Impact of aeroplane noise on mental and physical health: a quasi-experimental analysis

Scarlett Sijia Wang, Sherry Glied, Sharifa Williams, Brian Will, Peter Alexander Muennig

ABSTRACT

Objectives Historically, departures at New York City’s LaGuardia airport flew over a large sports complex within a park. During the US Open tennis games, flights were diverted to fly over a heavily populated foreign-born neighbourhood for roughly 2 weeks out of the year so that the tennis match was not disturbed (the ‘TNNS’ departure). In 2012, the use of the TNNS departure became year-round to better optimise flight patterns around the metropolitan area.

Methods We exploited exogenously induced spatial and temporal variation in flight patterns to examine difference-in-difference effects of this new exposure to aircraft noise on the health of individual residents in the community relative to individuals residing within a demographically similar community that was not impacted. We used individual-level Medicaid records, focusing on conditions associated with noise: sleep disturbance, psychological stress, mental illness, substance use, and cardiovascular disease.

Results We found that increased exposure to aeroplane noise was associated with a significant increase in insomnia across all age groups, but particularly in children ages 5–17 (OR=1.64, 95% CI=1.12 to 2.39). Cardiovascular disease increased significantly both among 18–44-year-old (OR=1.45, 95% CI=1.41 to 1.49) and 45–64-year-old Medicaid recipients (OR=1.15, 95% CI=1.07 to 1.25). Substance use and mental health-related emergency department visits also increased. For ages 5–17, rate ratio (RR) was 4.11 (95% CI=3.28 to 5.16); for ages 18–44, RR was 2.46 (95% CI=2.20 to 2.76); and for ages 45–64, RR was 1.48 (95% CI=1.31 to 1.67).

Conclusion We find that increased exposure to aeroplane noise was associated with an increase in diagnosis of cardiovascular disease, substance use/mental health emergencies and insomnia among local residents.

INTRODUCTION

LaGuardia’s airspace originally utilised departures over areas that were less populated, such as waterways, parks or areas with warehouses or manufacturing. However, as air traffic increased over time, the airspace used for traditional routes of arrivals and departures became crowded and conflicted with that of a nearby airport, John F. Kennedy. As with LaGuardia, other airports sometimes manage increases in traffic by optimising flight patterns with less regard for the populations on the ground. Almost invariably, these new flight patterns require routing aircraft over populated areas that were not previously exposed to aircraft noise.

Noise, and aircraft noise in particular, is associated with a number of health problems, particularly sleep disturbances, mental illness and substance use. The sleep disturbances and psychological stress associated with aircraft noise are, in turn, thought to produce a cascade of biological effects that result in premature ageing via endocrinologic changes.

Noise is thought to produce stress by activating the central nervous system and by interfering with sleep. This stress produces predictable changes in biochemical pathways in human and animal studies that accelerate the rate of ageing. This accelerated ageing process has been linked to the premature onset of age-related diseases, including cardiovascular disease.
While the pathways linking poor sleep and psychological stress to premature ageing and chronic disease are understood, few studies have experimentally examined interventions that alter noise exposure in human populations. Most of our knowledge about the health impact of aircraft noise in humans is based on associational studies. These studies suffer from a number of limitations. On one hand, people who live near airports may self-select, such that those who are less sensitive to noise can take advantage of lower home prices on purchases or rentals for homes. On the other hand, those who live near airports tend to have lower than average income, a major risk factor for premature disease and death. There is limited evidence based of the impact of aircraft noise on premature ageing and health based on experimental or quasiexperimental analysis.

Flight pattern changes afford a unique opportunity for studying the health impact of aircraft noise in humans. In the past decade, flight patterns have shifted, and these shifts have increasingly been accompanied by resident complaints. As they do so, it becomes possible to identify areas that are impacted by new aircraft noise. In general, the point of maximum noise from an aircraft happens immediately after take-off as the aircraft is on full power. This is the for the experimental group in our study.

We conducted a longitudinal case/control study of one well-documented flight pattern change in New York City. LaGuardia Airport (LGA) is one of three major airports in the greater New York City area. One of its departure patterns utilises Flushing Meadows Park, a route known as the ‘Whitestone Climb’. Because it is over green-space, the Whitestone Climb has little impact on humans living in nearby dwellings. However, this park is also the location of the US Open Tennis match. During games, a seldom-used departure route called the ‘Flushing Climb’ (now called TNNIS) was used so that the tennis games were not disturbed by aircraft noise. The TNNIS climb routes aircraft over densely populated Flushing, Queens, greatly increasing the exposure of residents to noise on the ground.

