**Title 1: Combining straight-line and map-based distances to quantify the impact of neighborhood-level food access on health**

**Title 2: A hybrid distance-based approach to quantify the impact of neighborhood-level food access on health**

By Sarah C. Lotspeich1, Lucy D’Agostino McGowan1, and Julie Velasquez Runk2

1Department of Statistical Sciences, Wake Forest University, Winston-Salem, North Carolina

2Department of XX Wake Forest University, Winston-Salem, North Carolina

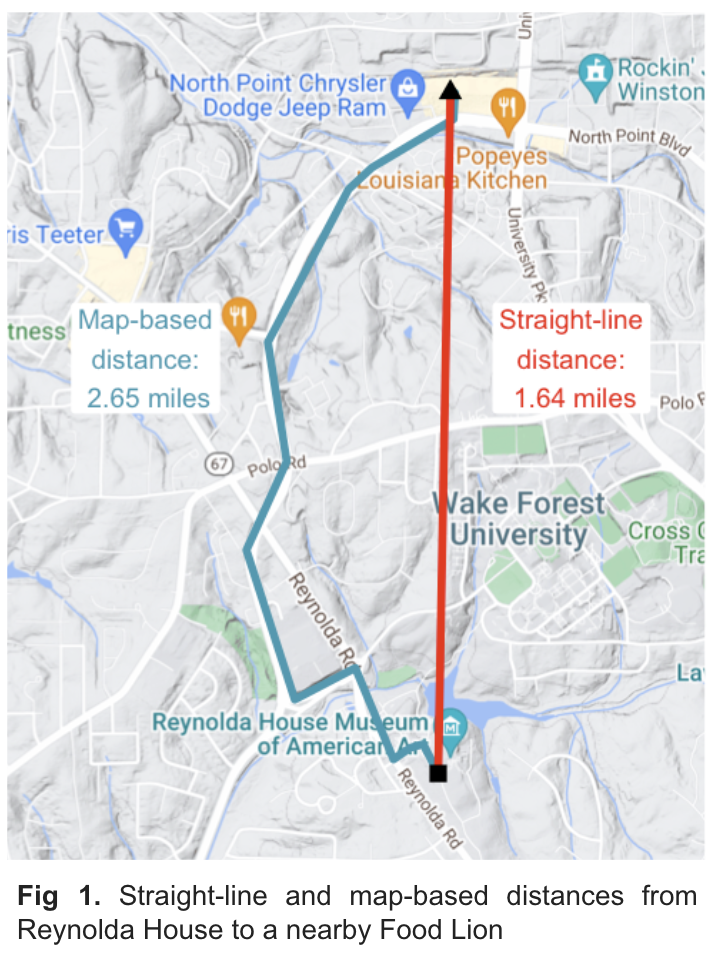
**Abstract**

Healthy foods are essential for a healthy life, but accessing healthy food can be more challenging for some people than others. With this disparity in food access comes disparities in well-being, leading to disproportionate rates of diseases in communities that face more challenges in accessing healthy food (i.e., low-access communities). Identifying low-access, high-risk communities for targeted interventions is a public health priority, but current methods to quantify food access rely on distance measures that are either computationally simple (the length of the straight-line route) or accurate (the length of the map-based route), but not both. Using data for Forsyth County, North Carolina, we first quantify and compare the associations between various health outcomes (coronary heart disease, diabetes, high blood pressure, and obesity) and neighborhood-level food access based on (i) straight-line distance versus (ii) map-based distance for all neighborhoods. Then, we propose a hybrid statistical approach to combine these methods, allowing researchers to harness the computational ease of one with the accuracy of the other. This model incorporates straight-line distances for all neighborhoods and map-based distances for just a subset, offering comparable estimates to the “gold standard” model using map-based distances for all neighborhoods.

Abstract word count: 186/XXX

**1. Background and Motivation**

Healthy eating is important to good health, yet not accessible to everyone.Eating healthy foods is critical to development in childhood and prevention of illnesses in adulthood, including cardiovascular disease,diabetes, and obesity.1-3 Not just preference for healthy foods, but access, goes into what people eat. For some people, predominantly those in low-income or minority communities and with disabilities, healthy foods are not always accessible.4-5 Access can be hampered by physical factors, like geographic proximity to stores and the availability of public transit,6 and social factors, like structural racism and discrimination.7-8 Thus, not just consumption of healthy foods but access to them can impact health,9-12 and disparities in access perpetuate disparities in health. For public health officials seeking to reduce the burden of a disease, understanding its connection to food access and the landscape of food access in a community can be informative. Quantifying food access (e.g., how far people travel to buy healthy foods) is an important place to start.

