Published in final edited form as:

Health Place. 2015 September ; 35: 95–104. doi:10.1016/j.healthplace.2015.08.003.

Comparison of Estimation Methods for Creating Small Area Rates of Acute Myocardial Infarction Among Medicare Beneficiaries in California

Laura C. Yasaitis^{a,1}, Mariana C. Arcaya^b, and S.V. Subramanian^b

^aHarvard Center for Population and Development Studies, Harvard University, 9 Bow St, Cambridge, MA 02138

^bDepartment of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115

Abstract

Creating local population health measures from administrative data would be useful for health policy and public health monitoring purposes. While a wide range of options – from simple spatial smoothers to model-based methods – for estimating such rates exists, there are relatively few side-by-side comparisons, especially not with real-world data. In this paper, we compare methods for creating local estimates of acute myocardial infarction rates from Medicare claims data. A Bayesian Monte Carlo Markov Chain estimator that incorporated spatial and local random effects performed best, followed by a method-of-moments spatial Empirical Bayes estimator. As the former is more complicated and time-consuming, spatial linear Empirical Bayes methods may represent a good alternative for non-specialist investigators.

Keywords

Medicare; Myocardial Infarction; Spatial Analysis; Local Disease Rates; Markov Chains

Introduction

With increasing availability of geocoded health data (Krieger et al., 2002; Vaughan et al., 2014; CDC 2013; and Losina et al., 2007) and more detailed and up-to-date local demographic data (Wong and Sun, 2013), geographic analyses at the local level are becoming increasingly popular in public health and health policy research. In the clinic, local population health measures may help inform clinicians' understanding of the social

Correspondence to: Laura C. Yasaitis.

¹Current address: Health Policy Division, Perelman School of Medicine, 1127 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, yasaitis@mail.med.upenn.edu

Conflicts of Interest: None disclosed.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

determinants of health; individuals may face increased risk of adverse health events due to factors beyond their easily discernible demographics. For public health surveillance, up-to-date, fine-grained disease rates may help policy makers and researchers track the effects of public health initiatives. Yet, population health measures for small areas are limited by potential disclosure concerns, as well as by the fact that rates derived from raw data may not act as stable estimates of true disease prevalence (Sakshaug and Raghunathan, 2014 and Anselin et al., 2006).

The statistical and spatial analysis methods for addressing this small area problem have rarely interested mainstream public health and policy researchers, but user-friendly software packages (Anselin et al., 2005; Rey and Anselin, 2010; and Lunn et al., 2000) are making spatially explicit analysis methods accessible to multiple disciplines. Recent examples include spatial estimation methods ranging from the relatively simple – merging together demographically similar contiguous areas or applying basic spatial smoothers (Greer et al., 2010 and Stevens et al., 2014) – to the far more complex Bayesian disease mapping using Monte Carlo Markov Chain (MCMC) estimation (Holowaty et al., 2010 and Asaria et al., 2012).

Yet, while there is a wide variety of methods and tools available, there is little guidance for investigators in choosing between them. Each approach has advantages and disadvantages, and they tend to differ in important ways in their assumptions about the spatial structure of the data, flexibility, and ease of implementation. While there have been numerous comparison studies within specific classes of estimators (e.g. model-based or non-model-based) (Marshall, 1991; Kafadar, 1994; Talbot et al., 2000; and Mungiole et al., 1999), including a recent study comparing more up-to-date regression methods for creating spatiotemporal estimates (Vaughan et al., 2015), there have been few side-by-side comparisons of the full range of methods.

One relatively comprehensive comparison was performed using simulated data (Lawson et al., 2000), but there are very few, if any, comparison studies using real-world data. Real-world data may differ from simulations in many important ways; even adjacent spatial units may differ greatly in size and shape, while population density or disease-relevant characteristics are unlikely to vary in ways that can be easily predicted. Even if prior simulations have explored specific irregularities, investigators may not have full knowledge of the factors unique to their data (e.g. underlying population characteristics). With little guidance regarding the potential advantages and disadvantages of each estimator, investigators would have to perform their own comparisons, or simply choose the method with which they were most comfortable or familiar.

In this paper, we compare a range of methods available for creating local disease rate estimates from georeferenced population health data, using ZIP code tabulation areas (ZCTAs) as our area of analysis. We examined Medicare claims data for all fee-for-service Medicare beneficiaries aged 65 years and older in the state of California, estimating the age-, sex-, and race-adjusted standardized incidence ratios (SIR) of acute myocardial infarction (AMI) for all ZCTAs during a single year (2010). Although a single state, California includes both widely varying local populations and geographical units with diversity in size

and shape, allowing the comparison of estimators in the face of several potential estimation challenges. We chose to use AMI for our analyses because this is a relatively well-measured, common event among Medicare beneficiaries. There tends to be less room for diagnostic ambiguity than for many common chronic diseases, while cardiovascular disease (of which AMI is one of the most dramatic manifestations) is the primary cause of death in the US. Therefore, AMI rates represent a relatively clearly defined measure of major illness burden.

