


BMJ Open Improving the geographical precision of rural chronic disease surveillance by using emergency claims data: a cross-sectional comparison of survey versus claims data in Sullivan County, New York

David C Lee,^{1,2} Justin M Feldman,² Marcela Osorio ,¹ Christian A Koziatek,¹ Michael V Nguyen,¹ Ashwini Nagappan,¹ Christopher J Shim,³ Andrew J Vinson,¹ Lorna E Thorpe,² Nancy A McGraw⁴

To cite: Lee DC, Feldman JM, Osorio M, *et al*. Improving the geographical precision of rural chronic disease surveillance by using emergency claims data: a cross-sectional comparison of survey versus claims data in Sullivan County, New York. *BMJ Open* 2019;**9**:e033373. doi:10.1136/bmjopen-2019-033373

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-033373>).

Received 02 August 2019
Revised 31 October 2019
Accepted 31 October 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr David C Lee;
David.Lee@nyulangone.org

ABSTRACT

Objectives Some of the most pressing health problems are found in rural America. However, the surveillance needed to track and prevent disease in these regions is lacking. Our objective was to perform a comprehensive health survey of a single rural county to assess the validity of using emergency claims data to estimate rural disease prevalence at a sub-county level.

Design We performed a cross-sectional study of chronic disease prevalence estimates using emergency department (ED) claims data versus mailed health surveys designed to capture a substantial proportion of residents in New York's rural Sullivan County.

Setting Sullivan County, a rural county ranked second-to-last for health outcomes in New York State.

Participants Adult residents of Sullivan County aged 25 years and older who responded to the health survey in 2017–2018 or had at least one ED visit in 2011–2015.

Outcome measures We compared age and gender-adjusted prevalence of hypertension, hyperlipidaemia, diabetes, cancer, asthma and chronic obstructive pulmonary disease/emphysema among nine sub-county areas.

Results Our county-wide mailed survey obtained 6675 completed responses for a response rate of 30.4%. This sample represented more than 12% of the estimated 53 020 adults in Sullivan County. Using emergency claims data, we identified 34 576 adults from Sullivan County who visited an ED at least once during 2011–2015. At a sub-county level, prevalence estimates from mailed surveys and emergency claims data correlated especially well for diabetes ($r=0.90$) and asthma ($r=0.85$). Other conditions were not well correlated (range: 0.23–0.46). Using emergency claims data, we created more geographically detailed maps of disease prevalence using geocoded addresses.

Conclusions For select conditions, emergency claims data may be useful for tracking disease prevalence in rural areas and providing more geographically detailed estimates. For rural regions lacking robust health

Strengths and limitations of this study

- Validates the use of emergency claims data to perform geographically detailed surveillance in rural settings.
- Provides a standard for estimating disease prevalence at a local level by performing a county-wide mailed survey.
- Limited by the accuracy of diagnosis codes found in claims data and is more accurate for conditions likely to be captured during emergency visits.
- Has the potential to improve rural health surveillance by using existing data to track the burden of chronic diseases.

surveillance, emergency claims data can inform how to geographically target efforts to prevent chronic disease.

INTRODUCTION

In New York State, Sullivan County has been ranked 61 out of 62 by the County Health Rankings based on the rates of premature death and quality of life (poor overall, physical, mental health and low birth weights) just behind Bronx County in New York City.¹ Located just 2 hours northwest of New York City, Sullivan County is rural and more than 70% of its residents are white. Like many rural areas of America, Sullivan County has faced significant economic challenges, along with disparities in healthcare access.^{2 3} Although some of the most pressing health problems can be found in rural America, their public health institutions lack timely data needed to provide geographically detailed chronic disease surveillance.^{4 5} Nationwide health



surveys, such as the Behavioral Risk Factor Surveillance System (BRFSS), often have inadequate coverage of these rural regions, and efforts to use models to extrapolate estimates of disease prevalence have questionable validity.^{6,7}

In recent years, there has been increasing interest in using alternative sources of data to track chronic disease prevalence.^{8–10} Approaches using claims data and electronic health records have emerged among the potential options.¹¹ These data are collected routinely by state agencies and may provide a cost-effective, ready-to-analyse alternative to expensive and time-intensive traditional survey methods.¹² For instance, one in five Americans report having visited an emergency department (ED) in the past year, which provides a 20% population sample with a single year of data.^{13,14} However, these approaches need to be validated before widespread dissemination because, unlike surveys, they are not random population samples and may therefore not be representative.

