on their respective websites or data portals. Updates to open data are immediately available to researchers and are free of charge, unlike periodically purchased data.

Open data may be accessed by direct download of datasets in the form of tables (e.g. comma-separated variables, CSV; tab-separated variables, TSV; Microsoft Excel worksheets, XLS). They may also be accessed by interacting with the action programming interfaces (APIs) of the respective data portals,

which are a software interface for data users to automatically request data from the open dataset via structured code queries.

In this paper, we describe a method of reconstructing the Singapore SHI from public-domain datasets provided by the Singapore Government (via data.gov.sg) and other government sources (Singapore Land Authority, Urban Redevelopment Authority, Ministry of Health, Attorney-General's Chambers). We aim to describe how the SHI and other building-level measures of socioeconomic status can be readily obtained from open data in Singapore, to improve accessibility of this tool to policy and research users. We compared the SHI obtained by this method with the original SHI dataset by Wong et al, which was curated from non-open data sources in 2018.

Methods

Data Retrieval Methods

Code for data retrieval was written in Python 3.6.9, using the Google CoLaboratory environment. Calls to open data APIs were made using the *requests* library and formatted into the standardised JavaScript Object Notation (JSON) format using the *json* library.

The data we sought to obtain related to housing block address, postal code and housing block composition. Other additional data that we deemed germane to SES included geographical coordinates (latitude and longitude) of the housing block, as well as recently transacted property prices. The data source of the respective data types is detailed in Table 1 below, and are classified into public housing (i.e. Housing Development Board (HDB) properties), private housing, and excluded addresses (i.e. welfare homes and residential care homes) to which the SHI does not apply. Open data sources included the HDB Property Information [13] and HDB Resale Flat Prices [14] datasets from data.gov.sg, the Singapore Land Authority (SLA) OneMAP API [15], the Urban Redevelopment Authority (URA) API [16]. Other public data sources were the Ministry of Health (MOH) HealthHub [17], and the Attorney-General's Chambers (AGC) Singapore Statutes Online (SSO) [18].

Table 1
Data source for various data types

	Public Housing	Private Housing	Excluded
Address	HDB Property Information dataset	Not available	Singapore Statutes Online (Destitute Homes) HealthHub (Nursing Homes)
Block Composition	HDB Property Information dataset	NA	NA
Postal Code	SLA OneMAP API	Not available	NA
Geographic Coordinates	SLA OneMAP API	SLA OneMAP API	NA
Recently Transacted Property Prices	HDB Resale Flat Prices dataset	URA API	NA
NA: Not Applicable			

The main data available in the public domain relates to public housing. We began with the HDB Property Information dataset via an API call to data.gov.sg, which contains exhaustive data on all HDB properties updated as of January 2021. This was filtered for residential properties only, with the block address and block composition (i.e. number of units of different size) available. No postal codes were available in this dataset. We hence processed the addresses into a structured text query, and made API calls to the OneMAP API to obtain the corresponding postal code and geographical coordinates.

As the HDB Property Information dataset is exhaustive and up to date, individuals whose address or postal code do not fall within it are not public housing residents. These individuals' addresses should be checked against an "Excluded Addresses Dataset" consisting of the exhaustive list of welfare homes (documented in the Destitute Persons Act [19]) and residential care homes in Singapore (documented in HealthHub [20]). An individual with an address or postal code that is in none of the above datasets is assumed to reside in private housing. The geographic coordinates of the given postal code or address can subsequently be queried from OneMAP.

An optional step is to verify that the given address or postal code is a residential property. This can be done via the URA API, but requires registration with a valid business email and a personalised token to access. This step is likely to be superfluous in Singapore, given that the National Registration Act [21] mandates all individuals with a registered identity card in Singapore (i.e. all residents in Singapore) to have a valid locally registered residential address.

Calculation of the SHI

The SHI ranges from 1 to 7. For public housing, the SHI described by Wong et al. [4] is derived by calculating the weighted mean number of rooms per apartment in the building. For each public housing block, SHI is given by the following formula: $\text{Sum}(\text{number of rooms in a flat} * \text{number of such flats per block}) / \text{total number of units in a block}$. This would yield an SHI value ranging from 1 to 6 for each public housing block. In the SHI formula calculation, the number of rooms in a flat is capped at 6 (i.e. executive condominiums, executive maisonettes, multigeneration flats, and any HDB category larger than a 5-room flat are all assigned a value of 6).