A Freedom of Information Act (FOIA) request by a local group opposed to aeroplane noise in Queens demonstrated that flight patterns using the TNNIS climb have increased to year-round use since 2012. Before that time, the use of the TNNIS climb was rare outside of the US Open.

Previous work found that the year-round use of the TNNIS climb was costly, both in terms of money and lives. However, this economic analysis was primarily based on associational data. Using data on flight patterns over Flushing obtained using the FOIA as well as Medicaid utilisation data, we conducted a quasiexperimental analysis of the health impacts of the aeroplane noise associated with this new route. In the USA, Medicaid is a safety-net health insurance programme for the low-income population. In New York State, over five million low-income individuals enrolled in the Medicaid program in 2012.

METHODS
Data
The data used in this study are New York City Medicaid claims prepared by the New York University Health Evaluation and Analytics Lab. The data include Medicaid member demographic information, address history, eligibility, medical services, and diagnostic information. The database consists of Medicaid fee for service claims and managed care encounters; both are comparable in quality.

A priori specifications and hypotheses
We hypothesised that exposure to aeroplane noise would increase healthcare utilisation, insomnia, mood disorder, anxiety, depression and cardiovascular disease depending on the age group. Specifically, exposure to aeroplane noise would produce sleep disorders across all age groups, would lead to emotional or behavioural disturbances including substance abuse, mood disorder, depression and developmental disorders among young adults aged 18–45 years who tend to be more at risk of these stress-associated disorders, and would produce or exacerbate cardiovascular disease among older adults aged 45 and over when heart disease begins to increase in prevalence. Noise studies suggest wide-ranging psychoneuroendocrinological effects (allostatic load) producing hypertension, hyperglycaemic and hypercholesterolaemia. These biological changes are linked to cardiovascular disease, a correlate of exposure to aeroplane noise as well as other forms of nighttime noise.

Study design
We used individual-level data at the member-cohort level for the analysis. We selected samples of Medicaid members residing in each of the two neighbourhoods at two points in time. The precohort was defined as Medicaid recipients living in the study neighbourhoods between 2009–2011 (precohort) and 2013–2016 (postcohort). About 64% of the Flushing precohort and 63% of the Sunset Park precohort were also in the postcohort. We used the difference-in-difference models to analyse the results.

Exposure
To determine exposure, we used data extracted under a FOIA request for flight patterns over Flushing, Queens, New York and from Part 150 Study Technical Advisory Committee Meeting No. 8 documents. These documents were derived from a 2014 study conducted and funded by the Port Authority of NY & NJ (Port Authority) in partnership with the Federal Aviation Administration (FAA). In these documents, the Port Authority presents estimated noise exposure in geographies surrounding LGA. Exposure is quantified using the Integrated Noise Model in DNL (day-night average sound level) units. We also visually inspected changes in sound related to aircraft flight over sound monitors on the ground in Flushing using Flight Aware, a publicly available flight-tracking website and visited the area. This was done to ensure that the estimates from the Port Authority had face validity.
These geographic regions or corridors were stratified according to intensity of noise exposure over a 24-hour period. We divided census tracts into >60 DNL, 55–60 DNL and <55 DNL (Figure 1). We aggregated census tracts and for regions of Flushing, Queens with noise exposure levels of 55 DNL or greater after 2012. These tracts after 2012 are therefore identified as the treatment condition in this quasi-experimental analysis.

Flushing, Queens is a vibrant, predominantly immigrant neighbourhood. It is increasingly populated by Asians immigrants, particularly those of Chinese descent. The English proficiency in 2018 was 49%, and the population was 54% Asian and 26% white. While the neighbourhood ranked as one of the poorest, the rates of education are higher than average and the rates of crime, obesity and hypertension are much lower than New York City as a whole.