We then computed the correlations and root mean squared error of each set of estimates with a gold standard consisting of SIRs calculated by pooling 5 years of data (2008-2012). This was chosen as the gold standard under the assumption that AMI rates change little over time- in fact, aggregating data over time is one common way to avoid a small numbers problem in publicly reported data (Casper et al., 2008 and Greer et al., 2010). Although there is sparse evidence on short-term stability of AMI rates in small areas, the disease is most often a manifestation of accumulated risk factors (e.g. smoking, physical inactivity, chronic stress) whose geographic distribution is unlikely to change dramatically within a 5 year time period. Finally, we also created maps for each set of estimates to provide a different perspective on potential biases introduced by each estimation method.

Methods

Medicare data

We examined Medicare claims data from the years 2008-2012. Beneficiaries between the ages of 65 and 99 on January 1 and without any months of coverage under managed care (HMO) Medicare during the entire year were included in that year's data. All fee-for-service beneficiaries were included, regardless of whether they had any claims on record for the study year. We recorded whether each beneficiary experienced an AMI during the calendar year, defining AMI by a claim indicating an inpatient stay with a primary diagnosis code 410.X0 or 410.X1 (X any integer 0-9), and length of stay greater than 1 day (CMS, 2011). These codes were specifically chosen to avoid capturing readmissions for previous AMI events. To estimate the proportion of the population that experienced an AMI during the year, an indicator variable was created with a value of zero if the beneficiary had no AMI visits during the year, and one if there were one or more visits. Additional variables included the beneficiaries' ZIP code of residence, age, sex, and self-identified race, which was coded as either White/Caucasian, or any other race/ethnicity.

Spatial data

As Medicare data are collected for administration purposes, the only spatial data associated with claims are the beneficiaries' mailing addresses; we used ZIP codes to geocode claims. First, an assignment file mapped every recorded ZIP code since 1990 to the physical location of a 2010 ZIP code. (Older ZIPs may be required if a Medicare enrollee had not updated his or her mailing address.) In some cases for the earlier years, there was no exact numerical match to a 2010 ZIP code, so the older ZIP was assigned to its nearest physical neighbor in 2010, resulting in a final dataset in which all available ZIP codes from 1990 to present were assigned a physical location. These ZIP codes were then mapped onto 2010 ZCTAs. For ZIPs whose numbers did not match a ZCTA, assignment was based on whether the centroid

of the ZIP was physically within the ZCTA polygon. Among all US ZIPS, a small minority of ZIP centroids (28) were located outside any ZCTA polygon; these were assigned to the nearest ZCTA. Finally, there were a few ZCTAs with no assigned Medicare beneficiaries. Each of these was merged with its largest (in population) bordering neighbor; there were 32,647 ZCTAs in the continental US. This paper focuses on the 1,741 ZCTAs within California. We focused on a single state because for some of the methods explored (e.g. MCMC Bayes estimators), the entire US comprised a prohibitively large sample size. California was selected among the states due to its larger area, diversity of urban and rural areas, and diversity of populations.

Study population

Basic summary statistics, including the overall AMI rate, were calculated for the Medicare beneficiaries both within the state of California and for the overall US population for 2010. The distribution of the number of beneficiaries per ZCTA was also determined for both populations.

Age-/sex-/race-adjustment

We used indirect standardization to adjust for differences in age, sex, and race across ZCTAs. Age was represented using 6 5-year categories (65-69, 70-74, 75-79, 80-84, 85-89, 90+), while race was represented with two categories. Our gold standard SIR values consisted of the number of observed beneficiaries with AMI over 5 years (2008-2012), divided by the age-/sex-/race-adjusted expected events for that time period.

Estimation methods

We used the expected number of beneficiaries with AMI and the observed number in each ZCTA in 2010 to create several sets of SIR estimates. The different estimation methods are summarized in Table 1.

Our first estimate, the fixed effect estimate, was the simple ratio of observed to expected events. This estimate can cause sparsely populated regions to appear as outliers, as a single event in an area with few expected events can cause its adjusted rate to be quite high (Leyland and Davies, 2005). We next explored a variety of alternative estimation methods, in four broad classes: spatially smoothed locally weighted averages, spatially smoothed medians, linear Empirical Bayes estimators, and Bayesian models implemented with MCMC. Most of the models included some type of spatial parameter, taking advantage of the strong spatial autocorrelation found in geographic data – neighboring regions are likely to be similar to each other (Tobler, 1970) – to improve estimation.

The spatial smoothing locally weighted average methods that we tested were all executed similarly, differing only in the way we defined each ZCTA's neighbors. In each of these methods, the estimate for each ZCTA is a weighted average of its own value and those of its neighbors. The behavior of this estimator depends on how a region's neighbors are defined; using too few neighbors results in under-smoothing such that nonsignificant outliers are retained in the final estimates, while too many neighbors will cause over-smoothing, and legitimate variation will be lost (Anselin et al, 2006).

We created locally weighted average estimators using four different neighborhood definitions. First, the queen contiguity neighborhood included all ZCTAs that shared a border, of any length, with the ZCTA whose rate was being estimated. The next definition included a constant number of nearest neighbors; we chose a number that resulted in the least populous neighborhood containing at least 5,000 beneficiaries. A minimum of 31 nearest neighbors was needed to ensure that the least sparsely populated neighborhood had at least 5,000 beneficiaries. Finally, we created neighborhoods that were designed to have a pre-specified minimum population of Medicare beneficiaries (either 5,000 or 10,000). To create these neighborhoods, we iteratively included each ZCTA's next nearest neighbor, and then assessed the total pooled number of beneficiaries, until that population was above the threshold (Talbot, 2000).