There are several challenges that make estimation of chronic disease prevalence in rural areas difficult. The Centers for Disease Control and Prevention (CDC) has started to use modelling approaches with Bayesian and spatial smoothing with BRFSS data to estimate county-level disease prevalence in rural areas.¹⁵ But, few traditional health surveys have been performed in these areas with sample sizes adequate for sub-county level area estimation.^{16,17} In addition, there is similarly sparse data on the sociodemographic composition of these rural regions (as evidenced by wide confidence intervals for sub-county estimates of race and ethnicity). Furthermore, ZIP codes, county borders and other geographical units are less likely to align in rural areas, limiting the possibility of attributing aggregated data to specific regions.¹⁸ In addition, certain traditional survey techniques used to refine estimates based on underlying demographic characteristics (eg, statistical weighting, adjustment or stratification) that are often performed with Census data cannot be readily applied in rural areas especially if there is insufficient data.¹⁹

The goal of this study was to perform a comprehensive health survey of a single rural county in the USA. We report the results of a geographically distributed health survey delivered by mail to households within Sullivan County, New York. We then compare the disease prevalence estimates obtained from these surveys with a novel method that uses emergency claims data to identify areas with a higher burden of chronic disease.¹⁴

METHODS

Study design

We administered a brief health survey by mail throughout Sullivan County during Fall 2017 and Spring 2018 to a random sample of residential addresses. We used survey data to estimate the age and gender-standardised prevalence of several chronic diseases at a sub-county level. We also estimated disease prevalence using the comprehensive, all-payer New York Statewide Planning and Research

Cooperative System (SPARCS) claims database. Our alternative measure was the proportion of ED patients with ≥ 1 diagnostic code for a given disease on ≥ 1 emergency visit during the period 2011–2015.¹⁴ In each method, residents with an address located at a nursing or correctional facility were excluded to estimate prevalence for the non-institutionalised population. This study was approved by NYU School of Medicine's Institutional Review Board.

Mailed health surveys

To generate a sampling frame for our mailed health survey, we obtained point and parcel data for all mailing addresses in Sullivan County from the New York State GIS Clearinghouse (www.gis.ny.gov).²⁰ This data source was selected because it contained property class and land use data. Addresses were filtered to include any residential listing not marked as seasonal housing. We also included commercial addresses listed as apartments. This list of mailing addresses was then refined using an address verification service (www.smartystreets.com) to select valid, non-vacant mailable addresses.²¹ As a substantial proportion of residents do not receive delivered mail in Sullivan County, we also queried the address verification service to find all valid, mailable PO boxes in the county. The final sampling frame consisted of 39084 households located across 56 ZIP codes within Sullivan County.

Given the sparse population in some areas of the rural county, less-populated ZIP codes were oversampled to maximise the geographical coverage of the survey over the entire county. To do so, we used a quota sampling strategy. We mailed surveys to a random sample of 750 households for each ZIP code in our sampling frame. In ZIP codes with fewer than 750 households, all households were mailed a survey. Each health survey consisted of questions that first confirmed residence within Sullivan County and age over 18 years, and then asked a brief selection of health and demographic questions derived from the BRFSS (see online supplementary file 1 for mailed health survey).²² For households with multiple residents, we asked that only one adult respond to the survey. We mailed a survey to 24 141 or 62% of the households in our sampling frame. Survey respondents were offered a \$10 gift card for participation and a stamped return envelope was enclosed in the surveys. The Sullivan County Public Health Department also made local news outlets aware of the survey and fielded phone calls from local residents to confirm that the survey was legitimate.

Emergency claims data

Using the SPARCS, all-payer claims database, we identified all adult patients who had visited an ED located at a general acute care hospital in New York State between 2011 and 2015. We included all patients with a PO box or home address located within the borders of Sullivan County. Patients with more than one ED visit either at the same hospital or different hospitals were counted as a single observation by collapsing multiple visits using unique identifiers from SPARCS. The result was a listing

of unique Sullivan County residents who had accessed emergency care at least once during the 5-year period.