For the current open data formulation of the SHI, the exact composition of each public housing building is known, hence the SHI may be directly calculated. We note that the original formulation by Wong et al. had an algorithm for assigning SHI values where the exact composition of the block was not known, but this algorithm need not apply here. (The algorithm assigned SHI 1.5 to a block of rental flats where the relative proportion of 1-room and 2-room rental flats was unknown, and an SHI of 2.5 where the block was a mixture of 3-room flats and an unknown proportion of rental flats.)

The SHI for private housing described by Wong et al. used the Singapore Land Authority postal code master plan, and classified the postal codes labelled as private housing (no government subsidy) as SHI 6 (condominium apartment building) or 7 (landed residence such as bungalows or terrace houses). In the current open data formulation of the SHI, no open dataset exists that exhaustively identifies private housing, condominiums or landed residences, and the distinction between private housing buildings of SHI 6 and 7 cannot be made. We thus recommend that that all individuals whose addresses do not fall into the exhaustive HDB Property Information dataset, or the Excluded Addresses Dataset, be coded as SHI 6.

Retrieval of Other SES Data

Although previous work has focused on the SHI as a measure of building-level SES, we recognise that other researchers may wish to examine data on property transacted prices as a surrogate measure of the individual's asset value, as has been done in other countries [22]. This data may be extracted from the HDB Resale Flat Prices dataset (data.gov.sg) for public housing, and the Urban Redevelopment Authority API for private housing.

General Overview

The overall workflow for open data is summarised in Fig. 1.

Validation of Open Data SHI

We performed validation of the open data derived SHI dataset against the original SHI dataset by Wong et al (data was obtained from the authors of that study). Only public housing buildings were validated, as differentiation between condominiums (SHI 6) and landed residences (SHI 7) was not possible from open data.

We assessed the overlap of properties in the two datasets by matching via postal codes and addresses. For properties that were included in both datasets, we assessed agreement between the SHI derived from the two sources by examining the intraclass correlation coefficient (ICC), and Bland-Altman plot analysis. ICC was calculated using the *pingouin* package, and the Bland-Altman plots were generated with base Python.

We performed further validation by investigating cases where buildings could not be matched by address and postal code. These were manually triangulated against two other public data sources - SingPost Find Postal Code search [23], and Google Maps [24].

Results

Dataset Matching

Details of 21272 HDB buildings were extracted from data.gov.sg, of which 10077 unique buildings were labelled as residential. API calls to OneMAP were made for all residential buildings. Valid geographical coordinates were obtained for all (100%) buildings. 10012 (99.36%) of the buildings had a valid postal code, as 65 (0.64%) returned a NIL result.

We compared the overlap of buildings included in the open dataset and the original dataset by Wong et al, which had 10012 unique HDB residential buildings. Buildings were first matched by postal code and address, and 9888 buildings were successfully matched in this fashion. They were then matched by the address only, and 123 more buildings were matched in this manner. The total number of matched buildings was thus 10011.

A total of 67 buildings could not be matched between the two datasets. Of these buildings, 47 were HDB blocks that were newly constructed in 2019 and 2020. As the dataset by Wong et al had only been updated in 2018, we decide not to pursue further validation for these buildings. The remaining 20 buildings could not be matched by any other means. They were not in the 'Excluded Buildings' dataset as well. The overall process of matching is depicted in Fig. 2.

Dataset Validation

Three types of validation were performed. SHI was compared between the original and open datasets in Validation A. Validation B examined cases where buildings could not be matched between original and open datasets via postal code, and Validation C examined buildings that could not be matched between original and open datasets at all.

Validation A: Validation of SHI

We performed validation of the SHI for the 10011 properties that could be matched via the combination of postal code and address, and the address alone. SHI was calculated for all properties in the open dataset. We initially mirrored the formula in Wong et.al. In doing so, we noted two potential limitations to

the original calculation by Wong et al. - First, that studio flats were treated as having 3 rooms (in fact, they often only have 1 room); Second, that an uncommon type of 4 room rental flat ("*other_room_rental*") present in 9 housing blocks had been excluded in the original calculation, resulting in some blocks having an SHI of 0.

