Title
Residuals in space: Potential pitfalls and applications from single-institution survival analysis

Authors
Arabadjis, Sophia D
Sweeney, Stuart H

Publication Date
2024-06-01

DOI
10.1016/j.sste.2024.100646

Residuals in Space: Potential Pitfalls and Applications from Single-Institution Survival Analysis

Sophia D Arabadjis[1], **Stuart Sweeney**[1]

[1]*Department of Geography*
University of California
Santa Barbara, CA 93106-2150

March 28, 2025

Corresponding author:
Sophia D Arabadjis
PhD Student
Department of Geography
University of California
Santa Barbara, CA 93106-4060
email: stuart.sweeney@ucsb.edu

Abstract

In practice, survival analyses appear in pharmaceutical testing, procedural recovery environments, and registry-based epidemiological studies, each reasonably assuming a known patient population. Less commonly discussed is the additional complexity introduced by non-registry and spatially-referenced data with time-dependent covariates in observational settings. In this short report we discuss residual diagnostics and interpretation from an extended Cox proportional hazard model intended to assess the effects of wildfire evacuation on risk of a secondary cardiovascular events for patients of a specific healthcare system on the California's central coast. We describe how traditional residuals obscure important spatial patterns indicative of true geographical variation, and their impacts on model parameter estimates. We briefly discuss alternative approaches to dealing with spatial correlation in the context of Bayesian hierarchical models. Our findings/experience suggest that careful attention is needed in observational healthcare data and survival analysis contexts, but also highlights potential applications for detecting observed hospital service areas.

keywords: survival analysis, electronic medical records, administrative data, spatial residuals, Bayesian hierarchical models, hospital service areas

1 Introduction

The assessment of residuals in the model fitting process for survival analysis is more complex than a traditional linear model (or even generalized linear model) framework (24; 19) and the complexity only increases with spatially-referenced data and time-dependent covariates (2; 12). In practice, survival analyses are often used in pharmaceutical testing, procedural recovery environments, and registry-based epidemiological studies, each reasonably assuming a set and known patient population (12; 4; 2; 3; 14; 5; 30). While interval censoring makes an (important) appearance in this literature (24; 19), less commonly discussed is the additional complexity introduced by non-registry and spatially referenced data combined with time-dependent covariates in observational settings. This short report was born of questions we encountered in implementing an extended Cox proportional hazard model to assess the effects of wildfire evacuation on risk of a secondary cardiovascular events (CVE) for patients of a singular specific healthcare system on the California's central coast (1).

While the main paper discusses the nuts and bolts of the findings and the model specifications, in this short report we present a concise narrative of how traditional residuals in the survival framework can obscure important spatial patterns indicative of true geographical variation, and their impacts on parameter estimates in the model. We briefly discuss alternative approaches to dealing with spatial patterns in the context of Bayesian hierarchical and spatial survival models. Our findings/experience suggest that careful attention is needed in observational healthcare data and survival analysis contexts, but also suggest that survival analysis may have applications for detecting observed hospital service areas using patient visit records.

2 Methods

This study has been approved by the Santa Barbara Cottage Hospital Institutional Review Board.

We collected electronic medical records of all patients who arrived at Santa Barbara Cottage Hospital system with a qualifying cardiovascular event between October 1, 2016 and June 1, 2019 and an in-county address(es) (n=2948). Qualifying cardiovascular events were identified using ICD-10 codes selected by a physician collaborator. Qualifying diagnoses included all child codes within the following: I10, I11, I13, I15, I20-I25, I40, I42-I52, I71, R00, R07.1, R07.2, R07.8, R07.9, and R94.3, which encompass a range of cardiac dysfunctions including severe diagnoses (stroke, cardiac arrest, acute myocardial infarction) and potentially less severe diagnoses (hypertensive heart disease and chest pain). These diagnosis codes align well with other studies using similar methods (20; 27; 9; 11; 7; 21; 13; 6; 22; 15; 29; 28). All patient addresses during the study period were geo-coded using 2016 Santa Barbara County Assessor parcel data.

For this set of patients, we also captured secondary cardiovascular events over the same period (n=473), and determined which patients had experienced an evacuation order for any of the three fires that occurred during the period (n=393). Patients were considered exposed to an evacuation if and only if their current address location at the time of the evacuation order fell within an evacuation zone polygon. The details of the data preparation is discussed elsewhere (1), but relevant to this analysis we only used a singular address for each patient in the sample and that address was the address associated with the period of the evacuation orders (no prior- or post-evacuation order addresses).

2.1 Statistical Analysis

Our model of interest takes the following form:

$$h(t|x_i) = h_0(t)e^{\beta x_i(t)}$$

where h_0 is the unestimated baseline hazard (or risk) of secondary cardiovascular events, and $x_i(t)$ is the value of the i^{th} patient's evacuation event indicator at any time point t, which is set to 0 until a patient experiences an evacuation order, at which point it takes the value 1. The estimate of interest is the value of $\hat{\beta}$, or the multiplicative effect of evacuation exposure on the risk of secondary CVE.

To assess general model fit, we use Schoenfeld residuals. To assess the model fit and sensitivity to spatial patterning, we use martingale and jackknife residuals calculated at several scales. Martingale residuals are defined as the observed count at a given time ($N_i(t)$) less the expected count per our model formulation at a given time ($\hat{E}_i(t)$). We sum these residuals by their unique patient ID, and use the maximum value to identify individuals who survived too long (or longer than expected) between cardiovascular events. The minimum value is indicative of individuals who "died too soon" or had survival times between events that were less than expected. (In this case, some of these lower-end outliers did not survive their second cardiovascular event.) We then sum or average these martingale residuals by city name, census tract, and zip code tabulation area (ZCTA) to inspect a-spatial and spatial patterning of the residuals.

To assess the sensitivity of the $\hat{\beta}$ (evacuation effect estimate) to spatial patterning, we rely on jackknife or case-deletion residuals. These residuals are calculated by "leaving-one-out", re-fitting the model and noting the change in the parameter estimate. In the survival context, "leaving one out" usually refers to a single individual. We exploit this and expand it to "leave one group out", which is common place in cross-validation techniques (10). We use both individual-level jackknife residuals in addition to city name, census tract, and zip code tabulation area (ZCTA) jackknife residuals to assess the sensitivity of the evacuation order effect.