Study outcomes

Our primary outcome was the prevalence of chronic disease at a sub-county level as identified by our mailed health survey or estimated using emergency claims data. In our mailed survey, respondents were asked if they had ever been diagnosed with hypertension, hyperlipidaemia, cancer, diabetes, asthma, chronic obstructive pulmonary disease (COPD) or emphysema. In our analysis of emergency claims data, all available primary and secondary diagnosis codes across visits were scanned by individual for the presence of ≥ 1 diagnosis code during ≥ 1 ED visit for these same conditions. The codes from the International Classification of Diseases (ICD-9 and ICD-10) used were: hypertension (401–405 or I10–I16), hyperlipidaemia (272 or E78), diabetes (250 or E10–E11), cancer (140–239 or C00–C96), asthma (493 or J45), and COPD/emphysema (491–492 or J43–J44). Thus, prevalence was estimated as a proportion: the number of unique ED patients with each of the listed conditions divided by the total number of unique ED patients.

Statistical analysis

To generate the sub-county areas in our analysis, we first grouped ZIP codes based on the Census-defined subdivisions (ie, town borders) within Sullivan County. ZIP codes were assigned to these subdivisions based on the largest area of overlap given that ZIP code boundaries do not exactly align with town borders.¹⁸ After grouping ZIP codes into these 15 subdivisions, it was found that 10 of these subdivisions had less than 2000 households who received a mailed survey and were thus unlikely to obtain the minimum 500 survey responses, a benchmark set by the CDC for obtaining acceptably narrow CIs for prevalence estimation (online supplementary figure 1).¹⁹ Therefore, these less populated subdivisions were systematically merged with each other based on proximity and population density to form four sub-county areas with a sufficient number of sampled households. The result was nine sub-county areas made of five subdivisions with adequate sampling and four areas combining neighbouring subdivisions to attain adequate sampling (see online supplementary table 1 for more details of aggregating ZIP codes into subdivisions and then sub-county areas).

In aggregating prevalence estimates between ZIP codes to create the sub-county areas, we used two weighting approaches. For the mailed survey, we applied design weights (the inverse probability of selection from the sampling frame) to account for our oversampling of less-populated ZIP codes. For the emergency claims data, we weighted ZIP code prevalence estimates by the inverse of the total number of unique ED patients divided by the Census estimate of adults aged 25 years and older for each ZIP code in Sullivan County to account for known differences in ED usage based on proximity to the nearest

hospital.²³ Prevalence estimates using both methods were then standardised to the overall age and gender distributions in Sullivan County from the 5-year 2012–2016 American Community Survey (ACS).²⁴ We then calculated Pearson correlation coefficients comparing the prevalence estimates obtained using the two methods at the sub-county level. By convention, the strength of correlation was graded as very strong (0.80–1.00), strong (0.60–0.79), moderate (0.40–0.59), weak (0.20–0.39) and very weak (0.00–0.19).

Geographical analysis

We also performed geographically detailed surveillance using the larger sample of Sullivan County residents identified in emergency claims data. For the subset of patients with a geocodable home address, we calculated unadjusted disease prevalence among their 100 nearest neighbours identified in the population of unique ED patients. We then interpolated rasters from this point data using the inverse squared distance technique. Chronic disease prevalence maps were generated from these unadjusted prevalence estimations for diabetes, asthma and hypertension, with categories based on SDs from the mean. For comparison, these maps were also created based on the 200 nearest neighbours to assess the influence of changing this parameter.

Statistical analyses were performed using Stata V.14.2 (Statacorp, 2015). Geographical analysis and mapping were performed using ArcGIS Desktop V.10.5.1 (ESRI; Redlands, California, USA, 2017).