For the matched properties, we performed SHI validation by examining the intraclass correlation coefficient (ICC). This was 0.999, indicating near-perfect agreement. We also generated the Bland-Altman plot (Fig. 3), which also showed near-perfect agreement between the two sources of SHI, as evidenced by the majority of the points falling within the 2 standard deviation band for difference. However, we noted a small number of outliers, with an SHI difference of > 0.1 between original and open datasets. Manual investigation of these revealed 5 properties that had an incorrect SHI of 6 assigned by the original dataset.

In lieu of Wong's original formula, researchers in future may consider using an updated formula with the following changes:

- Excluded addresses are assigned an SHI of 0
- Studio flats are counted as having 1 room
- Flats in the "*other_room_rental*" category are counted as having 4 rooms

Validation B: Validation of Postal Codes

Of the 123 buildings matched by address only, 65 had no valid postal code returned by the OneMAP API. These buildings did have valid postal codes from SingPost and Google Maps, and we interpreted these as errors made by the OneMAP API.

10 buildings had two postal codes available as per SingPost. For these buildings, we interpret this as both original and open datasets being correct despite the discrepancy. 46 buildings had postal codes that only differed by the third digit. These indicate buildings with the same address, but have different postal codes because of multiple postal delivery points [25]. For these buildings, we manually verified that the text addresses were similar, and also interpret these situations as both original and open datasets being correct.

Finally, we noted two buildings where a postal code for a different building was returned by the OneMAP API. This was confirmed by referencing both SingPost and Google Maps. All discrepancies noted in this section were reported in writing to the Singapore Land Authority, and were forwarded to the OneMAP technical team for review.

Validation C: Validation of Unmatched Buildings

Of the 20 unmatched buildings, 1 was from the original dataset, and 19 were from the open dataset. By referencing SingPost and Google Maps, we confirmed that all 20 buildings existed, with a valid postal code and address. We interpret these as unintended omissions, suggesting that the open dataset had

omitted 1 building, and the original dataset had omitted 19 buildings. Of these 20 buildings, based on search engine results, all appeared to be residential addresses.

Recent Transaction Prices

8599 public housing blocks had resale transaction data in the last 2 years (2019 and 2020). We matched this data to the open SHI dataset using the building address and calculated the median resale price over the 2 years. The boxplots of median log resale price, stratified by the SHI for each building, is shown in Fig. 4. There is a clear monotonically increasing relationship between the SHI of a building and the overall resale prices.

Discussion

In this paper, we described a method of using open data to obtain the SHI of an individual based on their address or postal code. We provided example Python code in the Google CoLaboratory environment for other researchers to replicate the workflow and to demonstrate use of the respective open data APIs. These software tools are all open source and free to use. An SHI dataset created by this method can be readily linked with clinical datasets via the postal code or address, with the merged dataset subsequently utilized for data analysis.

In describing these methods, we hope to increase awareness of the potential in open data among health services and epidemiology researchers. The main advantages of open data lie in the convenience of use, containment of research costs, and better comprehensiveness in the case of government sourced open data. Previous healthcare research has also successfully leveraged local open data [26–29]. Furthermore, usage of APIs improves recency, and reduces chance of error as no human curation is required. In this study, we discovered a small number of cases where a wrong SHI had been assigned in the original dataset.

Conversely, use of open data means that researchers are limited to the data fields provided by the source. This limitation was encountered in the current methodological description, where private residential addresses were determined through a process of elimination, as there was no existing open dataset for this. For the current SHI workflow, such assumptions are reasonable given that public and private housing in Singapore are essentially mutually exclusive, but this limitation may constrain other applications of open data locally. We note that in the current SHI application, open data reduced our ability to differentiate private residential addresses into condominium (SHI 6) and landed property (SHI 7). However, we were able to identify destitute homes, which was a limitation of the original methodology. Future researchers may consider assigning a separate code (e.g. SHI 0) to this group of patients, or grouping them together with rental flat occupants. We were also able to identify residential nursing homes, and future researchers may wish to consider this as a separate group, especially for research involving older residents. Information on whether these are voluntary or private nursing homes is also available from the MOH HealthHub data source, if added granularity is desired.