3 Results

The model result suggests a hazard ratio of 1.2833 ($e^{0.2494}$) or an increase in risk of 28%. (The hazard with robust standard error is shown in Table 1.) The effect is significant at $\alpha = 0.1$ level in this model.

The Schoenfeld residuals for the model are presented in Figure 1. The orange line is the $\hat{\beta}$ estimate of evacuation order plotted against time (t). The two grey bands demark the 95% confidence pointwise interval (24; 23). The relative stacking or linear effect in the plotted points is due largely to the simplicity of the model – evacuation order is a binary covariate. The relative flatness of the orange line suggests that the effect of the evacuation order is consistent across time, and proportionality assumptions are appropriate.

Using the summed martingale residuals (by patient identifier), we can assess the outliers. Negative outliers are those that survived not as long as expected; positive outliers are those that survived longer than expected. Figure 2 displays the martingale residuals sorted by size and direction of the residual. From the graphic, there appears to be around 50 or so individuals for whom the model would suggest should have survived longer, though the magnitude of the residual is relatively small. However, it is clear there are several hundred individuals that have survived longer than expected, for whom the model does not perform well.

This result is both interesting and concerning. The magnitude of the residual is high, indicating a large difference between expected and observed survival time. It is also unclear how exposure to an evacuation order enters into this subpopulation. If we assume no individuals in this outlier population were exposed to an evacuation order, then these patients survival times would contribute to a longer period between events in the un-estimated baseline hazard, which could inflate the effect of the exposure to an evacuation order. If, on the other hand, all of individuals in this outlier population were exposed to an evacuation order, then these elongated survival times would contribute (mostly) to the evacuation effect, suggesting a protective effect for evacuation exposure over the un-estimated baseline hazard. For a mix of evacuation exposed and un-exposed within the outlier subpopulation, it is harder to develop intuition.

Given the significant positive results of the model fit, there is reason to be concerned about the case where no patients, or only a small number, in the martingale outlier group were exposed to an evacuation order. In Figures 3 and 4, we plot the average martingale residual values by ZCTA and census tract. Lighter blue values represent higher average martingale residual values for patients in the area; darker blue values represent lower average martingale residual values in the area. The orange point is the location of the dominant hospital system. In Figure 3, there is some evidence of a distance decay– tracts with the highest values seem to be located further from the hospital system. In Figure 4, this effect is somewhat attenuated, though high average martingale residual areas are still generally located further from the hospital system.

In figure 5, we plot the summed positive and negative martingale residuals by Euclidean distance (km) from the hospital system. The clumping of individuals by distance is expected, as the settlements in the region are dispersed. The negative martingale residual values do not appear to have much of a relationship with distance. As distance increases, there is not much evidence to suggest a change in the residual martingale values (perhaps slight trend towards zero.) Surprisingly, the positive martingale residual values also do not appear to trend positively with increasing distance. However, we did not use network distance, which could further segregate the data.

We also investigate this spatial trend by testing the sensitivity of the $\hat{\beta}$ evacuation estimate to various individuals, city names, ZCTAs, or census tracts using the jackknife residuals. In Figure 6, panel (a) shows the sensitivity of evacuation estimate to removing single individuals. While some individuals would appear to make the estimate even higher, mostly individuals seem to be clustered around the estimated mean. In panel (b) we fit models while leaving out one of the sixteen city names in the patient addresses. We see that removing one specific city name lowers the beta estimate to almost zero (no effect) while removing a different specific city name would raise the beta estimate above 0.3. These bars are not weighted by the number of participants in each city (and the first bar is "Santa Barbara," which, by leaving out, decimates the sample size and makes that fitted model extremely unstable.) Panel (c) and (d) also show the beta estimates of models fit if leaving out one

ZCTA or one census tract. Again the ZCTA scale suggests that removing one or two areas would greatly increase the estimate of evacuation order effect, and removing one or two areas would decrease estimate of the effect. The census scale suggests less of a drop in removing any one census tract, but a similar increase in effect size if removing selected census tracts.

These same results for ZCTA and census tract are mapped in Figures 7 and 8 to assess spatial auto-correlation or trends. In Figure 7, there are two trends of interest. First, excluding the large southwestern ZCTA (that encompasses the southwestern tip of the county), would dramatically lower the evacuation effect, as would excluding the tract just east of the hospital system location. Second, excluding any of the ZCTAs just north or slightly east of the hospital system would substantially raise the evacuation effect estimate. The interpretation of these trends is somewhat difficult given that the estimates themselves are a product of both the sample size, the number of secondary CVEs observed, and the overall number of evacuation order exposed (which are also not uniformly distributed across these areas.) The effect of excluding the western tract (within which there were no evacuation orders) suggests that patients from that area are contributing longer than expected survival times, inflating the underlying hazard. Conversely, excluding any of the three lighter blue ZCTAs near the hospital system would also exclude large swaths of evacuated individuals (and secondary CVEs) which would make the estimate of the evacuation order effect based less stable and derived from much fewer cases (with less long survival times).

Using a different scale of aggregation (the census tract in Figure 8), the spatial differences in the area near to the hospital system is somewhat preserved, but the effect of removing the western-most areas is largely attenuated. This is likely due to the aggregation effect (aggregating at too small a scale means limited numbers of individuals per tract, and a less clear effect.)

4 Discussion

In this short report we explore the spatial patterning of evacuation effect model residuals from a full sample of Santa Barbara County residents with existing cardiovascular diagnoses. Summed martingale residual values flag several individuals with longer than expected survival times, which we explored spatially. We further investigated the sensitivity of the model fit to spatial patterning across several scales and found some evidence to suggest a problematic relationship with increasing distance and the magnitude of the evacuation order estimate.

The results of this investigation were a key part of our decision to geographically confine our final sample in the our primary paper and have a more conservative un-estimated baseline hazard (1). The statistical spatial patterning of these results has a parallel interpretation for clinical researchers, which is hospital service areas. If patients received care for their initial CVE at Cottage Hospital, and then had further treatments at other hospital systems – they would appear as censored survival times in our data (and potentially be quite long). There are other locations besides our dominant hospital system for immediate cardiovascular care - particularly in the western south central area of the county, and the northwest. These areas did appear in the statistical analysis, though the statistical narrative was less clear than we would have hoped.