To identify low-access, high-risk neighborhoods, food access is often quantified using the distance to grocery stores (i.e., how far they are away). When calculating distance there currently exists a trade-off between (i) computationally simple methods that are less accurate and (ii) more accurate methods that are computationally complex. Computationally simple options draw a straight line between the two points “as the crow flies.” However, this process ignores infrastructure like roads and topography like rivers, which underestimates true distances to grocery stores and overestimates food access (**Fig. 1**). Map-based distances incorporate real-world obstacles, like roads and rivers, but free options like Google Maps, or even paid options like ArcGIS, can quickly become time-intensive. ~~In Forsyth County, North Carolina (NC) there are 250 neighborhoods and 855 grocery stores; this would require 250 x 855 = 213,750 calculations to get the distances from every neighborhood to every store!~~ 

TO COMBINE: Current metrics of how close neighborhoods are to healthy foods have limitations.A neighborhood’s access to healthy food can be quantified by the number of grocery stores within some proximity, but this metric relies on calculating the distances between the neighborhood and possible stores. Current distance calculations are either computationally simple but less accurate (e.g., they draw a straight line through the shortest path) or more accurate but computationally complex (e.g., they draw a path following roadways). Computationally simple methods can underestimate actual distance to the grocery store by ignoring lack of roadways or natural obstacles like rivers. The United States Department of Agriculture (USDA) uses a more sophisticated grid-based calculation,13 but this method also ignores many realistic obstacles. These problems are likely exacerbated in rural areas, where there are generally fewer stores and fewer available straight-line routes.

Preliminary data using straight-line distance suggest that many neighborhoods in Forsyth County do not have many food options (**Fig. 2**), and that’s the best-case scenario.

This concern is echoed by a 2019 report from the USDA, which found that 68%–87% of neighborhoods in Forsyth County, NC had low access to food. There are 95 census tracts in Forsyth County, NC. The USDA identified 63 (68%) and 81 (87%) of them as having low access to food, due to the proportion of people living more than 1 mile and 1/2 mile from the nearest supermarket, respectively. These high proportions of low-access tracts put Forsyth among the hardest-hit counties in NC.20 Additional work is needed to determine if neighborhood variability in health (**Fig. 3**) can be explained by this variability in food access (**Fig. 2**).

Using map-based distance to quantify these neighborhoods’ food access would be more realistic.6 The Google Maps API is an incredibly powerful for calculations like this, and software can integrate the API into a statistical workflow.14-15 These tools can be used to calculate the map-based travel distance between two locations, offering a more accurate snapshot of a neighborhood’s access. Still, making all of the necessary calculations (i.e., the distances between all neighborhoods and grocery stores) would not be feasible due to time-intensive computations and monthly limits on API usage. With map-based distances for only some neighborhoods, a challenge emerges: what about neighborhoods with only straight-line distances available? How can they be included in analyses?

An ideal solution would offer the accuracy of the map-based distances for all neighborhoods without the added computational complexity. Unfortunately, on a fixed budget and barring software developments from Google, this solution is not currently feasible. However, a hybrid one is possible: to obtain map-based distances for a subset of neighborhoods and use straight-line distances for the rest. Then, treating the straight-line distances as error-prone versions of the map-based ones we create a manageable statistical problem instead of an unmanageable computational one.

Measuring access to healthy food is a measurement error problem. The straight-line and map-based methods seek to measure the same thing: distance between neighborhoods and grocery stores. Assuming that straight-line distance underestimates the more realistic map-based one, a measurement error model is a natural way to relate them.

REVISE: Consider estimating the neighborhood-level rate of coronary heart disease as a function of the number of grocery stores within 1 mile (food access). Then Assume that is the number of stores with map-based distances less than 1 mile to the neighborhood, while \*, the number of stores with straight-line distances less than 1 mile to the neighborhood, is an error-prone version of . Since straight-line distances are computationally simple, \* is available for all neighborhoods, but due to computational limits is only available for some.