Researchers using open data must trust the data source for veracity and completeness. This is an entirely reasonable assumption for local government sourced open data, given a stated commitment to providing timely and high-quality data [30]. Commercially sourced or other community sourced open data may not have such a commitment. In this study, we noted that the government sourced open data had valid results in the vast majority of cases, but did have a very small amount of data errors. For example, a NIL postal code was returned from the OneMAP API for some buildings. These were clarified with the data administrators of the relevant authority. As with any data source, researchers making use of open data need to perform validation checks prior to use, and we were able to resolve these cases by corroboration with other public government data sources.

In this study, we validated the SHI obtained from open data against the original SHI dataset by Wong TH et al. This showed near-perfect agreement and suggests that the open data version is practically equivalent to the manually curated version. We do however acknowledge that there are shortcomings to both methods. The current open dataset showed a small number of buildings with a wrong SHI computation by Wong et al. On the other hand, correct postal codes could not be retrieved for a small number of addresses via open data. These factors should be weighed by researchers who are contemplating either method of determining the SHI.

Other considerations for use of open data include research data governance - at present, usage of open data does not require Institutional Review Board (IRB) approval, as local IRBs do not have jurisdiction over data in the public domain. However, users of open data need to be familiar with the relevant licenses the data is provided under, and the acceptable terms of use. For example, the Singapore Open Data License [31] for data.gov.sg allows commercial and non-commercial use, but prevents users from assuming patent, trademark, or design rights.

Readers should also be aware that data in the public domain is not necessarily open data. In the current context of SHI determination, other information on property classification might be freely and publicly available on property agency websites. However, such websites are intended for human use and not for automated querying, and would generally not have APIs available. While data may still be programmatically obtained from such websites using web scraping software, this may not be the intention of the site owners and may be perceived as malicious online behaviour. Usage of web scraping tools is beyond the scope of this article, but we encourage fellow researchers to review the terms of use and the *robots.txt* file (a file describing acceptable use of automated web page retrieval for a given website) when interacting with web data sources that are not explicitly identified as open data.

The validity, strengths and limitations of SHI as a SES marker are beyond the scope of this study. The SHI can potentially be incorporated into composite indices using methods such as Principal Component Analysis. This approach of constructing building-level property-value indices as a SES marker could potentially be employed in a similar fashion outside Singapore.

Conclusion

We developed and described a workflow for re-constructing the Singapore Housing Index using open data. This provides a means for future researchers to obtain updated building-level markers of socioeconomic status for policy and research.

Declarations

Funding

AFWH was supported by the Estate of Tan Sri Khoo Teck Puat (Khoo Clinical Scholars Programme), Khoo Pilot Award (KP/2019/0034), Duke-NUS Medical School and National Medical Research Council (NMRC/CS_Seedfd/012/2018).

Ethical approval

Ethical approval was not required as all data used was derived from aggregated public sources and contained no individually identifying information.

Consent for publication

All authors give consent for publication.

Availability of data and materials

Data is available upon reasonable request.

Competing interests

Authors disclose no competing interests.

Authors' contribution

DYZL and AFWH conceived of the study. DYZL and WTH obtained the data. DYZL and FML conducted the analysis. Manuscript drafted and approved by all authors.

Acknowledgement

We are grateful for the help of Ms Tang Xuan Rong of Yong Loo Lin School of Medicine, National University of Singapore for coordinating the submission of the manuscript.

Supplementary Data

An example of the code used in this study may be viewed at:

https://colab.research.google.com/drive/1sdWtP4e-mlkDpU6pMdGrzpNvD_RJ8bBe?usp=sharing