One neighborhood definition that we did not pursue was that of a fixed radius disc smoother. In this approach, all ZCTAs within a fixed distance of the ZCTA of interest would be included as its neighbors. While this approach has the intuitive appeal of allowing a distance that might have some a priori interest, it has been shown to be less desirable for data based on populations with widely varying densities. The same radius that in some rural areas may not be large enough to include all the centroids of a ZCTA's contiguous neighbors, could encompass nearly an entire large city in urban areas. Previous work has suggested that a neighborhood based on a constant minimum population threshold is preferable to a fixed distance radius smoother (Talbot, 2000). We included the neighborhood based on a constant number of nearest neighbors due to the fact that this neighborhood is considerably easier to define operationally than one based on a minimum population threshold, yet still ensures a minimum population in rural areas, while lessening the potential for oversmoothing in urban areas (relative to a smoother with constant radius).

The second class of spatial smoothers included two median methods. In the first, the weighted median SIR in each ZCTA's neighborhood (inclusive of the ZCTA being estimated) was determined and assigned to that ZCTA. Only the neighborhood with a constant number of nearest neighbors was used for this smoother, as the queen contiguity and adaptive bandwidth definitions each might result in neighborhoods too small (i.e. less than three) to select a median.

The second median approach was a method called weighted headbanging. Figure 1 demonstrates several important aspects of this approach. Here, a set of collinear triples are selected from among the nearest neighbors. Each triple consists of the ZCTA under consideration and two others, where the angle between them is greater than 135° ; candidate triple 1 in Panel A of Figure 1 demonstrates a triple that would be rejected due to an angle that is too acute. The set of triples retained for analysis is determined by the distance between the central ZCTA and the line segment connecting the other two neighbors; in Panel A, triplet 2 would be retained before triplet 3. Neighboring ZCTA centroids can be reused in a new triple.

Among this set of triples, a "high screen" and a "low screen" are determined; each triple contributes its lowest value to a set of numbers whose median is the low screen, and its highest value to the set of numbers whose median is the high screen. In weighted

headbanging, this process is weighted by the local population; a simplified example with three triples is shown in Panel B of Figure 1. Here, the low screen is the median of the weighted set {2,2,4,5,5}, while the high screen is the median of the set {7,7,7,8,9}. The decision to adjust the central point to the new median is based on the weights- in this case, the sum of the end point weights ($1+1+1+2+2+3=10$) is greater than the product of the center point's weight and the number of triples ($2*3=6$), so the central point is now assigned the median of the set {2,4,7}, or 4. The process can be repeated for multiple iterations, with every point updated from the first iteration before beginning the next. We included weighted headbanging because previous work has shown that it retains both edges between dissimilar areas and significant spikes, while still smoothing over likely nonsignificant outliers (Mungiole et al., 1999), yet it has not been compared thoroughly to other spatial smoothers. We used each ZCTA's 20 nearest neighbors to define 10 sets of triples, and ran the estimator for 10 iterations.

The next approach we considered was a linear Empirical Bayes estimator. Bayesian estimators allow each observation's estimate to be adjusted towards the mean of a larger sample; the extent of that adjustment will depend on the variability of the estimate for that observation and the overall sample (Leyland and Davies, 2005). The linear Empirical Bayes estimator uses a method of moments approach to estimate the variability of each ZCTA's value relative to the larger sample mean (Marshall 1991). We considered both an aspatial version of this estimator and a spatial version. In the aspatial version, estimates of the overall mean and variability were based on the values from all the ZCTAs in California, while in the spatial version, these estimates were based on the ZCTAs in each neighborhood. This spatial version was tested with the adaptive neighborhood that ensured a minimum population of 10,000, and the constant number of nearest neighbors.

Finally, we considered two different Bayesian estimators implemented with MCMC methods. This approach models the number of observed events in each ZCTA as a function of the expected number of events and the SIR associated with that ZCTA. In our first specification, the SIR estimated for each ZCTA included a random effect, but no explicit spatial component. In the second specification, the SIR for each ZCTA was assumed to encompass both a ZCTA-specific random effect and a local effect related to the mean of the ZCTA's contiguous neighbors.

This second approach is the Besag-York-Mollie model, the most commonly used Bayesian spatial model for disease mapping (Besag et al., 1991). The MCMC models were run multiple times to estimate the number of iterations required for model convergence; for both, we discarded a burn-in of 50,000 iterations, and then retained every 10th iteration for the next 100,000 iterations; the mean estimated SIR was recorded as the estimated SIR.

Estimator evaluation

For all methods tested, we generated estimated SIR values for each ZCTA using data from 2010. We then compared these estimates to our gold standard 5-year SIR estimates. We calculated both root mean square error and correlations (linear/Pearson and rank/Spearman); the root mean square error and linear correlations were weighted, while the rank correlations were unweighted. State- and city-level (Los Angeles and San Francisco) maps were also

created for each set of estimates and the 5-year pooled values. For reporting purposes, only those values based on a minimum population of at least 2,500 Medicare beneficiaries are displayed for each map.

