Housing-Sensitive Health Conditions Can Predict Poor-Quality Housing

Daniel B. Neill, Ph.D.

Associate Professor of Computer Science, Public Service, and Urban Analytics
Director, Machine Learning for Good Laboratory
Robert F. Wagner Graduate School of Public Service
New York University
daniel.neill@nyu.edu

Joint work with:
Ougni Chakraborty
Kacie L. Dragan
Ingrid Gould Ellen
Sherry A. Glied
Renata E. Howland
Scarlett Wang

HealthAffairs

healthaffairs.org
6 May 2024

Acknowledgements and Disclaimers

- This work was supported by the Robert Wood Johnson Foundation (Grant Nos. 30422, 75823, and 80307).
- New York State Medicaid data were obtained from the New York State Department of Health.

- The views and opinions expressed in this material are those of the authors and do not necessarily reflect the official policy or position of the New York State Department of Health or RWJF. Examples of analysis performed in the material are only examples. They should not be utilized in real-world analytic products.

Summary of Research Questions

- Which health conditions are most strongly associated with poor housing conditions?
  - We identified 23 housing-sensitive health conditions (HSHCs) using NYS Medicaid, NYC Landlord Watchlist, and building data.
- How can Medicaid health data be used to target cities' building inspections to poor-quality housing that may impact health?
  - Using machine learning, we built a predictive model, the Housing Health Index (HHI), to predict a building's presence on the Landlord Watchlist using the HSHCs as predictor variables.

Background and Related Work

We use a data-driven analysis to identify specific health conditions associated with housing quality among children and adults.

- A growing literature suggests that housing quality is associated with many dimensions of health.

Pest management | Mold remediation | Lead paint remediation | Structural improvements | Heating system upgrades

improve

Asthma | Blood lead levels | Injuries | Mental health | Hypertension

We use multiple health conditions to build a richer model for predicting poor-quality housing.

- Elevated asthma rates are predictive of failed housing inspections (Samuels et al., 2022)

Background and Related Work

- This work was inspired by the literature on **ambulatory care-sensitive conditions** (Billings et al., 1993).
  - In the same way that ambulatory care-sensitive health conditions have been used to identify areas with inadequate access to primary care, increased prevalence of housing-sensitive health conditions can help local governments identify buildings with inadequate housing quality.

Datasets

- NYS Medicaid data: # of Medicaid recipients (adults vs. children) with each of 65 specific health conditions in each NYC rental building with 5+ units for each study year.
- NYC Landlord Watchlist data: the “worst” buildings & landlords each year, as identified by NYC Office of the Public Advocate.
- Additional building-level data:
  - For modeling: borough, community district, # of units, subsidization status.
  - For evaluation: # of housing violations, # of 311 resident complaints (by agency), and # of emergency repairs.

Methodology (Part 1: Identifying HSHCs)

- For each of **130 health variables** (65 health conditions x {adults, children}), we tested whether a building's presence on the Landlord Watchlist was associated with increased prevalence of that health condition for that subpopulation of Medicaid recipients.
- We used **binomial logistic regression** for this analysis.
  - Eight specifications with different control variables: subsidization status, borough, number of occupants on Medicaid, number of units.
- Health conditions for which the Landlord Watchlist coefficient was positive and at least weakly significant ($p < 0.1$) in at least one specification were chosen as housing-sensitive health conditions.

Housing-Sensitive Health Conditions

A = adults
C = children
* significance levels

Respiratory
- Asthma (A) ***
- Asthma (C) **

Cardiovascular
- Stroke (A) **
- Diabetes w/comp. (A)
- Hypertension (A)
- Hypertension w/comp. (A)

Substance Use
- Substance-related (A) ***
- Alcohol-related (A) **
- Substance-related (C)

Injuries
- Firearm-related (A) **
- Transport-related (C) **
- Other injuries (A) **
- Falls (A)
- Open wounds (A)

Mental Health
- ADHD (C) ***
- Adjustment disorders (C) ***
- Schizophrenia/psychosis (C) ***
- Mood disorders (C) **
- Mood disorders (A) *
- Adjustment disorders (A) *
- Impulse control (C) *
- Impulse control (A)
- Self-injury (C)

Methodology (Part 2: Housing Health Index)

- We used a statistical machine learning approach (**random forests**) to predict each building's probability of being included on the Landlord Watchlist in a given year.

  … as a function of its aggregate count of each HSHC (23 variables) and five additional control variables (borough, subsidization status, # of residential units, # of Medicaid-enrolled adults and children).

- We call this predicted probability its Housing Health Index.

Cross-Validation Methodology

- We use <u>stratified cross-validation</u> for model selection and evaluation:
  - We measure performance on held-out data not used to train the model.
  - We only compare a building to other similarly-sized buildings in the same borough with the same subsidization status.
- We measured the stratified area under the receiver operating characteristic curve (AUC), i.e., the probability that a Landlord Watchlist building has higher HHI score than a similar building that is not on the Watchlist.

Cross-Validation Results

- The HHI model using 23 HSHCs + 5 control variables:
  - Performed the same as using all 130 health variables + 5 control variables.
  - Performed much better than using just the 5 control variables.
  - Performed better using random forests than logistic regression.
- Overall performance varied from year to year, depending on the size and other characteristics of Landlord Watchlist buildings.
  - Many buildings have poor conditions but do not end up on the Watchlist.
  - Most of the useful information is in the top decile of HHI scores.

Evaluation Results: How well can we identify poor-quality housing using the prevalence of HSHCs?

- Compared to a similarly-sized building in its neighborhood, an NYC building with high rates of HSHCs (scoring in the top decile of HHI) is twice as likely to be on the Landlord Watchlist, and has on average:
  - **71%** more 311 calls to NYC Dept. of Housing Preservation and Development
  - **65%** more emergency repairs
  - **52%** more class C (“immediately hazardous”) housing violations
  - **50%** more total housing violations
  - **50%** more 311 calls to NYC Dept. of Buildings
  - **38%** more 311 calls to NYC Dept. of Health and Mental Hygiene

Limitations

- Results are predictive rather than causal. We do not attempt to identify causal effects of poor housing on health.
- Our analysis is limited to the Medicaid population (lower income, typically under age 65) and geographically limited to NYC.
- Various confounders may impact whether a building is placed on the Landlord Watchlist (size, location, reporting bias, politics…)
- Sample size limitations: the HHI can only identify a building for inspection if that building has enough Medicaid-enrolled residents to identify increased rates of HSHCs.

How can cities use these results in practice?

- **Prioritize** (reactive) housing inspections for high-HHI buildings, since these buildings are most likely to have housing issues that may impact health.
- **Inform** housing inspectors of specific risks to residents' health, e.g., air quality for a building with high asthma rates.
- **Discover** unreported housing issues by proactive inspections.
- To mitigate risks, pair code inspections with tenant protections, provision of social services, and community-based advocacy.

Thanks so much for listening!

Any questions?

Please feel free to get in touch at: daniel.neill@nyu.edu.

For more work from our Machine Learning for Good Lab, please see: https://wp.nyu.edu/ml4good.

Health Affairs