This two-phase study design is commonly used in measurement error work,16-17 and many statistical methods can be used to analyze the resulting data with no missing \* but some missing . Replacing missing values with predictions based on \* (i.e., imputation) is a promising option,18 because (i) it offers nice statistical precision and (ii) after replacing the missing s we can fit standard statistical models.19

**Aim 1:** Accurately quantify food access using map-based distance for all neighborhoods and estimate the impact of food access on rates of adverse health outcomes.

**Aim 2:** Illustrate how badly error-prone straight-line distances can bias estimates of the impact of food access.

**Aim 3:** Combine food access calculated using straight-line distance for all neighborhoods and map-based distance for a subset of neighborhoods to obtain accurate estimates at reduced computational cost.

Upon completion, we will have demonstrated that the accuracy of the map-based distance calculations can be captured without querying all neighborhoods through a measurement error framework. This project will provide preliminary results for a future NIH R21 to develop new statistical methods to design and analyze food access and health disparities studies across NC.

This project will propose a new statistical approach to accurately quantify food access and model its impact on health with less computational strain. Through the adoption of a measurement error framework, the proposed approach offers improved accuracy with computational ease.

**2. Methods and Data**

*2.1 Data Collection*

To assess the impact of food access on health in Forsyth County, NC, publicly available data will be combined from three sources using open-source tools in the statistical computing language R. All R code for data collection and processing are available on GitHub at <https://github.com/sarahlotspeich/Forsyth-Food-Access/>.

First, neighborhood-level prevalence estimates of health outcomeswere available from the Centers for Disease Control and Prevention in the 2022 PLACES dataset.23 [More description of how PLACES data are collected.] Based on prior studies, coronary heart disease, diabetes, high blood pressure, and obesity will be considered.9,11 A neighborhood’s prevalence can be interpreted as the rate of the event across the population (e.g., the proportion of a neighborhood’s population that has been diagnosed with coronary heart disease). These data were available at the census tract level (one level above the census block groups), which will be adopted as the “neighborhood” unit for analysis. Per the US Census, census tracts are “small, relatively permanent statistical subdivisions of a county.” The average population per census tract nationally is approximately 4000 people, with minimum and maximum populations of 1200 and 8000 people, respectively.REF

Second, as neighborhoods**,** the 95 census tracts (CTs) in Forsyth County, NC were used.22 The latitude and longitude centroid coordinates for the neighborhoods’ centroids (i.e., their geographic centers) were used as the origins for all distance calculations. These coordinates were obtained in the following way.

1. Shapefiles for the Forsyth County census tracts were loaded using the *tidycensus* package.REF
2. Latitude and longitude coordinates for the centroids of the shapefiles were calculated using the *sd* package.REF However, there is no guarantee that these “geometric” centroid coordinates correspond to real, accessible places in space (e.g., they could fall in the middle of a lake).
3. Coordinates from Step 2 were converted to the nearest street addresses (i.e., reverse-geocoded) with a rooftop-level match using the *ggmap* package.REF
4. Street addresses from Step 3 were converted back to latitude and longitude coordinates (i.e., geocoded) for “map-aligned” centroids using the *ggmap* package.REF

The geometric and map-aligned centroids from Steps 2 and 4, respectively, tended to be very close together (Supplemental Figure #). The median straight-line distance between them was 0.02 miles (interquartile range [IQR] 0.01–0.08 miles).

Third, grocery store data were collected from Forsyth and its bordering counties (Davidson, Davie, Guilford, Randolph, Rockingham, Stokes, Surry, and Yadkin) in the following way. The *googleway* packageREF can be used to search Google Maps for search phrase (“Grocery Stores”) within some radius (1 kilometer) of a specified location (the centroid of a census block group). Each search returns the top 20 results for the parameters. Thus, searches around the 1023 census block groups (CBGs)–rather than census the 95 census tracts–were used here to collect the grocery store data to achieve better coverage of Forsyth and its bordering counties. Compiling the top 20 search results from all CBGs led to an initial sample of 855 unique results for “Grocery Stores” in Forsyth and its bordering counties.