Sensitivity Analyses

Several sensitivity analyses were performed to examine alternate neighborhood definitions and smoother specifications. For example, we tested the effect of using a constant number of nearest neighbors that ensured a minimum population of 1,000 or 10,000, or an adaptive neighborhood with minimum population as small as 1,000 or as large as 20,000. We also tested the median smoother using different neighborhood sizes (or numbers of triples for weighted headbanging), as well as different numbers of iterations. A more detailed explanation of each specification of the estimators we tested is available in Supplementary Appendix 1.

Statistical Software

Summary statistics and age-/sex-/race-expected events were calculated using Stata 13 software. All spatially smoothed estimates, including the linear EB estimates, were calculated using Python, specifically the PySAL library (Rey and Anselin, 2010). The MCMC models were estimated using WinBUGS software (Lunn et al., 2000). ArcGIS version 10.2 was used to create all maps and perform geographic data management. Supplementary Appendix 2 includes further details on how we implemented several of the spatial smoothers, including creating the adaptive bandwidth weights and implementing the linear spatial EB.

Results

Study Population

There were 2.43 million beneficiaries aged 65-99 receiving fee-for-service (traditional) Medicare in California in 2010; pooling data from 2008 – 2012 resulted in 12.19 million person-years of observation. Table 2 presents summary statistics of the study population, as well as its distribution across 1,741 California ZCTAs. For reference, the table includes statistics from the overall US population of Medicare fee-for-service beneficiaries, which was used as the standard for calculating expected AMI events among demographic subgroups. The overall US rates were used as the standard so that estimated incidence ratios for California ZCTAs could be compared not only to each other, but also to those expected in the rest of the country. In California in 2010, there was a mean of 1,393 beneficiaries per ZCTA, with a range of 1 – 7,229 (25th percentile: 177, 75th percentile: 2,313). Across all US ZCTAs, there was a mean of 853 beneficiaries, with a range of 1 – 20,925. In 2010, 4.9 per 1,000 Medicare beneficiaries in California experienced an AMI, compared to 6.5/1,000 in the United States.

Spatial neighborhood definitions

For the spatial smoothers, each estimate was based on a population that depended on the neighborhood used. The number of beneficiaries in each neighborhood defined by queen contiguity averaged 8,264, with a median of 7,747 and ranged from 0 (by default, islands are

assigned a value of 0) to 39,604. To ensure at least 5,000 beneficiaries in the smallest neighborhood defined by a constant number of nearest neighbors, we included the nearest 31 ZCTAs. The mean number of beneficiaries among these neighborhoods was 46,665, and the median 47,915. Among the neighborhoods created with a threshold of 5,000 in the adaptive approach, the mean and median numbers of beneficiaries were 6,346 and 6,061; among those with a threshold of 10,000, the mean and median were 11,263 and 10,982.

Estimator performance

In Table 3, we present the population-weighted root mean square error for SIR estimates created under each approach, relative to the gold standard SIRs created by pooling data from the years 2008-2012. Two sets of root mean square error are presented; the first column shows the statistics for all ZCTAs, and in the second, for only those ZCTAs whose 5-year estimates were based on a population of at least 1,000 beneficiaries (1,287 ZCTAs). Across both the entire universe of ZCTAs and those ZCTAs with at least 1,000 beneficiaries, the Besag-York-Mollie model (a Bayesian model implemented with MCMC estimation) produced the lowest root mean square error, followed by the spatial linear Empirical Bayes estimator.

In Table 4 we present the correlations between each set of SIR estimates and the 5-year gold-standard SIRs, both in terms of population-weighted Pearson (linear) correlation coefficients, as well as unweighted Spearman's correlations, which measure how closely the ZCTA rankings agree. For both the complete set of ZCTAs and the subset of more highly populated ZCTAs, the highest linear correlations were found for the estimates created using the Besag-York-Mollie model, closely followed by those from the spatial linear Empirical Bayes models. In the Spearman correlations, the raw data estimates display the highest correlation among all ZCTAs, but the BYM and spatial linear EB models outperform the raw data in predicting ranks among more populous ZCTAs.

Figure 2 displays maps of each of the estimates for the state of California, and Figures 3 and 4 display maps for the Los Angeles and San Francisco/Bay Area regions, respectively. The categories chosen for display are those that separate the gold standard 5-year SIRs into 7 roughly equal categories. Due to confidentiality concerns related to the reporting of health data for small populations, we only represent data for ZCTAs whose estimates are based on a population of at least 2,500 beneficiaries (1,103 ZCTAs for the 5-year estimates).

Sensitivity Analyses

In sensitivity analyses examining alternative neighborhood definitions or alternative specifications for the median smoothers and weighted headbanging, our findings were consistent; in general, the neighborhood definitions and specifications we present here were among the best performing within each class of smoother. The full results for each different specification of each smoother are presented in Supplementary Appendix 1. In particular, the spatial linear Empirical Bayes estimators performed nearly identically when applied using neighborhoods with a constant number of nearest neighbors that ensured either a smaller (1,000) or larger (10,000) minimum population.

Discussion

In this study, we explored several different methods for creating local estimates of population health measures. Such methods may be necessary when an investigator wishes to explore local health phenomena where the populations are too small to create statistically stable estimates using raw data. While aggregating data to larger regions such as states is one solution, the methods we study here allow the exploration of variations within relatively small areas.