Sunset Park in Brooklyn, New York was identified as an appropriate control neighbourhood as the neighbourhood (1) did not experience a change in exposure to aircraft noise after 2012 when the TNNIS climb began frequent use, and (2) is similar to Flushing with respect to the distribution of sociodemographic and economic characteristics. Like Flushing, Sunset Park is increasingly populated by those of Chinese descent with 32% of the population identifying as Asian and 23% identifying as white. About 48% of the residents were born outside of the USA and the English proficiency in 2018 was 51%. Sunset Park also has high poverty rates with relatively low rates of crime, obesity and hypertension, and high levels of education relative to New York City as a whole. Census tracts in Sunset Park were matched to those identified in Flushing based on race, foreign-born status, and age distribution.

Key outcomes
We used International Classification for Disease revision (ICD-9 and ICD-10) codes as well as Clinical Classifications Software (CCS) codes for diagnostic groups to identify the following conditions of interest: insomnia (ICD-9=3270, 78052 or ICD-10=G470), cardiovascular disease (CCS=109–113), alcohol use disorder (CCS=660), substance use disorder (CCS=661), anxiety (CCS=651), depression (ICD-9=311 or ICD-10=F33), mood disorder (CCS=657) and disorders usually diagnosed in infancy, childhood adolescence (CCS=655), which includes autism, childhood emotional disorder and separation anxiety.

We linked census tracts of Flushing and Sunset Park to geocoded Medicaid addresses. If a recipient had a Medicaid-registered address within a given census tract, they were assigned to that census tract. Participants were excluded if they had invalid addresses, dual Medicare status, did not have a valid date of birth or were not officially enrolled in Medicaid during the study period (Table 1). Participant samples were then defined as Medicaid recipients in the period 2009–2011 (pre-implementation period) and 2013–2015 (the TNNIS use period) and who resided within census tracts in Flushing and Sunset Park.

For these identified records, indicator variables were created to identify type of medical claims inpatient, outpatient and emergency department (ED) visits as well as prescription drugs, both overall and for visits related to substance use and mental health disorders (650–663, 670). We additionally obtained information on the age of the subscriber associated with each record. Because we did not have access to Medicare records, and did not include dual eligible participants due to the high likelihood of
| Table 1  Demographic characteristics and outcome prevalence and rates for Flushing, New York and Sunset Park, New York before (earlier than 2012) and after (after 2013) aeroplane noise increased in Flushing, New York |
|---------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                                | Pre period: 2009–2011 | Post period: 2013–2015 |
|                                | Age 5–17 | Age 18–44 | Age 45–64 | Age 5–17 | Age 18–44 | Age 45–64 | Age 5–17 | Age 18–44 | Age 45–64 |
| Demographics                    | Flushing | Sunset Park | Flushing | Sunset Park | Flushing | Sunset Park | Flushing | Sunset Park | Flushing | Sunset Park |
| Total N                         | 20,120   | 21,597     | 57,089  | 52,016      | 36,472   | 18,681      | 24,552   | 26,009     | 76,278  | 60,774      | 50,806  | 24,421      |
| Age (mean)                      | 11.69    | 11.12      | 31.07   | 29.86       | 53.44    | 53.29       | 11.43    | 10.71      | 30.95   | 30.35       | 53.99   | 54.22       |
| Age (SD)                        | 3.94     | 3.92       | 8.19    | 7.54        | 5.59     | 5.5         | 3.94     | 3.87       | 7.84    | 7.25        | 5.44    | 5.57        |
| Female (%)                      | 48       | 48         | 58      | 57          | 54       | 51          | 48       | 48         | 56      | 54          | 54     | 52          |
| Asian (%)                       | 50       | 46         | 60      | 62          | 63       | 60          | 52       | 47         | 59      | 59          | 63     | 62          |
| Black (%)                       | 6        | 2          | 5       | 1           | 4        | 2           | 4        | 1          | 3       | 1           | 3      | 1           |
| Hispanic (%)                    | 17       | 15         | 11      | 9           | 11       | 11          | 14       | 14         | 7       | 7           | 7      | 8           |
| White (%)                       | 11       | 24         | 10      | 15          | 10       | 14          | 10       | 24         | 8       | 14          | 8      | 12          |
| Other (%)                       | 5        | 3          | 5       | 4           | 6        | 7           | 5        | 3          | 4       | 3           | 5      | 6           |
| Unknown (%)                     | 12       | 11         | 10      | 8           | 6        | 5           | 14       | 12         | 20      | 16          | 13     | 11          |
| Average months on Medicaid per year | 9       | 10         | 8       | 8           | 9       | 10          | 9       | 10         | 8       | 8           | 9      | 9           |
| Total Medicaid spending per person per year | US$1911 | US$1904   | US$3818 | US$3954     | US$6754  | US$6076     | US$1783  | US$1972    | US$3398 | US$3914     | US$6520 | US$6115     |
| Prevalence per 100,000           |          |            |         |             |          |             |          |            |         |             |        |             |
| Insomnia                        | 398      | 477        | 4208    | 6096        | 8036     | 9143        | 623      | 450        | 4755    | 5873        | 11,034  | 10,843       |
| Cardiovascular disease*         | NA*      | NA*        | 1955    | 1576        | 9934     | 9073        | NA*      | NA*        | 3575    | 2040        | 13,260  | 10,786       |
| Alcohol use disorder            | NA*      | NA*        | 2114    | 1173        | 2470     | 2184        | NA*      | NA*        | 2264    | 1358        | 2870    | 2199         |
| Substance use disorder          | NA*      | NA*        | 2265    | 1517        | 1799     | 2098        | NA*      | NA*        | 3799    | 2926        | 4250    | 4058         |
| Anxiety                         | NA*      | NA*        | 5124    | 4639        | 6279     | 6279        | NA*      | NA*        | 5726    | 5265        | 7537    | 7416         |
| Depression                      | NA*      | NA*        | 3782    | 2874        | 6007     | 5867        | NA*      | NA*        | 3191    | 2272        | 5637    | 4656         |
| Mood disorder                   | NA*      | NA*        | 6371    | 4900        | 9399     | 8891        | NA*      | NA*        | 5607    | 4410        | 8375    | 7297         |
| Disorders diagnosed young       | 1983     | 1394       | 289     | 212         | 170      | 112         | 2480     | 2219       | 307     | 244         | 163     | 188          |
| Continued                       |          |            |         |             |          |             |          |            |         |             |        |             |
pre-existing medical conditions and smaller sample size, our sample does not include adults aged 65 or older. Age in years was defined as the calculated age on 1 January 2009 and 1 January 2013, and stratified into three age cohorts, 5–17, 18–44 and 45–64 years.