RESULTS

Mailed survey responses

Of the 24 141 surveys that we mailed to addresses within Sullivan County, approximately 20% were returned to sender even after using an address verification service (online supplementary figure 2). Of the 7241 survey responses received, 216 were missing key demographic information or were otherwise incomplete, 248 were not residents of Sullivan County and 22 were located at a nursing or correctional facility. In addition, only 80 respondents were aged 18–24 years old, which we deemed too few for inclusion in the study. Therefore, we limited study results to adults aged 25 years and older. Using the AAPOR RR2 definition for mail surveys of unnamed persons, our response rate was 30.4%.²⁵

Population characteristics

The county-wide mailed survey received valid responses for 6675 adults or 12.6% of the adult population 25 years and older in Sullivan County. Using 5 years of emergency claims data, we were able to identify 65.2% of the Census-estimated adult population 25 years and older in Sullivan County. In comparison with ACS 2012–2016 Census estimates, survey respondents were notably older (42.5% vs Census estimate of 23.9% aged 65 years and older). In comparison, the population of unique ED patients was

**Table 1** Demographic comparisons among Census estimates and data sources

Demographic comparisons	2012–2016 Census estimates	County-wide mailed survey	Emergency claims data
Total	53 020	6675	34 567
Age (years)			
25–44	33.0%	15.9% (15.0%–16.8%)	39.2% (38.7%–39.7%)
45–64	43.1%	41.6% (40.4%–42.7%)	36.9% (36.4%–37.5%)
65 and older	23.9%	42.5% (41.3%–43.7%)	23.9% (23.4%–24.3%)
Sex			
Male	50.8%	39.3% (38.1%–40.4%)	49.5% (48.9%–50.0%)
Female	49.2%	60.7% (59.6%–61.9%)	50.5% (50.0%–51.1%)
Race/ethnicity			
White	73.0%	88.7% (87.4%–89.0%)	75.4% (74.9%–75.8%)
Black	7.7%	2.3% (1.9%–2.6%)	7.0% (6.7%–7.3%)
Hispanic	15.0%	5.3% (4.8%–5.9%)	10.5% (10.1%–10.8%)
Asian	1.6%	0.9% (0.7%–1.2%)	0.5% (0.4%–0.6%)
Other	2.7%	2.8% (2.4%–3.2%)	6.6% (6.3%–6.9%)

slightly younger (39.2% vs Census estimate of 33.0% aged 25–44 years old). A higher proportion of survey respondents were women (60.7% vs Census estimate of 49.2%). Also, a higher proportion of survey respondents were non-Hispanic white (88.7% vs Census estimate of 73.0%). However, the sex and race/ethnicity distributions of the unique ED patient population were similar to Census estimates (table 1).

Prevalence estimates adjusted for age and gender

The county-wide prevalence estimates using emergency claims data was higher than the mailed survey for diabetes, but lower for asthma (table 2). The correlation by sub-county area was very strong for these two conditions at $r=0.90$ (95% CI: 0.60 to 0.98) and $r=0.85$ (95% CI: 0.44 to 0.97), respectively. For all other conditions except for diabetes, the county-wide prevalence estimates using emergency claims data was lower than the mailed survey. These correlations were graded across conditions: moderate for hypertension ($r=0.46$, CI: -0.30 to 0.86) and COPD/emphysema ($r=0.42$, CI: -0.34 to 0.85), and weak for cancer ($r=0.39$, CI: -0.37 to 0.84) and hyperlipidaemia

($r=0.23$, CI: -0.51 to 0.78). Graphs of these correlations are found in online supplementary figure 3, which demonstrate the variability between prevalence estimates especially for conditions with poor correlation by sub-county area. We displayed maps of prevalence estimates for diabetes, asthma and hypertension based on survey results for the sub-county areas analysed in figure 1.

ED surveillance

Among the 34 567 unique patients identified from emergency claims data, 76% had a geocodable home address, 20% were PO box only and 4% were not geocodable but had a ZIP code located fully within Sullivan County. Using the 100 nearest neighbours among patients with a geocodable home address, we estimated unadjusted prevalence at the geocoded location of each patient and created interpolated rasters to provide a more geographically detailed maps of diabetes, asthma and hypertension prevalence (figure 2). These maps were able to identify localised clusters of disease throughout the county with greater geographical detail.