The approaches we compared ranged in complexity and ease of implementation. Although they require more time and effort to implement, Bayesian models with MCMC estimation produced the estimates with the smallest root mean square error and the largest linear correlation with the 5-year estimates. Despite advances in computing speeds and software, the Besag-York-Mollie model took several hours to run, and we only used data for a single state. The spatial linear Empirical Bayes estimators had the next smallest root mean square error and largest correlations with 5-year estimates, suggesting that this more accessible method may be a reasonable alternative, especially for those with particularly large datasets. These estimates, implemented using a method-of-moments approach to estimate the variance of each ZCTA's data, took only a few minutes to calculate.

Among the spatial smoothers, investigators also face choices in defining the local neighborhood used to create estimates. The adaptive neighborhood, which required a minimum population threshold, is intuitively appealing in that it allows the inclusion of greater or fewer neighbors in response to local data. Among the locally weighted average smoothers, the adaptive neighborhood performed better than a constant number of neighbors, consistent with previous research (Talbot et al., 2000). Yet, within the context of a linear Empirical Bayes estimator, the neighborhood based on a constant number of neighbors performed similarly. Among the spatial smoothers, the spatial linear Empirical Bayes smoother appears to be more robust to neighborhood misspecification, as it allows the identification of local variations even with the use of a neighborhood consisting of 31 nearest neighbors. The linear Empirical Bayes estimator shrinks each estimate to the neighborhood's mean to an extent dictated by the data; a more populous ZCTA will have its value adjusted less, even in a large neighborhood. As the adaptive neighborhood definition requires additional custom coding, the constant number of neighbors may be an attractive alternative, at least in the context of a linear Empirical Bayes smoother. However, this neighborhood choice is probably not the best when using a locally weighted average smoother, as it will likely lead to oversmoothing, especially in urban areas.

While the raw data displayed relatively high correlations with the 5-year SIRs, the root mean square error was also higher than that for all other estimators, indicating inaccuracy. The differences in findings for the linear and rank correlations likely result from the fact that the weighted linear correlations gave higher emphasis to ZCTAs with more observations, thereby emphasizing those whose 5-year estimates were more likely to be stable. Sparsely populated ZCTAs are more likely to give rise to outlier SIR values (Leyland and Davies, 2005); even when aggregated with an additional 4 years' worth of data, an outlier in one year may affect the 5-year pooled estimate. Such outliers may be driving the high rank

correlations found between the raw estimates and our gold standard. While these high rank correlations may seem to suggest that estimation methods are unnecessary if one's goal is simply to rank the regions, the high root mean square error indicates inaccuracy.

Additionally, among the more highly populated ZCTAs (where the gold standard is likely more stable), both types of Bayesian estimators (linear and MCMC), displayed higher rank correlations.

Mapping the estimates also highlights the characteristics of each estimator. The state maps show that the two aspatial estimators (global linear Empirical Bayes and Random Effects MCMC Bayes) tend to adjust the values to a global mean; more sparsely populated areas are adjusted more dramatically towards that mean. Additionally, it seems likely that the locally weighted average and median smoothers with a constant number of 31 nearest neighbors are oversmoothing the data, given that large swaths of each of the Los Angeles and San Francisco metropolitan regions are represented with similar values, while most of the other estimators depict substantial variation within those regions.

The maps also demonstrate that different estimators can vary dramatically in their estimates for the same ZCTAs. This may be due to the fact that even with spatial smoothers, individual estimates can be based on relatively few observations. One possible solution to this problem would be to combine several years of data, and then apply an estimation method (e.g. Besag-York-Mollie MCMC or linear Empirical Bayes). Pursuing more stable estimates may be especially important if policy decisions or public health interventions will be based on the estimates created. Such uses of the data may also suggest the use of an MCMC Bayes estimator, as this approach produces credible interval estimates (akin to confidence intervals); the other approaches produce point estimates of the SIR.

Our study has several limitations. First, as our gold standard, we used estimates created by combining 5 years' worth of data, under the assumption that AMI rates would vary little over time. Yet, for many ZCTAs, even 5 years' worth of data may be unlikely to provide a large enough population for accurate estimates. We partly addressed this issue by calculating the root mean square error and correlations both for all ZCTAs and for the subset with at least 1,000 beneficiaries included in the 5-year estimates. Even with this exclusion, the 5-year estimates are likely imperfect representations of the true risk of AMI in the local population. One possible solution to this problem might have been to use an alternative data source as our gold standard, but such an approach would have had other drawbacks. All-payer registries, (such as the SPARCS registry in New York) usually cover different populations than what we studied here (i.e. all adults, rather than Medicare beneficiaries over the age of 65), and aren't available for all areas. Self-reported data on AMI incidence, available from some health surveys, may not be very accurate, especially among older populations (Author et al., 2015). For neither of these sources are fine-grained geographic data publicly available.

Despite the imperfections in our gold standard, we chose to pursue this example because the use of real-world data allows an examination of how each estimator performs outside the artificial confines of simulated examples. Previous simulation studies have attempted to create situations akin to real-world data by including structured areas of high or low rates, often with additional statistical noise (e.g. Kafadar, 1994 and Lawson et al., 2000), but data

representing human processes rarely follow patterns that can be defined by a function of x- and y- coordinates. Having to rely on an imperfect gold standard is a drawback of using real-world data.