References

1. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health*. 1997;18. doi:10.1146/annurev.publhealth.18.1.341
2. Earnest A, Ong ME, Shahidah N, Chan A, Wah W, Thumboo J. Derivation of indices of socioeconomic status for health services research in Asia. *Preventive medicine reports*. 2015;2. doi:10.1016/j.pmedr.2015.04.018
3. Havard S, Deguen S, Bodin J, Louis K, Laurent O, Bard D. A small-area index of socioeconomic deprivation to capture health inequalities in France. *Soc Sci Med*. 2008;67. doi:10.1016/j.socscimed.2008.09.031
4. Wong TH, Skanthakumar T, Nadkarni N, Van Nguyen H, Gopalakrishna Iyer N. Survival of patients with head and neck squamous cell carcinoma by housing subsidy in a tiered public housing system. *Cancer*. 2017. pp. 1998–2005. doi:10.1002/cncr.30557
5. Wong TH, Lau ZY, Ong WS, Tan KB, Wong YJ, Farid M, et al. Cancer patients as frequent attenders in emergency departments: A national cohort study. *Cancer Med*. 2018;7: 4434–4446.
6. Wei WE, Wong CH, Matchar DB, Earnest A, Wah W, Ong MEH, et al. Effect of Housing Type and Neighborhood Socioeconomic Indicators on Survival After Low Falls in Older Adults. *Journal of the American Medical Directors Association*. 2019. pp. 646–649. doi:10.1016/j.jamda.2018.12.016
7. Wong FY, Wong RX, Zhou S, Ong WS, Pek PP, Yap Y-S, et al. Effects of housing value and medical subsidy on treatment and outcomes of breast cancer patients in Singapore: A retrospective cohort study. *The Lancet Regional Health - Western Pacific*. 2021. p. 100065. doi:10.1016/j.lanwpc.2020.100065
8. Tan AL, Iyer NG, Putri N, Nadkarni N, Skanthakumar T, Wong TH, et al. Factors driving frequent attendance at emergency departments for patients with head and neck cancer. *Head & Neck*. 2019. pp. 3798–3805. doi:10.1002/hed.25916
9. Zhuang Q, Lau ZY, Ong WS, Yang GM, Tan KB, Ong MEH, et al. Sociodemographic and clinical factors for non-hospital deaths among cancer patients: A nationwide population-based cohort study. *PLOS ONE*. 2020. p. e0232219. doi:10.1371/journal.pone.0232219
10. Wong TH, Wong YJ, Lau ZY, Nadkarni N, Lim GH, Seow DCC, et al. Not All Falls Are Equal: Risk Factors for Unplanned Readmission in Older Patients After Moderate and Severe Injury—A National Cohort Study. *Journal of the American Medical Directors Association*. 2019. pp. 201–207.e3. doi:10.1016/j.jamda.2018.08.006
11. Open Government, Treasury Board Secretariat of Canada. Open Data 101. [cited 4 Mar 2021]. Available: <https://open.canada.ca/en/open-data-principles>
12. Data.gov.sg. [cited 24 Mar 2021]. Available: <https://data.gov.sg/>
13. HDB Property Information-Data.gov.sg. [cited 6 Mar 2021]. Available: <https://data.gov.sg/dataset/hdb-property-information>