Table 2 Age and gender-adjusted county-level disease prevalence and correlation at a sub-county level

Chronic disease	County-wide mailed survey	Emergency claims data	Correlation coefficient
Diabetes	12.7% (11.5%–13.8%)	14.7% (14.3%–15.1%)	0.90 (0.60 to 0.98)
Asthma	15.6% (14.1%–17.1%)	8.7% (8.4%–9.0%)	0.85 (0.44 to 0.97)
Hypertension	38.3% (36.7%–39.8%)	36.1% (35.6%–36.6%)	0.46 (-0.30 to 0.86)
COPD/emphysema	7.0% (6.2%–7.9%)	4.0% (3.8%–4.2%)	0.42 (-0.34 to 0.85)
Cancer	10.7% (9.8%–11.5%)	3.9% (3.7%–4.2%)	0.39 (-0.37 to 0.84)
Hyperlipidaemia	33.9% (32.2%–35.5%)	21.2% (20.8%–21.7%)	0.23 (-0.51 to 0.78)

COPD, chronic obstructive pulmonary disease.

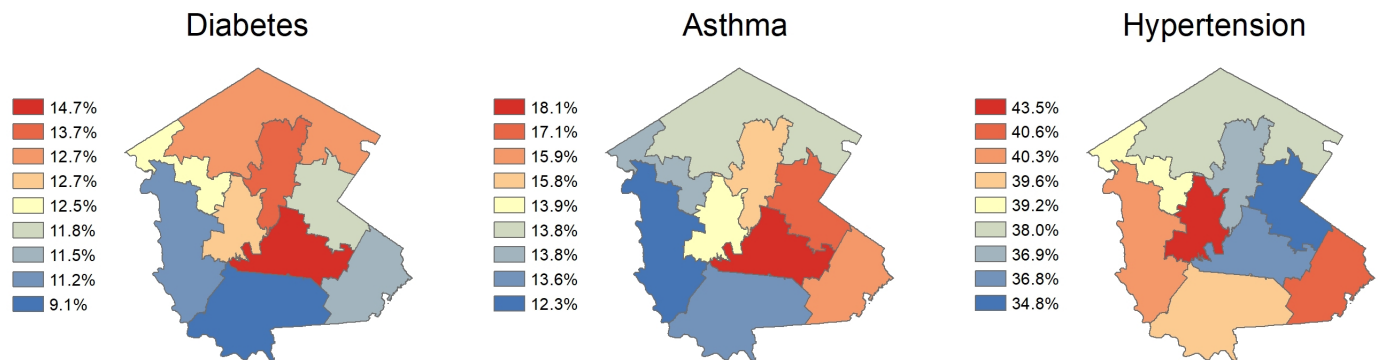


Figure 1 Sub-county estimates of adjusted disease prevalence based on mailed survey responses.

DISCUSSION

The intensity of health problems experienced by residents living in rural areas of the country underscores the need for improving our methods of health surveillance.²³ Our study findings demonstrate a novel solution that uses emergency claims data to estimate chronic disease prevalence at a sub-county level. These estimates are important for identifying key hotspots of disease, which may reveal previously unexplored risk factors that increase disease burden in rural America and guide efforts to prevent chronic disease in specific geographical areas that experience the worst health outcomes.²⁶ Current health surveillance techniques rely on traditional methods such as telephone-based surveys. Not only are these methods costly and time-intensive, but also due to dramatic shifts in phone use response rates over the past two decades have dropped dramatically from around 36% to 9%.^{27,28} The sample size of a large national health survey such as the BRFSS is inadequate for generating precise estimates of disease prevalence even at the county level for much of rural America, which is why the CDC has started to use alternative estimation methods to impute prevalence among rural counties.²⁹

Recent efforts to provide greater geographical coverage have focused on approaches that use the data in adequately sampled areas and statistical models to extrapolate disease estimates for poorly sampled areas largely based on sociodemographic factors.¹⁶ But many of these techniques have not been validated, and in the few instances when they have been compared, these

approaches do not always work as well as expected.^{6,7} Our mailed health survey found that adjusted diabetes prevalence in Sullivan County was 12.7%. This estimate is much higher than the CDC's most recent estimate of 9.5% in 2015, which is based on a modelling approach. For a given area, these modelling approaches can be especially imprecise when used to estimate disease prevalence in areas with low response rates, which includes many rural regions.