Then, each grocery store result was manually reviewed in Google Maps to classify it as one of the following: (i) major chains (e.g., Food Lion and Lowe’s Foods), (ii) local stores (e.g., Asian and Indian markets), (ii) specialty stores (e.g., produce stands and butcher shops), (iv) dollar stores (e.g., Family Dollar and Dollar Tree), (v) convenience stores, or (vi) other (e.g., grocery distributors and community pantries). Results in the “other” category were excluded to focus on more traditional store options, and the remaining results were grouped into healthy foods stores (major grocery chains, local grocery stores, or specialty stores) and unhealthy foods stores (dollar stores or convenience stores). This exclusion left a total of 774 validated grocery stores (425 healthy and 349 unhealthy); a more detailed breakdown of the validated grocery stores by type can be found in **Table XX**. Finally, for distance calculations the *ggmap* package was used to geocode grocery store addresses to latitude and longitude coordinates.14

| **Grocery Store Type** |  |
| --- | --- |
| *Healthy Foods Stores* | *425 (55%)* |
| Major Chains | 247 (32%) |
| Local Stores | 139 (18%) |
| Specialty Stores | 39 (5%) |
| *Unhealthy Foods Stores* | *349 (45%)* |
| Dollar Stores | 92 (12%) |
| Convenience Stores | 257 (33%) |

**Table XX.** Frequency of grocery store types in the processed dataset.

*2.2 Distance Calculations*

Two distance calculations are considered to quantify each neighborhood’s food access: (i) straight-line distances and (ii) map-based distances. Straight-line distances are calculated using the Haversine Formula, which measures the shortest distance between two sets of latitude and longitude coordinates (adjusting for the slight curvature of the earth). The Haversine Formula is implemented in the *geosphere* packageREF and can be used to quickly compute the distances between all neighborhoods and all grocery stores. Map-based distances are calculated using the *ggmap* package to query the driving route between two sets of latitude and longitude coordinates (i.e., between a neighborhood centroid and grocery store) from the Google Maps API.14

*2.3 Model and Notation*

For the health outcomes, let denote the number of events (e.g., the number of people diagnosed with coronary heart disease) and denote the population in a neighborhood, such that is the “rate” of that outcome in the neighborhood. To measure food access, let be the number of grocery stores within 1 mile of the neighborhood. The data collected include neighborhoods and grocery stores. Assume that true food access, , can only be obtained using the map-based distance calculations, while the straight-line distance calculations yield an error-prone measure of food access, \*.

Due to computational strain, is only observed for m neighborhoods (), while \* is observed for all. This setup creates three analytical datasets: (i) unqueried (i.e., ), with missing for all neighborhoods, (ii) fully-queried (i.e., ), with measured for all neighborhoods, and (iii) partially-queried (i.e., ), with measured for half of the neighborhoods and missing for the rest (**Fig. 2**). 

Constructing the unqueried dataset is computationally simple, as the *geosphere* package can be used to calculate all 250 x 775 = 193,750 calculations in 0.055 seconds on a 2020 Macbook Pro24 ~~but using this error-prone \* in place of in our analysis will lead to incorrect results.~~~~25~~ Still, constructing the fully-queried dataset with for all neighborhoods is computationally expensive. Calculating the map-based distance between one neighborhood and one grocery store takes longer than calculating all of the straight-line distances for all neighborhoods and stores (e.g., 0.42 seconds vs. 0.055 seconds). Thus, creating the fully-queried dataset is a key challenge for this project, as resources are needed to overcome the barriers of the map-based calculations (e.g., slow computation and monthly query limits). However, these data are essential to tackling the aims of the project.

*2.4 Selecting Neighborhoods to Undergo Map-Based Distance Calculations*

The partially-queried dataset will include straight-line food access \* from the unqueried dataset for all neighborhoods and a map-based food access (from the fully-queried dataset) for a subset of neighborhoods. To ensure a wider geographic spread, the m = 125 to-be-queried neighborhoods were chosen using a stratified random sample on the <CENSUS TRACT?>

*2.5 Primary Analysis: Model of Health Outcomes and Food Access*

Poisson regression will be used to assess the impact of food access on the neighborhood-level rate of various adverse health outcomes. Inference and conclusions were based on estimates of the model . Specifically, the prevalence ratio for food access was of primary interest, where a would indicate that greater food access led to greater rates of the health outcomes, would indicate that greater food access had no impact on the rates of the health outcomes, and would indicate that greater food access led to lower rates of the health outcomes. The hypothesis that access to more healthy foods

For comparison, this model of health outcomes and food access was fit to each of the three datasets. In the partially-queried data, some s are missing, and in the unqueried data, all s are. To overcome the missingness, the models fit to these datasets were modified in the following ways.