Alternatively, it is also possible to directly account for the spatial correlation of the residuals using Bayesian hierarchical models. These models extend frailty models to include a spatial frailty that can be modeled from direct patient locations or aggregated areal counts (*geostatistical* or *lattice/areal* respectively). Essentially, these models extend random effects (frailty terms) to non-$i.i.d.$ settings, where one can model spatial correlation structures using, for example, a powered exponential, or Matèrn (in the geostatistical approach) or a network graph approach (lattice/areal approach). Banerjee, Wall and Carlin (2003) provide excellent statistical theory development and application for infant mortality in Minnesota (4) as do Henderson, Shimakura and Gorst in with leukemia in Northwest England

(2002) (12). In extensions/adaptations, Bastos and Gamerman (2006) describe use dynamic models to consider models without proportional hazards assumption (5); Banarjee and Dey consider semi-parametric approaches (3); Zhang and Lawson assess other spatial parametric approaches (accelerated failure times)(30).

Despite their strengths, in the context of this short report, the Bayesian hierarchical models present some additional challenges – the spatial frailty effects for each area (latticed or otherwise) are assumed to be smoothly spatially varying, and the statistical theory still rests on a random (or completely enumerated) sample within each area. While the first assumption may hold, given the fragmented landscape of healthcare within the United States, the second assumption likely does not.[1] Patients with addresses further from Santa Barbara (the city) have other options for cardiovascular care, and those who do appear in the data may appear for systematic reasons (such as commuting patterns or severity). This likely would bias the inference of estimated hazards across space (and the interpretation of relative areas), but also may reproduce the spatial patterning directly in the hazard rates themselves as opposed to the residuals from the Cox model.

Given the somewhat murky spatial interpretation of the residual martingale values and leave-one-out and leave-one-group-out analyses, we still believe there might be applications of survival analysis (both Bayesian hierarchical and otherwise) with hospital data to describe functional hospital service areas, particularly for diagnoses that require somewhat regular care. The hospital service area literature is rich in floating catchment area methods to define service units (8; 17; 16; 25; 26; 18), but to our knowledge survival methods remain absent. While clearly not ready for prime time, simple Kaplan-Meier estimates at various scales/distances might reveal natural edges for regular care in the healthcare landscape. Similarly, loss-to-follow up trends (censoring) or disease exacerbation analyses could reveal

[1] Additionally, U.S. census tracts, by design, typically are more homogeneous than ZCTAs. In general, they range in population size from 1,200 to 8,000 people with an optimal size of 4,000. In Santa Barbara County, populated census tracts range in size from 1,245 to 9,519 whereas ZCTAs range in size from 154 to 55,126. These differences in population can be seen in Figure 9, and, coupled with our results further exacerbate potential issues with areal units, spatially structured survival data, and reconciling heterogeneity in the healthcare system.

edge distances where patients might benefit from satellite clinics or coordinated care across systems. Though tricky to implement, such analyses could prove a fruitful compliment to the existing floating catchment methods.

5 Conclusion

In observational studies with spatially referenced data, careful model fitting is critical. A thorough check for spatial autocorrelation and/or patterning in the residual values is a necessary step, and multiple scales may reveal different narratives and conclusions. In survival models and especially extended Cox models, these spatial effects may play out in model estimates that are artificially high (or low) depending on the structure of the unestimated baseline hazard. Conservative and meticulous decision making in the construction of the models is vital for interpretation, and Bayesian hierarchical models that directly incorporate spatial structures should be explored. However, despite the difficulties, there are also potential applications in other areas from these methods, and in particular hospital service area research. Survival methods to assess changing survival times between regular care visits could be used to determine functional catchment areas and target existing resources, and would compliment existing floating catchment methods.

References

[1] Arabadjis, S. D., Sweeney, S. H., Kenner, C. E., and Tedesco, D. J. Wildfire, evacuation, and cardiovascular events: A spatial exposure approach. *Applied Geography 159* (2023), 103033.

[2] Banerjee, S. Spatial data analysis. *Annual Review of Public Health 37*, 1 (2016), 47–60.

[3] Banerjee, S., and Dey, D. K. Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis 11*, 2 (2005), 175–191. Publisher: Springer Nature.

[4] Banerjee, S., Wall, M. M., and Carlin, B. P. Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics 4*, 1 (2003), 123–142.

[5] Bastos, L. S., and Gamerman, D. Dynamic survival models with spatial frailty. *Lifetime Data Analysis 12*, 4 (2006), 441–460.

[6] Chan, C., Elliott, J., Troughton, R., Frampton, C., Smyth, D., Crozier, I., and Bridgman, P. Acute myocardial infarction and stress cardiomyopathy following the Christchurch earthquakes. *PloS One 8*, 7 (2013), e68504.

[7] Cohen, O., Shapira, S., and Furman, E. Long-Term Health Impacts of Wildfire Exposure: A Retrospective Study Exploring Hospitalization Dynamics Following the 2016 Wave of Fires in Israel. *International Journal of Environmental Research and Public Health 19*, 9 (2022), 5012.

[8] Delamater, P. L. Spatial accessibility in suboptimally configured health care systems: A modified two-step floating catchment area (m2sfca) metric. *Health & Place 24* (2013), 30–43.

[9] Delfino, R. J., Brummel, S., Wu, J., Stern, H., Ostro, B., Lipsett, M., Winer, A., Street, D. H., Zhang, L., Tjoa, T., and Gillen, D. L. The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003. *Occupational and Environmental Medicine 66*, 3 (2009), 189–197.

[10] Hastie, T., Tibshirani, R., and Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed ed. Springer series in statistics. Springer, 2009.

[11] Heaney, A., Stowell, J. D., Liu, J. C., Basu, R., Marlier, M., and Kinney, P. Impacts of Fine Particulate Matter From Wildfire Smoke on Respiratory and Cardiovascular Health in California. *GeoHealth 6*, 6 (2022).

[12] Henderson, R., Shimakura, S., and Gorst, D. Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association 97*, 460 (2002), 965–972. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214502388618753.

[13] Henderson, S. B., Brauer, M., MacNab, Y. C., and Kennedy, S. M. Three Measures of Forest Fire Smoke Exposure and Their Associations with Respiratory and Cardiovascular Health Outcomes in a Population-Based Cohort. *Environmental Health Perspectives 119*, 9 (2011), 1266–1271.

[14] Lawson, A. B., Choi, J., and Zhang, J. Prior choice in discrete latent modeling of spatially referenced cancer survival. *Statistical Methods in Medical Research 23*, 2 (2014), 183–200. Publisher: SAGE Publications Ltd STM.