Other efforts to advance health surveillance methods have experimented with the use of claims data and electronic health records to provide estimates of disease prevalence. A recent study demonstrated that emergency claims data could be used to estimate chronic disease prevalence in New York City, and this approach was validated with results obtained from an annually performed citywide health survey. In this urban study, it was found that conditions including diabetes, hypertension and asthma had correlations of 0.86, 0.88 and 0.77, respectively, when analysed among 34 sub-county areas.¹⁴

With our novel method of using emergency claims data to estimate chronic disease burden, we identified health records for a substantial majority of all adults in Sullivan County using 5 years of emergency claims data. Furthermore, the demographic patterns among this population of unique ED patients were much closer to Census estimates than our county-wide mailed survey. Underrepresentation of certain demographic groups, especially minorities, is a common problem of traditional survey methods that can be adjusted for, as long as geographically

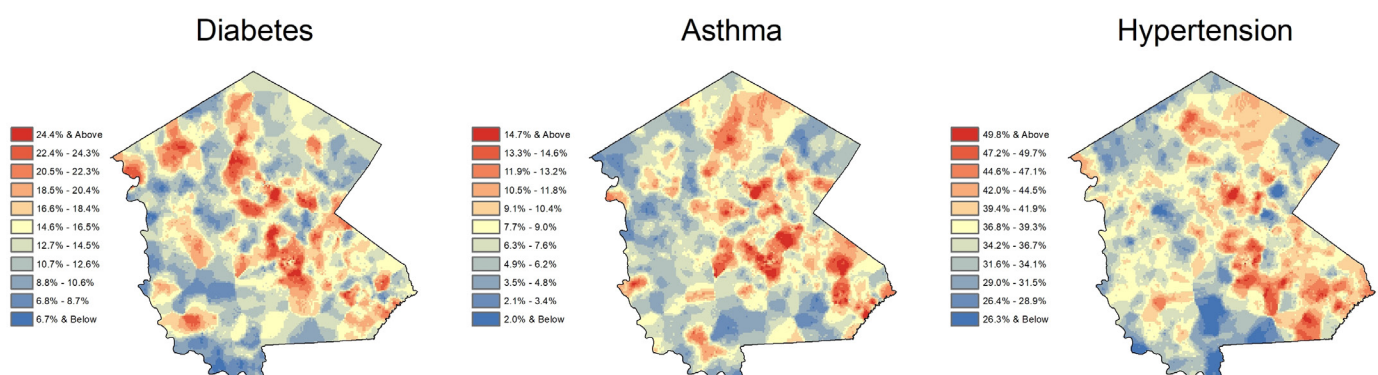


Figure 2 Geographically detailed estimates of unadjusted disease prevalence based on emergency claims data.



matched sociodemographic data exists.³⁰ In rural areas where Census estimates for race and ethnicity often have wide CIs, emergency claims data may provide an alternative population sample that closely mirrors the underlying population in a given region.

For some conditions such as diabetes and asthma, we found strong correlation between the two estimation methods for sub-county disease prevalence. For the other conditions studied, the strength of correlation was weaker. This may be attributable to disease-specific differences in the validity of both ED claims data and self-reported survey data. Prior research has shown that, for both data sources, validity is routinely higher for diabetes and asthma but lower for other conditions such as hyperlipidaemia, with low sensitivity (ie, under-reporting) being the reason for poor correlation.^{8–10} It should be noted that although some conditions such as COPD are a frequent primary diagnosis for a patient's ED visit, COPD may not be frequently accounted for as one of the secondary diagnoses, which are included in this ED-based surveillance approach.

Emergency claims data are already widely collected around the country, can capture a large population sample and in some areas include address data that can be used to precisely identify where patients live. By geocoding these addresses, more precise health surveillance can provide detailed maps of disease burden. This granular level of geographical detail is important because localised hotspots of disease might otherwise be hidden as they are averaged out by neighbouring areas of low disease prevalence. However, some important caveats should be understood before employing these methods. There is some variation in how accurately some hospitals capture chronic disease conditions using diagnosis codes.³¹ In addition, for some parts of rural America, mail is only delivered to PO boxes, therefore the more geographically detailed maps of disease prevalence based on geocoded data may not be accurate in these regions where mail is not delivered directly. Furthermore, our study found substantial variability in prevalence estimates for conditions that may not be well captured by emergency claims data. More research may be needed to determine the best approaches for estimating disease prevalence in rural areas.