A second limitation to our study was that we were not able to compare all possible methods for estimating disease rates for small areas. There have been many different approaches proposed over time in the literature, and it would not have been possible to compare all of them. In particular, one small area estimation method involving the combination of detailed survey and demographic data has been used relatively often to create estimates of health risks and behaviors such as smoking, hypertension, or physical activity (Twigg et al., 2000, Olives et al., 2013, and Dwyer-Lindgren et al., 2013). However, the very limited demographic data available in Medicare claims precluded this approach. We selected the estimators compared in this paper because they represent many of the approaches used most often, and are accessible through the use of freely available software packages (Anselin et al., 2005, Rey and Anselin, 2010, and Lunn et al., 2000). We believe these approaches provide a good overview of the range of methods commonly implemented.

A third limitation to our study is our reliance on fee-for-service Medicare claims data for our population health data; basically, our study population was limited to Medicare beneficiaries aged 65 or older, whose health care had not been contracted to a managed care provider during the year of study. Different patient populations in different areas may be more or less likely to be covered under a managed care plan; those covered by managed care plans may be healthier, and therefore less likely to experience an AMI (Mello et al., 2003, and Brown et al., 2011). In areas with high managed care penetration, therefore, we may not be estimating the true underlying health of the local population. Yet, as our intent was foremost to compare the performance of the estimators, this is less of a concern if we can assume that managed care penetration did not change greatly over the period of our study. In fact, there was, on average, only about a 4% difference between the maximum and minimum HMO enrollment rates within each county over the time period studied (Medicare, 2015).

Due to the nature of the data we used, we were unable to create estimates for areas smaller than ZCTAs. It is possible that analyses at other levels of aggregation, such as census tracts, may have revealed even more fine-grained patterns of variation, especially in the more densely populated urban regions. On the other hand, had we chosen to first aggregate our data to larger regions, such as counties, each region might have had more observations with which to create more reliable estimates, but this would have likely masked meaningful finer grained variations. Our assumption that applying spatial smoothers to ZCTA-level data would allow larger patterns to appear if present appears to have been valid; in some cases, multiple neighboring ZCTAs show similar values. We believe that this approach was the best we could pursue given the limitations of our data.

Finally, our estimation approaches did not include any adjustments for benchmarking to the total population rates, as has been used in some small area estimation work (Datta et al., 2011). Including such adjustments may have improved our estimates, but it is difficult to speculate which estimators would have been affected most, and these additional analyses were not the focus of the current work.

In conclusion, our findings suggest that the Besag-York-Mollie model implemented with MCMC estimation outperforms most basic spatial smoothers, but the next best performing estimator, the spatial linear Empirical Bayes model, may represent a good option for investigators with limited MCMC estimation experience or very large datasets. For more accurate estimates, investigators may wish to combine multiple years of data before applying a spatial smoother or estimator.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge funding support from the National Institute on Aging (Grant No. P01 AG19783-11).

Funding: This work was supported by the National Institute on Aging (Grant No. P01 AG19783-11).

References

- Anselin L, Syabri I, Kho Y. GeoDa: An introduction to spatial data analysis. *Geogr Anal.* 2005; 38(1): 5–22.
- Anselin, L.; Lozano, N.; Koschinsky, J. [Accessed February 15, 2015] Rate transformations and smoothing. https://geodacenter.asu.edu/pdf/smoothing_06.pdf. Revised January 31, 2006
- Asaria P, Fortunato L, Fecht D, et al. Trends and inequalities in cardiovascular disease mortality across 7932 english electoral wards, 1982-2006: Bayesian spatial analysis. *Int J Epidemiol.* 2012; 41(6): 1737–49. discussion 1750-2. [PubMed: 23129720]
- Author et al., 2015 [details removed for review].
- Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics.* 1991; 43(1):1–20.
- Brown J, Duggan M, Kuziemko I, Woolston W. How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program. *NBER Working Paper Series.* 2011; 16977
- Casper, ML.; Nwaise, IA.; Croft, JB.; Nilasena, DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2008.
- Centers for Disease Control and Prevention (CDC). SMART: BRFSS city and county data. 2013
- Centers for Medicare and Medicaid Services (CMS). Medicare hospital quality chartbook. 2011
- Datta GS, Ghosh M, Steorts R, Maples J. Bayesian benchmarking with applications to small area estimation. *Test.* 2011; 20(3):574–588.
- Dwyer-Lindgren L, Freedman G, Engell RE, et al. Prevalence of physical activity and obesity in US counties, 2001-2011: A road map for action. *Popul Health Metr.* 2013; 11(7)
- Greer, SA.; Nwaise, IA.; Casper, ML. Atlas of heart disease hospitalizations among medicare beneficiaries. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2010.
- Holowaty EJ, Norwood TA, Wanigaratne S, et al. Research feasibility and utility of mapping disease risk at the neighbourhood level within a canadian public health unit: An ecological study. *Int J Health Geo.* 2010; 9(21)
- Kafadar K. Choosing among two-dimensional smoothers in practice. *Computational Statistics & Data Analysis.* 1994; 18(4):419–439.
- Krieger N, Chen JT, Waterman PD, et al. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level

matter?: The public health disparities geocoding project. *Am J Epidemiol.* 2002; 156(5):471–482. [PubMed: 12196317]

Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Clark A, et al. Divino F. Disease mapping models: an empirical evaluation. *Disease Mapping Collaborative Group. Statistics in medicine.* 2000; 19(17-18):2217–41. [PubMed: 10960849]

Leyland AH, Davies CA. Empirical bayes methods for disease mapping. *Stat Methods Med Res.* 2005; 14(1):17–34. [PubMed: 15690998]

Losina E, Wright EA, Kessler CL, et al. Neighborhoods matter: Use of hospitals with worse outcomes following total knee replacement by patients from vulnerable populations. *Arch Intern Med.* 2007; 167(2):182–186. [PubMed: 17242320]

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and computing.* 2000; 10(4):325–337.