[15] Lim, Y.-H., Hong, Y.-C., and Kim, H. Effects of diurnal temperature range on cardiovascular and respiratory hospital admissions in Korea. *Science of The Total Environment 417-418* (2012), 55–60.

[16] Luo, W., and Qi, Y. An enhanced two-step floating catchment area (e2sfca) method for measuring spatial accessibility to primary care physicians. *Health & Place 15*, 4 (2009), 1100–1107.

[17] Luo, W., and Whippo, T. Variable catchment sizes for the two-step floating catchment area (2sfca) method. *Health & Place 18*, 4 (2012), 789–795.

[18] McGrail, M. R., and Humphreys, J. S. Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method. *Applied Geography 29*, 4 (2009), 533–541.

[19] Moore, D. F. *Applied Survival Analysis Using R*, 1st ed. 2016 ed. Use R! Springer International Publishing : Imprint: Springer, Cham, 2016.

[20] Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., and Elliott, C. T. Critical Review of Health Impacts of Wildfire Smoke Exposure. *Environmental Health Perspectives 124*, 9 (2016), 1334–1343.

[21] Reid, C. E., Jerrett, M., Tager, I. B., Petersen, M. L., Mann, J. K., and Balmes, J. R. Differential respiratory health effects from the 2008 northern California wildfires: A spatiotemporal approach. *Environmental Research 150* (2016), 227–235.

[22] Swerdel, J. N., Janevic, T. M., Cosgrove, N. M., Kostis, J. B., and Myocardial Infarction Data Acquisition System (MIDAS 24) Study Group. The effect of Hurricane Sandy on cardiovascular events in New Jersey. *Journal of the American Heart Association 3*, 6 (2014), e001354.

[23] Therneau, T. M. *A Package for Survival Analysis in R*, 2023. R package version 3.5-3.

[24] Therneau, T. M., and Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000.

[25] Wan, N., Zou, B., and Sternberg, T. A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science 26*, 6 (2012), 1073–1089.

[26] Wang, F. Measurement, optimization, and impact of health care accessibility: A methodological review. *Annals of the Association of American Geographers 102*, 5 (2012), 1104–1112.

[27] Wellenius, G. A., Schwartz, J., and Mittleman, M. A. Air pollution and hospital admissions for ischemic and hemorrhagic stroke among medicare beneficiaries. *Stroke 36*, 12 (2005), 2549–2553.

[28] Wen, B., Wu, Y., Xu, R., Guo, Y., and Li, S. Excess emergency department visits for cardiovascular and respiratory diseases during the 2019-20 bushfire period in Australia: A two-stage interrupted time-series analysis. *The Science of the Total Environment 809* (2022), 152226.

[29] Yang, J., Zhou, M., Ou, C.-Q., Yin, P., Li, M., Tong, S., Gasparrini, A., Liu, X., Li, J., Cao, L., Wu, H., and Liu, Q. Seasonal variations of temperature-related mortality burden from cardiovascular disease and myocardial infarction in China. *Environmental Pollution 224* (2017), 400–406.

[30] Zhang, J., and Lawson, A. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics 38*, 3 (2011), 591–603. Publisher: Routledge.

6 Tables and Figures

	Dependent Variable:
	Risk of Secondary CVE
Evacuation Order Exposure	0.2494*
	(0.1486)
	Robust SE
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 1: Evacuation and Full County Population Reduced Model

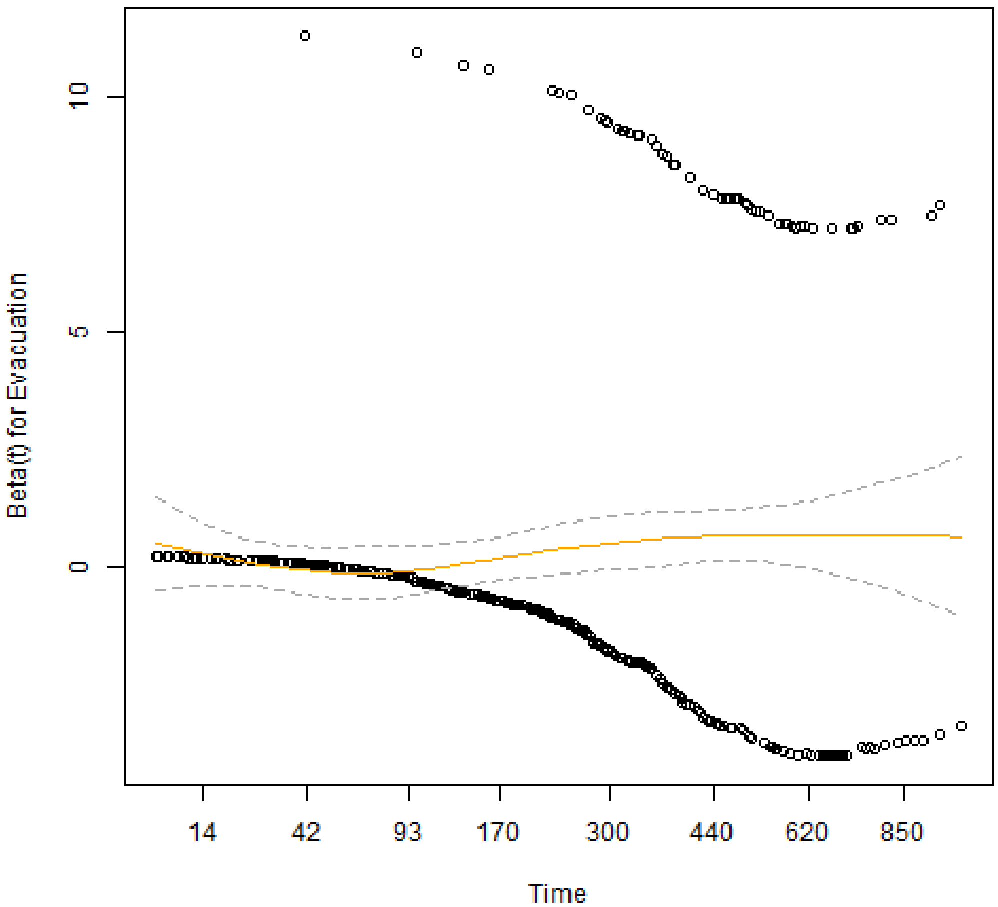


Figure 1: The Schoenfeld residuals for the evacuation order parameter estimate by time for the simple model with the full sample.