14. Resale Flat Prices-Data.gov.sg. [cited 6 Mar 2021]. Available: <https://data.gov.sg/dataset/resale-flat-prices>
15. OneMap API Docs. [cited 6 Mar 2021]. Available: <https://docs.onemap.sg/>
16. URA API Reference. [cited 4 Mar 2021]. Available: <https://www.ura.gov.sg/maps/api/#introduction>
17. HealthHub. [cited 6 Mar 2021]. Available: <https://www.healthhub.sg/>
18. Home - Singapore Statutes Online. [cited 6 Mar 2021]. Available: <https://sso.agc.gov.sg/>
19. Destitute Persons (Welfare Homes) (Consolidation) Notification - Singapore Statutes Online. [cited 4 Mar 2021]. Available: <https://sso.agc.gov.sg/SL/DPA1989-N1?DocDate=20171026>
20. Nursing Homes – Directory – HealthHub. [cited 4 Mar 2021]. Available: <https://www.healthhub.sg/directory/nursing-homes>
21. National Registration Act - Singapore Statutes Online. [cited 6 Mar 2021]. Available: <https://sso.agc.gov.sg/Act/NRA1965>
22. Tzeng C-F, Lu C-H, Lin C-H. Community Socioeconomic Status and Dispatcher-Assisted Cardiopulmonary Resuscitation for Patients with Out-of-Hospital Cardiac Arrest. *Int J Environ Res Public Health*. 2021;18: 1207.
23. SingPost - Find Postal Code. [cited 22 Mar 2021]. Available: <https://www.singpost.com/find-postal-code>
24. Google Maps. [cited 22 Mar 2021]. Available: <https://www.google.com/maps>
25. Chia JYJ. Six-digit postal code system. National Library Board Singapore; 29 Apr 2010 [cited 23 Mar 2021]. Available: https://eresources.nlb.gov.sg/infopedia/articles/SIP_1006_2010-05-27.html
26. Ho AFW, Wah W, Earnest A, Ng YY, Xie Z, Shahidah N, et al. Health impacts of the Southeast Asian haze problem - A time-stratified case crossover study of the relationship between ambient air pollution and sudden cardiac deaths in Singapore. *Int J Cardiol*. 2018;271: 352–358.
27. Ho, Ho, Zheng, Cheong, En, Pek, et al. The Relationship Between Air Pollution and All-Cause Mortality in Singapore. *Atmosphere*. 2019. p. 9. doi:10.3390/atmos11010009
28. Ho AFW, Zheng H, Earnest A, Cheong KH, Pek PP, Seok JY, et al. Time-Stratified Case Crossover Study of the Association of Outdoor Ambient Air Pollution With the Risk of Acute Myocardial Infarction in the Context of Seasonal Exposure to the Southeast Asian Haze Problem. *Journal of the American Heart Association*. 2019. doi:10.1161/jaha.118.011272
29. Ho AFW, Zheng H, De Silva DA, Wah W, Earnest A, Pang YH, et al. The Relationship Between Ambient Air Pollution and Acute Ischemic Stroke: A Time-Stratified Case-Crossover Study in a City-State With Seasonal Exposure to the Southeast Asian Haze Problem. *Annals of Emergency Medicine*. 2018. pp. 591–601. doi:10.1016/j.annemergmed.2018.06.037
30. FACTSHEET - Data.gov.sg. [cited 4 Mar 2021]. Available: <https://www.tech.gov.sg/files/products-and-services/9-Datagovsg-Factsheet.pdf>
31. Singapore Open Data Licence. [cited 4 Mar 2021]. Available: <https://data.gov.sg/open-data-licence>