Limitations

Since surveys did not ask respondents to report household size, single-adult households are likely over-represented. Furthermore, we did not specify a method of randomly selecting an adult in households with multiple residents, which may have contributed to bias in our sample. However, we estimated age and gender-standardised rates to the overall population in Sullivan County, which may partially reduce this bias. While our adjustment methods reduced age and gender-specific non-response bias, we were unable to standardise by race and ethnicity due to very small proportion of minorities in several areas of Sullivan County. Given that minorities often have higher

rates of chronic disease and tend to have lower response rates, our mailed survey may have underestimated disease prevalence. Although the groups that frequently seek emergency care and those who respond to surveys tend to diverge, there may have been some sort of parallel bias that accounted for the correlations or disease prevalence identified in our study. Our method that used emergency claims data to estimate disease prevalence is subject to many of the limitations associated with the use of administrative data. Fidelity of coding some variables including race, ethnicity and diagnosis codes can vary by hospital and may impact resulting disease prevalence estimates. Also, these claims data are often available about a year after they have been filed, thus there is some lag in reporting. In this study, emergency claims data were collected for 2011–2015, whereas the county-wide survey was performed in 2017–2018.

CONCLUSION

We found that for select conditions, emergency claims data may be useful for tracking disease prevalence in rural areas and may provide more geographically precise estimates. Given the infrastructure already in place to collect this data, efforts could be focused on collecting more accurate diagnosis codes and more detailed geographical data. This approach could potentially help match the limited public health resources of a rural county to the geographical areas with the highest burden of disease.

Author affiliations

¹Ronald O. Perelman Department of Emergency Medicine, New York University School of Medicine, New York City, New York, USA

²Department of Population Health, New York University School of Medicine, New York City, New York, USA

³California Northstate University College of Medicine, Elk Grove, California, USA

⁴Sullivan County Public Health Services, Liberty, New York, USA

Acknowledgements The content of the study reflects the views of the authors and not the official position of the Sullivan County Public Health Department nor the NYU School of Medicine.

Contributors Study conception and design: DCL, LET and NAM; acquisition of the data: DCL, MO, MVN, AN and AJV; analysis and interpretation of the data: DCL, JMF, MO, CAK, CJS, AJV and LET; drafting of the manuscript: DCL; critical revision of the manuscript for intellectual content: JMF, MO, CAK, MVN, AN, CJS, AJV and LET. DCL is the guarantor of this work and had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding This work was supported by the New York State Health Foundation grant number 16-04083.

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by NYU School of Medicine's Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Marcela Osorio <http://orcid.org/0000-0001-7315-8341>