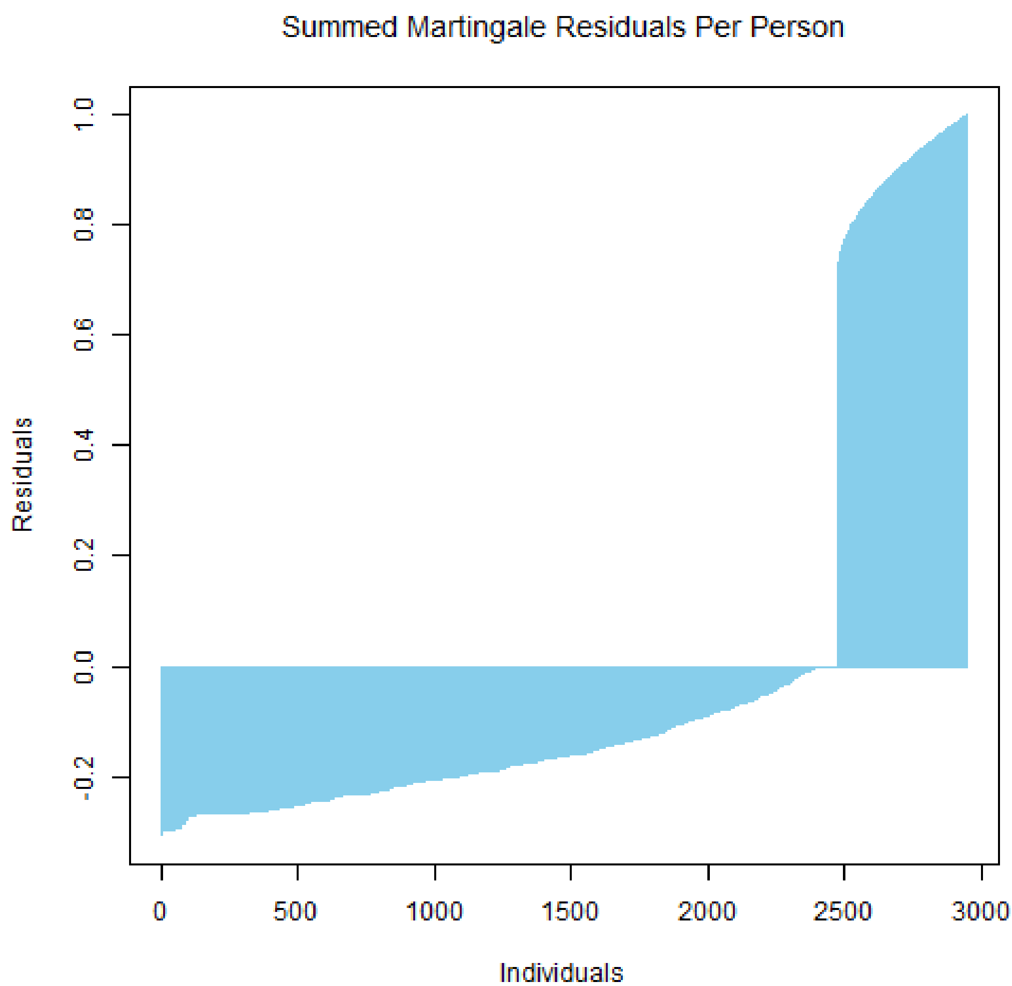


Figure 2: The summed martingale residual values by individual, sorted by size.

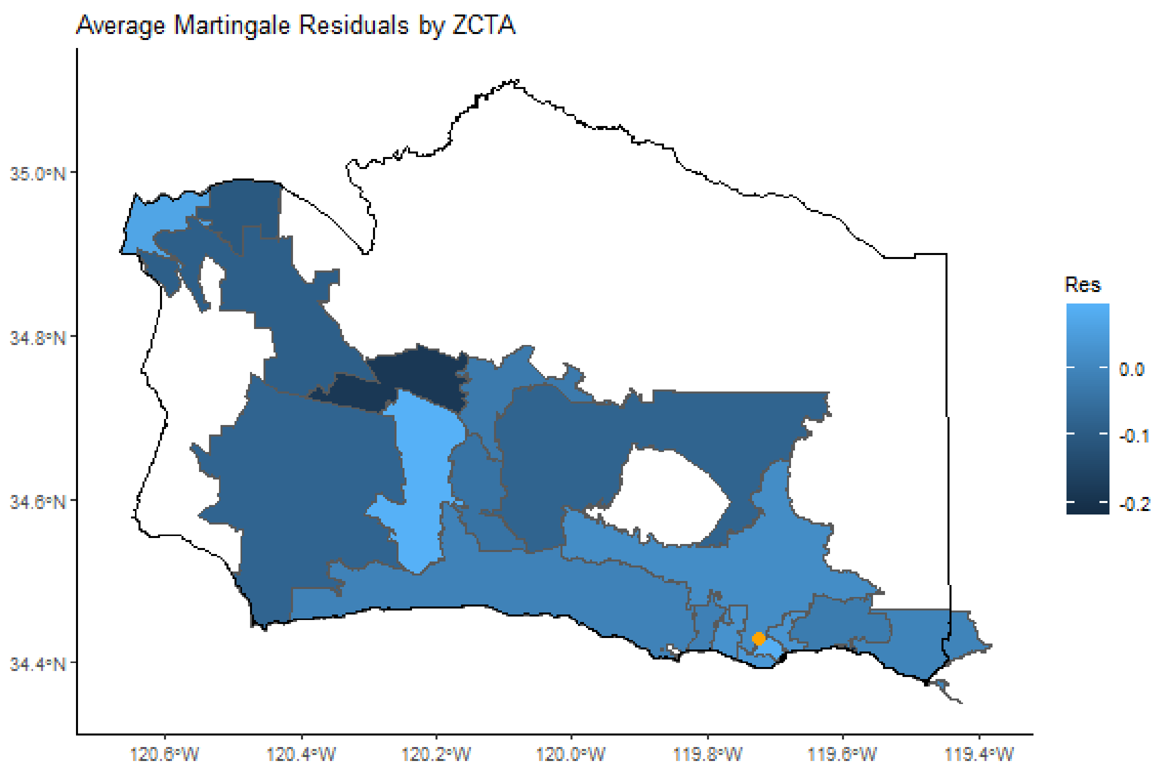


Figure 3: Map of the summed martingale residual values per individual, associated to one address and aggregated by ZCTA.