Marshall RJ. Mapping disease and mortality rates using empirical bayes estimators. *Applied Statistics.* 1991; 49(2):283–294.

Medicare Advantage Enrollees as a Percent of Total Medicare Population. The Henry Kaiser Family Foundation; <http://kff.org/medicare/state-indicator/enrollees-as-a-of-total-medicare-population/> [Last accessed: July 30, 2015]

Mello MM, Stearns SC, Norton EC, Ricketts TC. Understanding biased selection in medicare HMOs. *Health Serv Res.* 2003; 38(3):961–992. [PubMed: 12822921]

Mungiole M, Pickle LW, Simonson KH, White AA. Application of a weighted head-banging algorithm to mortality data maps. *Stat Med.* 1999; 18(23):3201–3209. [PubMed: 10602145]

Olives C, Myerson R, Mokdad AH, Murray CJ, Lim SS. Prevalence, awareness, treatment, and control of hypertension in united states counties, 2001–2009. *PLoS One.* 2013; 8(4):e60308. [PubMed: 23577099]

Rey, SJ.; Anselin, L. PySAL: A python library of spatial analytical methods. In: Fischer, MM.; Getis, A., editors. *Handbook of applied spatial analysis.* New York, NY: Springer; 2010. p. 175-193.

Sakshaug JW, Raghunathan TE. Generating synthetic microdata to estimate small area statistics in the american community survey. *Statistics in Transition.* 2014; 15(3):341–369.

Stevens CD, Schriger DL, Raffetto B, et al. Geographic clustering of diabetic lower-extremity amputations in low-income regions of california. *Health Aff (Millwood).* 2014; 33(8):1383–1390. [PubMed: 25092840]

Talbot TO, Kulldorff M, Forand SP, Haley VB. Evaluation of spatial filters to create smoothed maps of health data. *Stat Med.* 2000; 19(17-18):2399–2408. [PubMed: 10960861]

Tobler WR. A computer movie simulating urban growth in the detroit region. *Economic geography.* 1970; 46:234–240.

Twigg L, Moon G, Jones K. Predicting small-area health-related behaviour: A comparison of smoking and drinking indicators. *Soc Sci Med.* 2000; 50(7):1109–1120. [PubMed: 10714931]

Vaughan AS, Rosenberg E, Shouse RL, Sullivan PS. Connecting Race and Place: A County-Level Analysis of White, Black, and Hispanic HIV Prevalence, Poverty, and Level of Urbanization. *Am J Public Healh.* 2014:e1–e8.

Vaughan AS, Kramer MR, Waller LA, Schieb LJ, Greer S, Casper M. Comparing methods of measuring geographic patterns in temporal trends: an application to county-level heart disease mortality in the United States, 1973 to 2010. *Ann Epidemiol.* 2015; 25(5):239–335.

Wong D, Sun M. Handling data quality information of survey data in GIS: A case of using the american community survey data. *Spatial Demography.* 2013; 1(1):3–16.

Highlights

We compared methods to estimate local rates of Acute Myocardial Infarction (AMI).

ZCTA-level rates were examined among Medicare beneficiaries in California.

Single-year estimates were compared to those created from five years' pooled data.

Bayesian analyses accounting for spatial relationships performed best.

Linear methods performed nearly as well as those implemented using MCMC estimation.

Figure 1. Demonstration of Weighted Headbanging Approach

Panel A: Comparison of candidate triples to be included in estimating the central point's value.

Panel B: Illustration of values and weights (in parentheses) for the central point and its neighboring triples.

Figure 2. Maps of Standardized Incidence Ratio Estimates for Acute Myocardial Infarction Rates Among Medicare Beneficiaries in California

Standardized incidence ratios (SIRs) were estimated for Medicare beneficiaries aged 65-99 residing in each ZCTA in the state of California. Observed subpopulation rates in the overall US were used to calculate the expected number of events per ZCTA. The Gold Standard estimates reflect the O:E ratio in each ZCTA, using data pooled across years 2008-2010. All other estimates were created using a single year of data (2010). Only those estimates based on at least 2,500 beneficiaries are displayed.

Figure 3. Maps of Standardized Incidence Ratio Estimates for Acute Myocardial Infarction Rates Among Medicare Beneficiaries, Los Angeles Metropolitan Region

Standardized incidence ratios (SIRs) estimated for Medicare beneficiaries aged 65-99 residing in each ZCTA in the state of California. Total and subpopulation rates in the overall US were used to calculate the expected number of events per ZCTA. The Gold Standard estimates reflect the O:E ratio in each ZCTA, using data pooled across years 2008-2010. All other estimates were created using a single year of data (2010). Only those estimates based on at least 2,500 beneficiaries are displayed.