**Statistical analyses**

Our focus is on the rate of diagnoses for the hypothesised conditions. We first assess whether there were significant changes in utilisation overall between the baseline and TNNIS use periods and whether the observed changes differed by neighbourhood (ie, exposure) after considering other changes over time between these neighbourhoods. We use Poisson regression (see equation 1) to model the number of overall and substance use and mental health-related inpatient, ED and outpatient visits for those months in which participants were enrolled in Medicaid.

For our primary analyses, we use logistic regression (see equation 2) to examine the odds of receiving a diagnosis for the hypothesised conditions. Before implementing these regression analyses, we examined trends in sociodemographic characteristics as well as trends in Medicaid enrolment to ensure that no divergent patterns were noted around 2012. Because racial composition varied somewhat between the two neighbourhoods (table 1), we controlled for race in our analyses to ensure that compositional changes by race did not influence the analysis. We also stratified by age so that we could better test our a priori hypotheses by condition. For chronic diseases, that is, cardiovascular disease, we adopted a longer follow-up period 2008–2011 (preimplementation) and 2013–2016 (TNNIS implementation) to allow for lag time in disease manifestation.

\[
\log(E(Y|X)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_3 \quad (1)
\]

where \( Y \) = number of Medicaid claims for condition of interest and offset = number of Medicaid enrolment month.

\[
\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_3, \quad (2)
\]

where \( p = \Pr(Y=1 \text{ right}) \) is the probability of having Medicaid claim for condition of interest

Here, \( x_1 \) was the indicator for neighbourhood exposure condition (Sunset Park=0 vs Flushing=1); \( x_2 \) indicated implementation period (preimplementation=0 vs TNNIS implementation=1); and \( x_3 \) race/ethnicity group membership (Asian=1, black=2, Hispanic=3, white=4 (reference), other=5, nknown=6).

**Patient and public involvement**

The research question was inspired by the work of a non-profit community organisation called Queens Quiet Skies that works to mitigate aeroplane noise. One of the coauthors of the paper was a member of this organisation and obtained the FOIA requests for FAA documents. These documents were used to identify the treatment
RESULTS

Participants were generally similar across both groups over the two points in time (table 1), but healthcare utilisation varied over time by age group and treatment status.