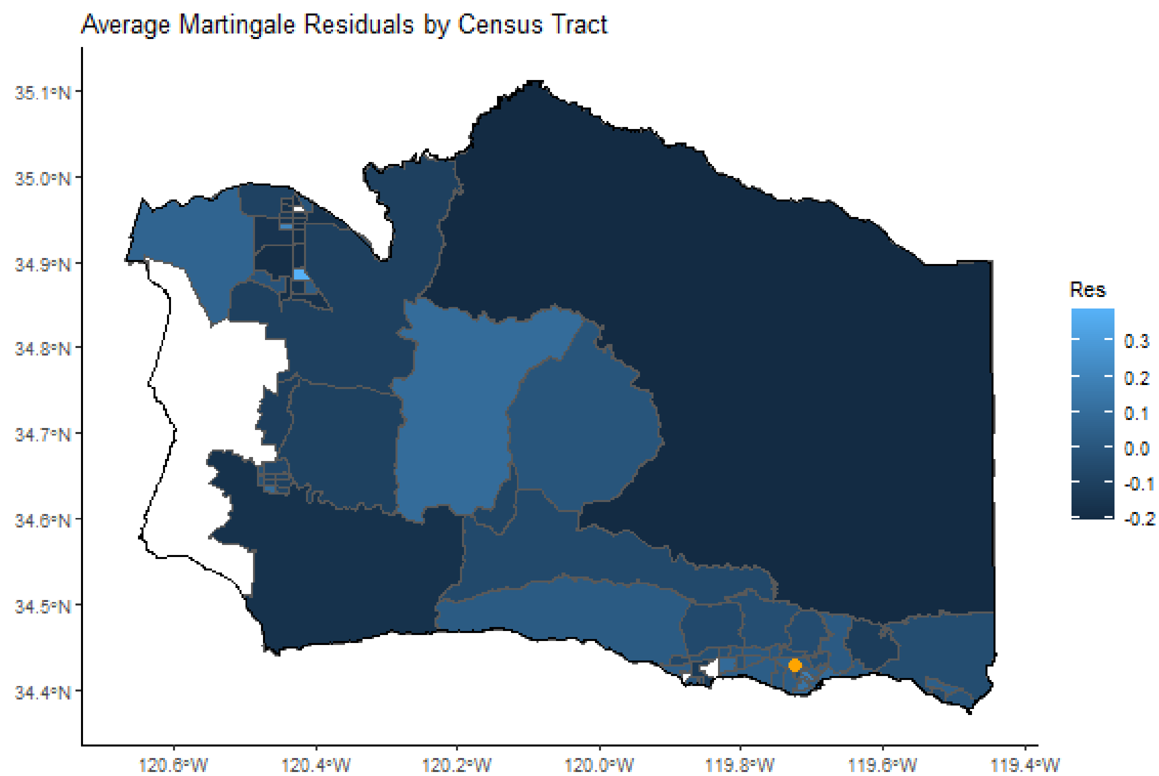
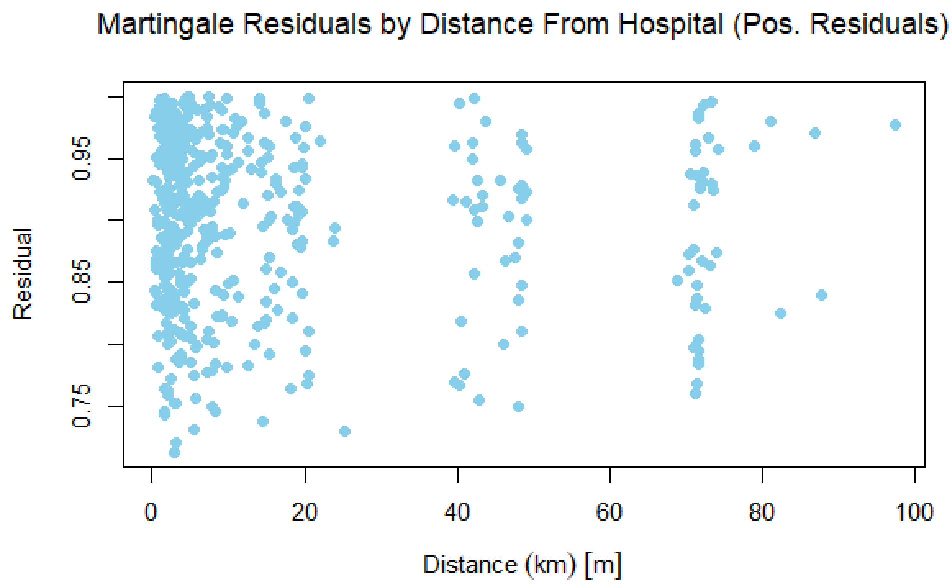
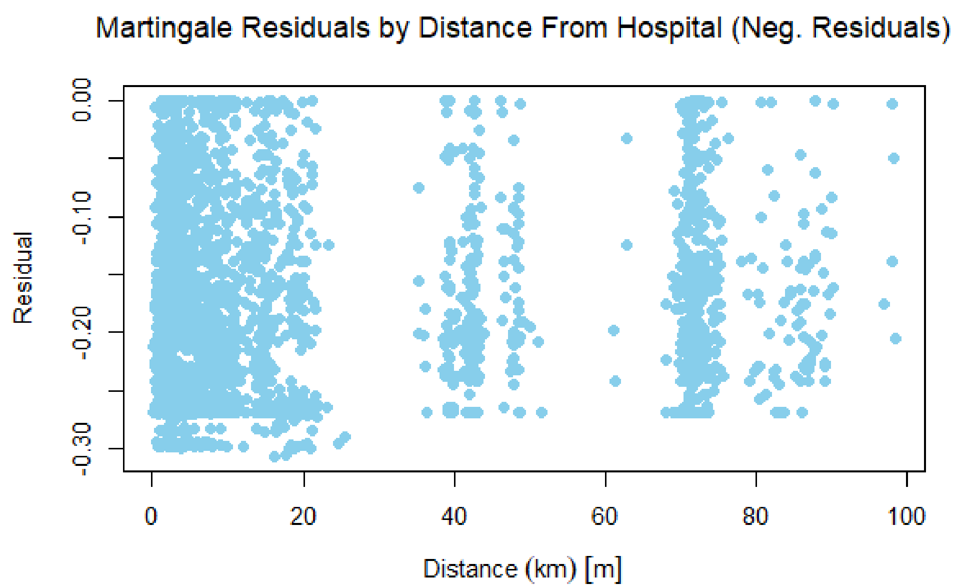


Figure 4: Map of the summed martingale residual values per individual, associated to one address and aggregated by census tract.