Figures

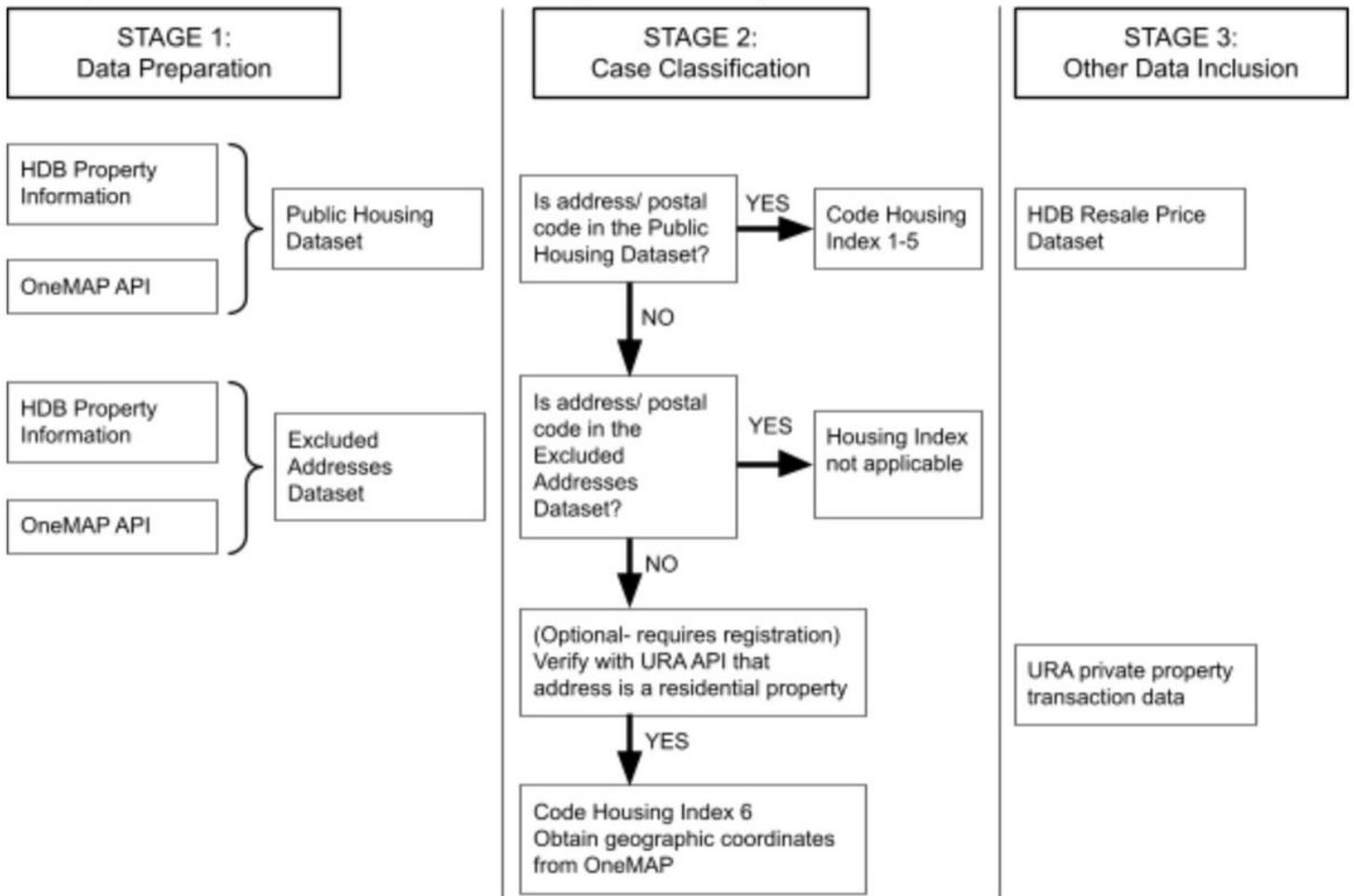


Figure 1

Algorithmic representation of open data workflow, depicting the stages of data preparation from open data sources, case classification process, and potential inclusion of other data



Figure 2

Description of the staged matching of buildings from the Original and Open datasets, with description of validation steps

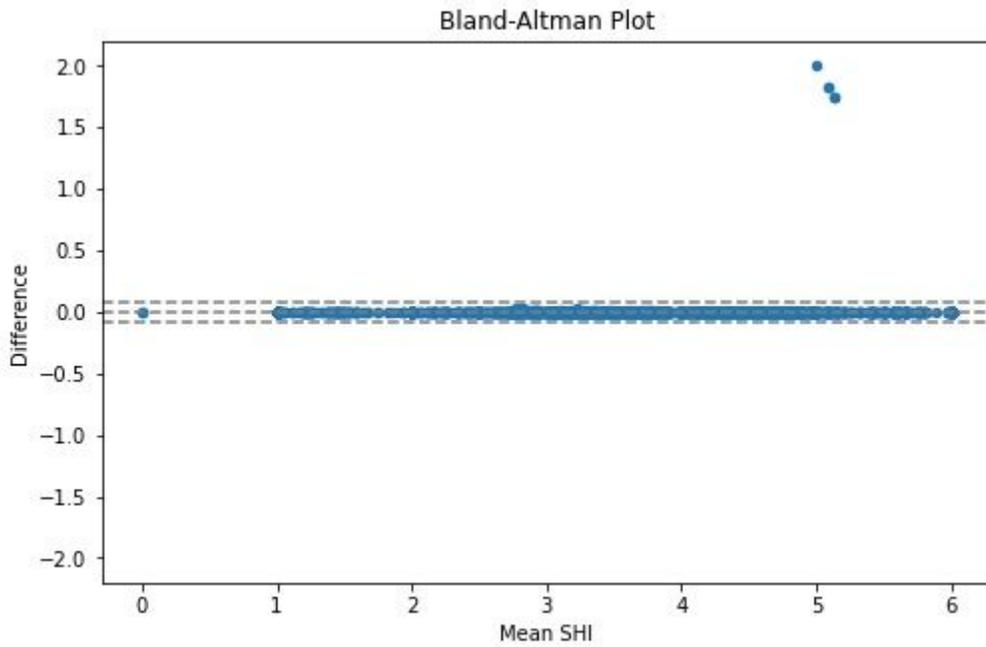


Figure 3

Bland-Altman Plot of SHI Difference against Mean SHI, with 2 Standard Deviation Confidence Limits