The increased use of the TNNIS climb occurred in 2012. Prior to that date the climb was only used for the US Open or unexpected weather/runway repairs. We were only able to obtain data on TNNIS departures after the New York Port Authority’s Fiscal Year 2013 because the Port Authority indicated that the earlier data had been lost. There were roughly 1278 TNNIS departures/year on average during US Open events in the 2013–2019 period, providing a point of reference. Since 2013, the total annual number of TNNIS climbs ranged between 9349 and 29 676, with an average of 18 653/year. The DNL figures (figure 1) reflect the average noise exposure by census tract across the 2013–2019 period, and may not reflect the actual change in aircraft noise in Flushing, New York in the pre-2012 and post-2012 periods.

Overall medical utilisation

Table 2 provides results from regression models assessing period-related changes in medical utilisation and diagnoses. The effects of the change in flight patterns on overall utilisation were inconsistent across types of utilisation and age. Overall, outpatient visits increased slightly in Flushing relative to Sunset Park for those aged 18–44 (rate ratio (RR)=1.04, 95% CI=1.04 to 1.05). Prescription drug claims also increased by a similar amount for this group (RR=1.06, 95% CI=1.06 to 1.06).

However, outpatient visits and prescription drug use for children in Flushing aged 5–17 (outpatient RR=0.86, 95% CI=0.85 to 0.87; prescription drug claims RR=0.94, 95% CI=0.92 to 0.93) as well as for older adults 45–64 declined (outpatient RR=0.92, 95% CI=0.92 to 0.93; prescription drug claims RR=0.93, 95% CI=0.92 to 0.93).

While the general pattern for outpatient visits indicates decreased medical utilisation in Flushing compared with Sunset Park over time, ED visits in Flushing increased in the post TNNIS period among all age groups. For ages 5–17, the RR was 1.31 (95% CI=1.24 to 1.37); for ages 18–44, the RR was 1.45 (95% CI=1.41 to 1.49); and for ages 45–64, the RR was 1.16 (95% CI=1.11 to 1.21). Substance use and mental health-related ED visits also increased in Flushing.

| Table 2 | Model results and 95% CI |
|----------------|-------------------------|
| **Rate ratios from the difference-in-difference Poisson model** | | |
| **Age 5–17** | **Age 18–44** | **Age 45–64** |
| Inpatient visits | 0.92 (0.83 to 1.03) | 1.05 (1.02 to 1.08) | 0.93 (0.88 to 0.97) |
| Emergency department visits | 1.31 (1.24 to 1.37) | 1.45 (1.41 to 1.49) | 1.16 (1.11 to 1.21) |
| Outpatient visits | 0.86 (0.85 to 0.87) | 1.04 (1.04 to 1.05) | 0.92 (0.92 to 0.93) |
| Pharmacy claims | 0.94 (0.94 to 0.95) | 1.06 (1.06 to 1.06) | 0.93 (0.92 to 0.93) |
| **Rate ratios from the difference-in-difference Poisson model substance use and mental health related** | | |
| **Age 5–17** | **Age 18–44** | **Age 45–64** |
| Inpatient visits | NA* | 1.11 (1.01 to 1.22) | 1.19 (1.04 to 1.36) |
| Emergency department visits | 4.11 (3.28 to 5.16) | 2.46 (2.20 to 2.76) | 1.48 (1.31 to 1.67) |
| Outpatient visits | 1.12 (1.09 to 1.16) | 0.93 (0.92 to 0.95) | 0.87 (0.85 to 0.89) |

*These diseases and conditions are rare or difficult to diagnose in children.
†We adopted a longer follow-up period 2008–2011(pre-implementation) and 2013–2016 (TNNIS implementation) to allow for lag time in disease manifestation.
in the post period relative to Sunset Park, with rate ratios ranging between 2.5 and 4.1. For ages 5–17, RR was 4.11 (95% CI=3.28 to 5.16); for ages 18–44, RR was 2.46 (95% CI=2.20 to 2.76); and for ages 45–64, RR was 1.48 (95% CI=1.31 to 1.67).

Relative to Sunset Park, inpatient visits in Flushing also show statistically significant increases for overall visits for ages 18–44 (RR=1.05, 95% CI=1.02 to 1.08). However, statistically significant decreases were observed for ages 45–64 (RR=0.93, 95% CI=0.88 to 0.97).