(a) Positive Summed Residuals



(b) Negative Summed Residuals

Figure 5: Summed martingale residual values per individual separated by positive (longer than expected survival time) and negative (shorter than expected survival time), and plotted by Euclidean distance from the hospital system location.

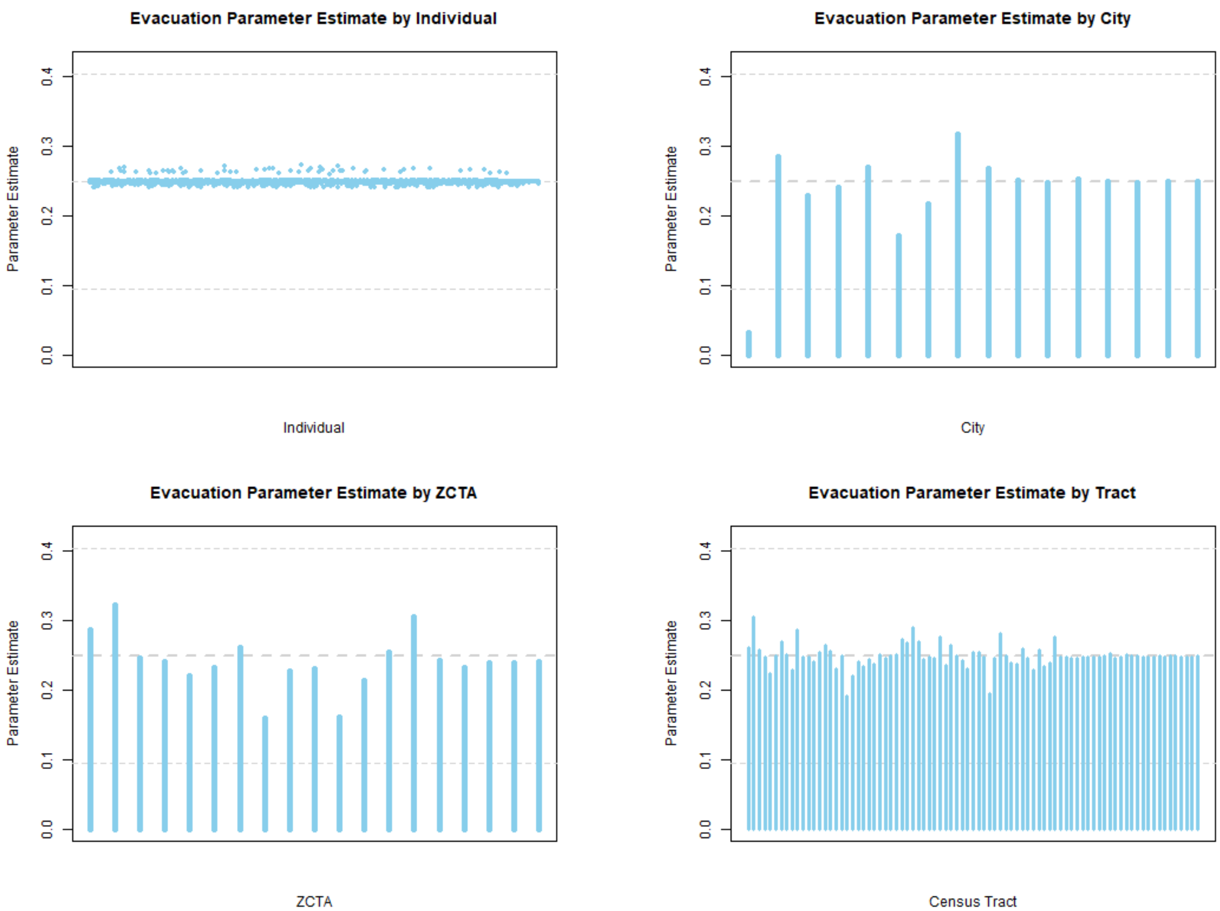


Figure 6: Jackknife residuals from leave-one-out analyses at various scales: individuals, city names, ZCTA, census tract.

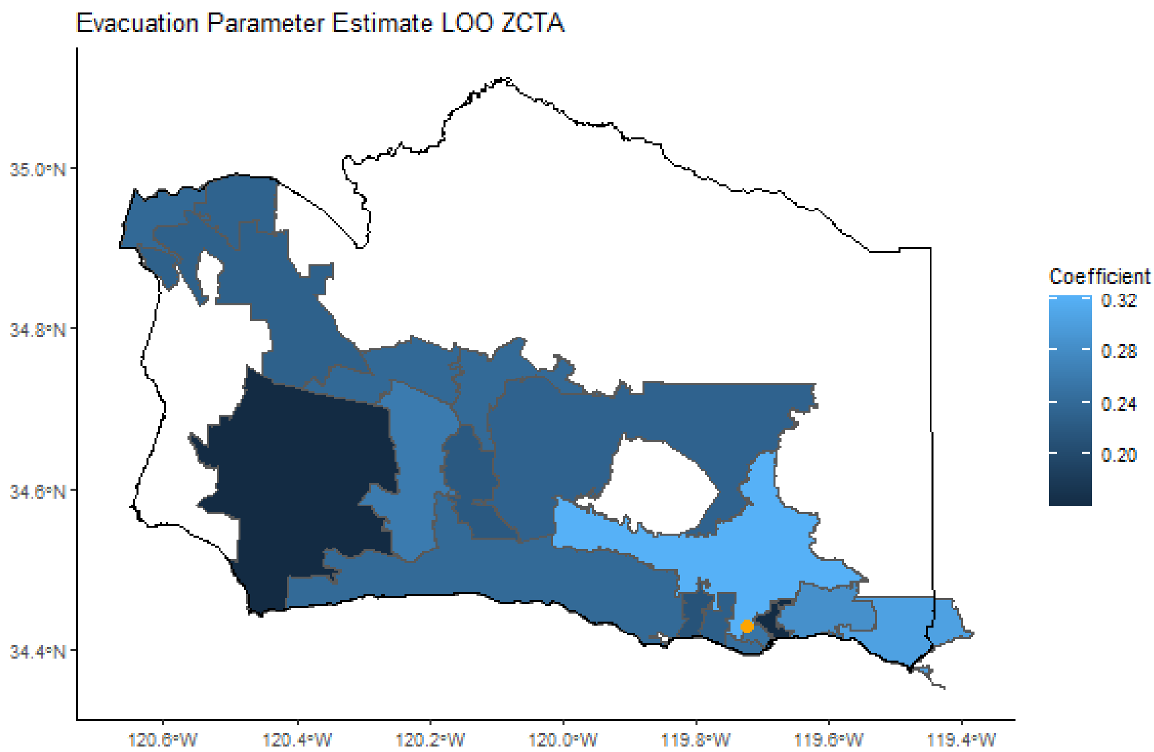


Figure 7: Mapped jackknife residuals for leave-one-out ZCTA and coefficient estimates.