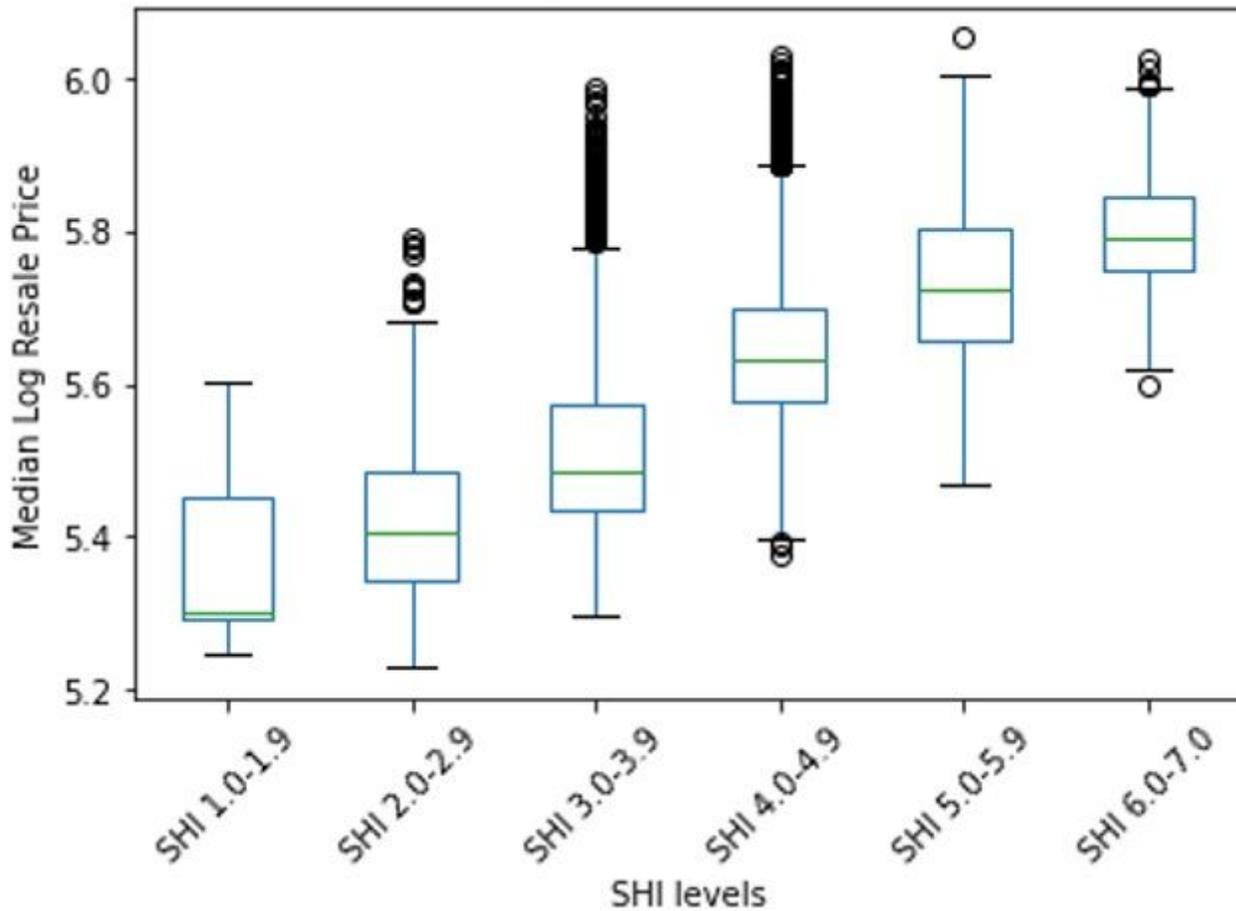


Figure 4

Boxplots of Median Log Resale Price at Different SHI Levels