Figure 4. Maps of Standardized Incidence Ratio Estimates for Acute Myocardial Infarction Rates Among Medicare Beneficiaries, San Francisco/Bay Area Metropolitan Region

Standardized incidence ratios (SIRs) estimated for Medicare beneficiaries aged 65-99 residing in each ZCTA in the state of California. Total and subpopulation rates in the overall US were used to calculate the expected number of events per ZCTA. The Gold Standard estimates reflect the O:E ratio in each ZCTA, using data pooled across years 2008-2010. All other estimates were created using a single year of data (2010). Only those estimates based on at least 2,500 beneficiaries are displayed.

Table 1
Summary of Estimation Approaches

Estimator	Description	Neighborhoods tested
Fixed effects/Raw	Ratio of observed:expected beneficiaries with AMI events	N/A; observed and expected events based on data from each ZCTA alone
Locally weighted average	Weighted average of the ratio of observed:expected beneficiaries with AMI events	Queen contiguity Constant # nearest neighbors (31) Adaptive size, minimum population [*] (5,000) Adaptive size, minimum population [*] (10,000)
Median	Median ratio of observed:expected beneficiaries with AMI events	Constant # nearest neighbors (31)
Weighted headbanging	Weighted median that takes into account the 2-dimensional nature of spatial data	20 nearest neighbors- among these, ZCTAs selected to create 10 unique triplets with the widest angle (the central ZCTA serving as the vertex)
Aspatial Linear Empirical Bayes	Bayesian estimator that uses a linear (method of moments) approximation of each ZCTA's variability relative to that of the overall mean	N/A; each ZCTA compared to the pooled mean across all ZCTAs
Spatial Linear Empirical Bayes	Bayesian estimator that uses a linear (method of moments) approximation of each ZCTA's variability relative to that of the neighborhood mean	Constant # nearest neighbors (31) Adaptive size, minimum [*] (10,000)
MCMC Bayes RE	# Observed beneficiaries w/AMI events modeled as a function of the expected; each ZCTA has a random effect associated; estimation carried out with Monte Carlo Markov Chain modeling	N/A; aspatial model
MCMC Bayes BYM	MCMC Bayes RE, with an additional random effect associated with the mean of each ZCTA's neighbors	Queen contiguity defines each ZCTA's neighbors

^{*} population refers to the number Medicare beneficiaries

Table 2
Medicare Beneficiary Summary Statistics and Distribution Across ZCTAs in California
and the United States

	California	United States
Number beneficiaries	2,425,411	28,085,479
Male (%)	44.7	43.5
African-American/other race (%)	27.6	14.1
Age distribution (% in each category):		
65 – 69	31.8	30.8
70 – 74	22.7	22.6
75 – 79	17.3	17.9
80 – 84	13.8	14.4
85 – 89	9.5	9.4
90 – 99	4.9	4.9
Overall AMI rate	4.9/1,000	6.5/1,000
# ZCTAs	1,741	32,647
Distribution of beneficiaries/ZCTA		
Mean	1,393	853
Min	1	1
25 th perc	177	95
50 th perc	1,051	318
75 th perc	2,313	1184
Max	7,229	20925

Table 3
Root Mean Square Error Between Each Set of Standardized Incidence Ratio Estimates
and 5-Year Pooled Values

	All ZCTAs (n = 1,741)	Subset of larger ZCTAs (n = 1,287)
Raw SIR estimate	0.813	0.328
Locally weighted average		
Queen contiguity	0.640	0.328
Constant number nearest neighbors (31)	0.508	0.241
Adaptive bandwidth, minimum population (5,000)	0.530	0.312
Adaptive bandwidth, minimum population (10,000)	0.518	0.268
Median (31 nearest neighbors)	0.519	0.231
Weighted Headbanging	0.535	0.251
Linear Empirical Bayes		
Global (aspatial)	0.504	0.210
Constant number nearest neighbors (31)	0.498	0.196
Adaptive bandwidth, minimum population (10,000)	0.498	0.201
MCMC Bayes		
Random effects only	0.504	0.209
Spatial and random effects (Besag-York-Mollie)	0.497	0.191

Table 4
Pearson (Linear) and Spearman (Rank) Correlations Between Each Set of Standardized Incidence Ratio Estimates and the Gold Standard 5-Year Pooled Values

	Pearson (linear)		Spearman (rank)	
	All ZCTAs (n = 1,741)	Subset of larger ZCTAs (n = 1,287)	All ZCTAs (n = 1,741)	Subset of larger ZCTAs (n = 1,287)
Raw SIR estimate	0.699	0.695	0.565	0.625
Locally weighted average				
Queen contiguity	0.397	0.429	0.260	0.420
Constant number nearest neighbors (31)	0.473	0.492	0.352	0.467
Adaptive size, minimum population (5,000)	0.492	0.510	0.432	0.530
Adaptive size, minimum population (10,000)	0.515	0.537	0.423	0.541
Median (31 nearest neighbors)	0.556	0.588	0.339	0.522
Weighted Headbanging	0.579	0.607	0.403	0.537
Linear Empirical Bayes				
Global (aspatial)	0.666	0.696	0.512	0.638
Constant number nearest neighbors (31)	0.702	0.735	0.471	0.665
Adaptive size, minimum population (10,000)	0.713	0.745	0.495	0.674
MCMC Bayes				
Random effects only	0.668	0.698	0.505	0.638
Spatial and random effects (Besag-York-Mollie)	0.735	0.772	0.481	0.705