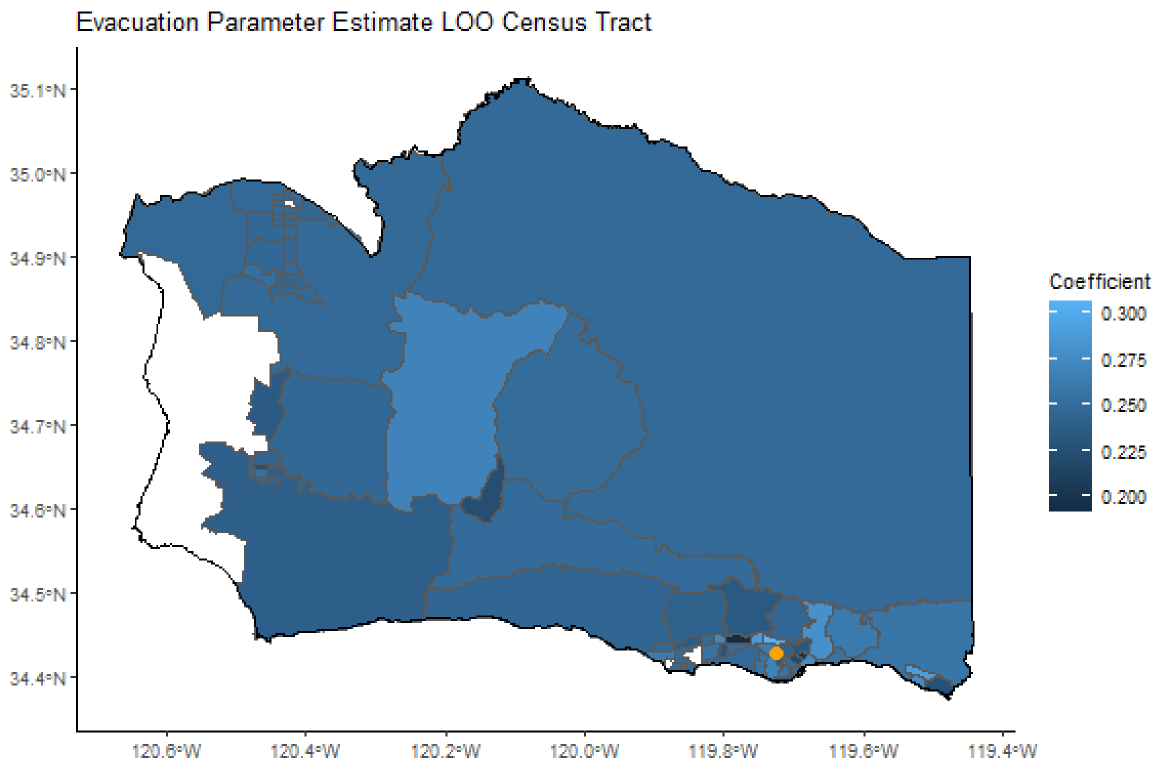
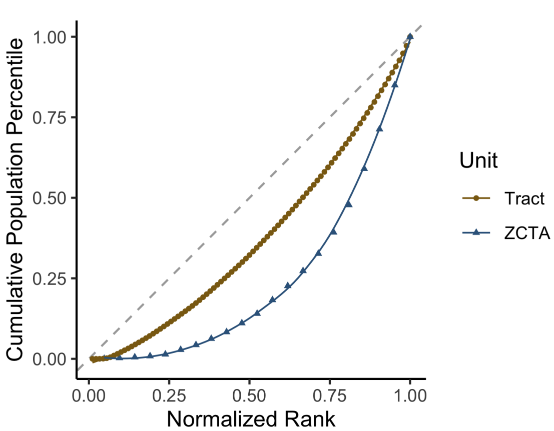
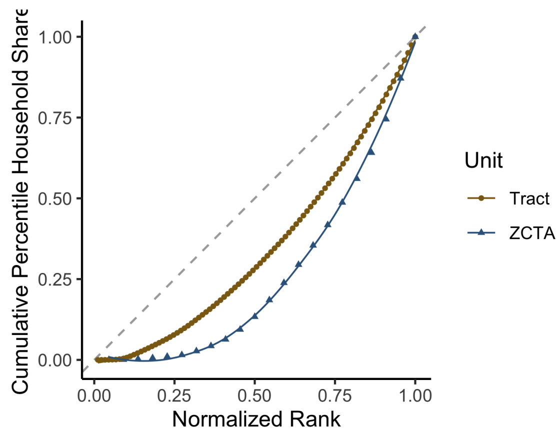


Figure 8: Mapped jackknife residuals for leave-one-out census tract and coefficient estimates.



(a) Total population



(b) Households with at least one member aged 65 years or older

Figure 9: The Lorenz curves display the cumulative percentiles and normalized ranks of both U.S. census tracts and zip code tabulation areas (ZCTAs) in Santa Barbara County. In each case, the curves are compared to a 45°-line which would indicate perfect alignment (e.g. 25% of the population would have a normalized rank of 0.25.) The further the curve from the 45°-line, the less the measure (variable) is distributed homogeneously (or equally) across areal units. For both the total population (panel A) and the share of households with at least one member aged 65 years or older (panel B) the census tracts show more homogeneity in construction than the ZCTAs. As age is a particular risk factor for cardiovascular events, and we can see from the panel B that while neither ZCTAs nor census tracts are perfectly aligned, the census tracts are better than the ZCTAs. Specifically, while 50% of the census tracts account for approximately 25% of the share of households with at least one member over the age of 65, nearly 70% of the ZCTAs account for only 25% of the share of households with at least one member over the age of 65.