REFERENCES

- County Health Rankings and Roadmaps. *2018 County health rankings report*. Madison, WI: University of Wisconsin Population Health Institute, 2018. http://www.countyhealthrankings.org/sites/default/files/state/downloads/CHR2018_NY.pdf
- Harris JK, Beatty K, Leider JP, et al. The double disparity facing rural local health departments. *Annu Rev Public Health* 2016;37:167–84.
- Brundisini F, Giacomini M, DeJean D, et al. Chronic disease patients' experiences with accessing health care in rural and remote areas: a systematic review and qualitative meta-synthesis. *Ont Health Technol Assess Ser* 2013;13:1–33.
- Puma JE, Belansky ES, Garcia R, et al. A community-engaged approach to collecting rural health surveillance data. *J Rural Health* 2017;33:257–65.
- Bellamy GR, Bolin JN, Gamm LD. Rural healthy people 2010, 2020, and beyond: the need goes on. *Fam Community Health* 2011;34:182–8.
- Feldman JM. Re: obesity and place: chronic disease in the 500 largest U.S. cities. *Obes Res Clin Pract* 2018;12.
- Wang Y, Holt JB, Zhang X, et al. Comparison of methods for estimating prevalence of chronic diseases and health behaviors for small geographic areas: Boston Validation Study, 2013. *Prev Chronic Dis* 2017;14:E99.
- Hebert PL, Geiss LS, Tierney EF, et al. Identifying persons with diabetes using Medicare claims data. *American Journal of Medical Quality* 1999;14:270–7.
- Tessier-Sherman B, Galusha D, Taiwo OA, et al. Further validation that claims data are a useful tool for epidemiologic research on hypertension. *BMC Public Health* 2013;13:51.
- Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. comparing administrative data and self-reports. *Med Care* 1997;35:932–47.
- Colantonio LD, Levitan EB, Yun H, et al. Use of Medicare claims data for the identification of myocardial infarction: the reasons for geographic and racial differences in stroke (REGARDS) study. *Med Care* 2018;56:1051–1059.
- Hope KG, Merritt TD, Durrheim DN, et al. Evaluating the utility of emergency department syndromic surveillance for a regional public health service. *Commun Dis Intell Q Rep* 2010;34:310–8.
- Centers for Disease Control and Prevention. *Health, United States, 2013: with special feature on prescription drugs*. Hyattsville, MD: National Center for Health Statistics, 2014.
- Lee DC, Long JA, Wall SP, et al. Determining chronic disease prevalence in local populations using emergency department surveillance. *Am J Public Health* 2015;105:e67–74.
- Centers for Disease Control and Prevention. Public Health Surveillance in Rural Communities, 2019. Available: <https://www.cdc.gov/chronicdisease/resources/publications/factsheets/surveillance-in-rural-communities.htm>
- Jia H, Muennig P, Borawski E. Comparison of small-area analysis techniques for estimating county-level outcomes. *Am J Prev Med* 2004;26:453–60.
- Zhang X, Holt JB, Yun S, et al. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. *Am J Epidemiol* 2015;182:127–37.
- Coburn AF, MacKinney AC, McBride TD, et al. *Choosing rural definitions: implications for health policy*. Iowa City, IA: Rural Policy Research Institute Health Panel, 2007.
- Centers for Disease Control and Prevention. SMART BRFSS MMSA Methodology, 2016. Available: https://www.cdc.gov/brfss/smart/2016/2016_SMART_BRFSS_MMSA_Methodology.pdf [Accessed 8 Nov 2018].
- NYS ITS GIS Program Office. NYS GIS Clearinghouse. Available: www.gis.ny.gov [Accessed 8 Nov 2018].
- Mullen PD, Savas LS, Bundy Łucja T, et al. Minimal intervention delivered by 2-1-1 information and referral specialists promotes smoke-free homes among 2-1-1 callers: a Texas generalisation trial. *Tob Control* 2016;25:i10–18.
- Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med Res Methodol* 2013;13:49.
- Lee DC, Doran KM, Polsky D, et al. Geographic variation in the demand for emergency care: a local population-level analysis. *Health Care* 2016;4:98–103.
- Klein RJ, Schoenborn CA. *Age adjustment using the 2000 projected U.S. population. healthy people statistical notes*. Hyattsville, Maryland: National Center for Health Statistics, 2001.
- American Association for Public Opinion Research. Mailed surveys of unnamed persons, 2016. Available: [https://www.aapor.org/getattachment/Standards-Ethics/Standard-Definitions-\(1\)/Mail_Surveys_Unnamed_Persons_042216.pdf.aspx](https://www.aapor.org/getattachment/Standards-Ethics/Standard-Definitions-(1)/Mail_Surveys_Unnamed_Persons_042216.pdf.aspx) [Accessed 8 Nov 2018].
- Brownson RC, Fielding JE, Green LW. Building capacity for evidence-based public health: reconciling the Pulls of practice and the push of research. *Annu Rev Public Health* 2018;39:27–53.
- Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007;17:643–53.
- Keeter S, Hatley N, Kennedy C, et al. *What low response rates mean for telephone surveys*. Washington DC: Pew Research Center, 2017. <http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/>
- O'Connor A, Wellenius G. Rural–urban disparities in the prevalence of diabetes and coronary heart disease. *Public Health* 2012;126:813–20.
- Link MW, Mokdad AH, Stackhouse HF, et al. Race, ethnicity, and linguistic isolation as determinants of participation in public health surveillance surveys. *Prev Chronic Dis* 2006;3:A09.
- Lee DC, Swartz JL, Koziatek CA, et al. Assessing the reliability of performing citywide chronic disease surveillance using emergency department data from sentinel hospitals. *Popul Health Manag* 2017;20:427–34.