Changes by diagnosis
Relative to Sunset Park, implementation of the TNNIS climb was associated with increases in insomnia diagnoses, particularly for children. For example, the crude prevalence of insomnia for children increased by 57% from 398 per 100,000 in Flushing, compared with a 6% decrease from 477 per 100,000 in Sunset Park. For children in this age group, the OR for insomnia was 1.64 (95% CI=1.12 to 2.39). For older ages, the effect sizes were somewhat less striking (ie, for the 18–44 age group, the OR was 1.17 (95% CI=1.09 to 1.26), and for ages 45–64 the OR=1.1 (95% CI=1.09 to 1.28)).

Cardiovascular disease diagnoses increased significantly in Flushing relative to Sunset Park in the post-2012 period. For 18–44-year-olds, the crude prevalence of cardiovascular disease increased in both neighbourhoods due to ageing of the samples, by 83% from 1955 per 100,000 in Flushing and by 29% from 1576 per 100,000 in Sunset Park. The OR for cardiovascular disease diagnoses in Flushing relative to Sunset Park in this age group was 1.45 (95% CI=1.30 to 1.62). For 45–64-year-olds, the crude prevalence increased by 33% from 9934 per 100,000 in Flushing and 19% from 9073 per 100,000 in Sunset Park. For this age group, the OR was 1.15 (95% CI=1.07 to 1.25). Substance use disorder only increased significantly for the 45–64 age group in Flushing relative to Sunset Park (OR=1.24, 95% CI=1.07 to 1.44).

Figure 2 shows the monthly prevalence of insomnia and cardiovascular disease diagnoses for the 45–64 age group. Age is measured at the beginning of each period, 1 January 2009 for the pre period and 1 January 2013 for the post period. The numerator is the number of unique individuals with one or more diagnosis from inpatient, emergency room or outpatient claims and the denominator is the number of Medicaid-enrolled patients. The trends of both conditions increased throughout the study periods, because people are getting older, but Flushing showed increases that were larger in magnitude in the post period relative to Sunset Park.

Figure 2 Prevalence of insomnia and cardiovascular disease diagnoses per 100,000 among 45–64 age group. *Age is measured at the beginning of each period, 1 January 2009 for the pre period and 1 January 2013 for the post period.
Results for other conditions were more mixed. Clinical depression diagnoses increased for the two older age groups (ages 18–44 OR=1.12, 95% CI=1.02 to 1.24; ages 45–64 OR=1.20, 95% CI=1.08 to 1.33). Broader mood disorder diagnoses, however, only showed statistically significant increases for the 45–65 age group (OR=1.10, 95% CI=1.00 to 1.20). For 5–17-year-olds, developmental disorder diagnoses significantly decreased (OR=0.80, 95% CI=0.66 to 0.97) in Flushing relative to Sunset Park after the implementation of TNNIS.

**DISCUSSION**

We find that increases in aeroplane noise at DNL levels greater than 55 were associated with increases in insomnia, depression, substance abuse and cardiovascular disease across most age groups. These diagnoses are generally consistent with our a priori hypotheses regarding the relationship between exposure to aeroplane noise and health, health-care, and health-related quality of life and cardiovascular disease rates for those who live close to airports is much more comprehensive quasiexperimental and economic analyses are required to determine the extent to which policymakers may wish to act. The costliest options—building airports far from populated areas and providing high-speed transit and freeways—can increase the cost of mitigation by billions of dollars.

**Acknowledgements** The authors thank NYU Health Evaluation and Analytics Lab and the New York State Department of Health for making the Medicaid claims data available and gratefully acknowledge the funding for this research from the Robert Wood Johnson Foundation’s Policies for Action program. Disclaimer: The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the New York State Department of Health. Example of analysis performed within this article are only examples. They should not be utilized in real-world analytic products.

**Contributors** SSW, SG, SW and PAM approved the final draft and agreed to be accountable for all aspects of the work. SSW contributed to study design, data linkage, analysis, interpretation of the data, drafting the methods and results sections. SG contributed to the acquisition of data, study design, analysis and interpretation of data. SZW contributed to data analysis and interpretation of data. PAM contributed to study conception, study design, analysis, interpretation of data and drafting the introduction and discussion sections. Though BW passed away prior to the submission of the manuscript, he had significant contributions in the study conception, the acquisition of data and sample identification. PAM is the guarantor of the study.

**Funding** Robert Wood Johnson Foundation (75822).