1. As the gold standard analysis, the fully-queried data were used to fit the model using , , and from all neighborhoods.
2. As the naive analysis, the unqueried data were used to fit a the model using , , and \* (in place of ) from all neighborhoods.
3. As the complete-case analysis, the partially-queried data were used to fit the model using using , , and from only the subset of queried neighborhoods.
4. As the hybrid analysis, the partially-queried data were used to fit the model using , , and imputed (in place of ) from all neighborhoods. For queried neighborhoods, let . For unqueried neighborhoods, food access was imputed as \*, with and estimated using an “imputation model.” The imputation model was a linear regression of \* fit to the subset of neighborhoods with both and \* measured. Multiple imputed datasets were created, the same model fit to each, and then these models pooled to obtain final results in a multiple imputation framework.18

These models

Still, these fully-queried data are critical to show (i) the extent to which food access impacts health (full cohort analysis), (ii) how badly error-prone measures to food access can bias the impacts on health (naive analysis), and (iii) how well the proposed approach captures the true impacts of food access on health while reducing computation (hybrid analysis).

*2.6 Secondary Analyses: Sensitivity to Unhealthy Foods Store and Food Access Threshold*

To investigate sensitivity of the primary analyses to additional assumptions, two secondary analyses were included. First, food access was re-calculated excluding the unhealthy food stores, and model results compared to those from the primary analysis that included these stores. Second, food access based on the number of grocery stores located within 1/2, 3, and 5 miles will also be considered, and model results compared to those from the primary analysis that measured food access within 1 mile.

**~~C3. Alternative Strategies~~** ~~If the imputation model for is inadequate, we will consider more complex options. For example, the model can be more flexible (e.g., using splines to allow a nonlinear relationship between and \*) or incorporate more information (e.g., a neighborhood indicator of rural/urban status). Depending on the performance of the imputed analysis, the size of the partially queried dataset could be modified (e.g., if it performs well with m = 125, how well would it perform with a smaller m = 100 or 75?).~~

3. Results

*3.1 Comparing Food Access by Distance Measures*

*3.2*

4. Conclusions

Healthy food is a critical determinant of healthy living, and yet many communities suffer from inadequate access to fresh, nutritious food.

**Broader impacts**: By overcoming the computational hurdles, the proposed methods will make large scale collection of accurate distance-based measures feasible. Therefore, the geographic scope of studies of access (to food, medical care, etc.) can expand to answer questions and drive decision-making for larger communities. All data collected and code written will also be made publicly available to propel further research.

Future directions: Building upon preliminary data from Forsyth County, we will develop new statistical methods for food access and health disparity studies on statewide scale in a future NIH R21. We will build on our prior measurement error work16-17 to develop a more robust analytical approach that makes fewer assumptions about the relationship between straight-line and map-based distances for broader generalizability. We will derive new optimal designs to decide the most informative neighborhoods to query using Google Maps to achieve the best statistical precision, modifying our prior work to tackle a new data challenge.27-28 With map-based travel time as an another metric of food access that is expensive to obtain,33 we will consider imputing travel time from straight-line distance. Additionally, the computational gains from the proposed methods can be used to more easily expand the study area and quantify the impact of food access on health beyond our community here in Forsyth County to the entire state of North Carolina.

**References**

[1] Liu, S., Manson JE, Lee IM, Cole, SR, Hennekens CH, Willett WC and Buring, JE. (2000). Fruit and Vegetable Intake and Risk of Cardiovascular Disease: the Women's Health Study. *The American Journal of Clinical Nutrition*, 72(4), 922–928.

[2] Harding AH, Wareham NJ, Bingham SA, Khaw K, Luben R, Welch A and Forouhi NG. (2008). Plasma Vitamin C Level, Fruit and Vegetable Consumption, and the Risk of New-Onset Type 2 Diabetes Mellitus: the European Prospective Investigation of Cancer–Norfolk Prospective Study. *Archives of Internal Medicine*, 168(14), 1493–1499.

[3] California Center for Public Health Advocacy, PolicyLink, and the UCLA Center for Health Policy Research. (2008) *Designed for Disease: The Link Between Local Food Environments and Obesity and Diabetes.*

[4] Brucker DL and Coleman-Jensen A. (2017) Food Insecurity Across the Adult Lifespan for Persons with Disabilities, *Journal of Disability Policy Studies*, 28(2): 109–118.

[5] Bower KM, Thorpe RJ, Rohde C, Gaskin DJ. (2014) The Intersection of Neighborhood Racial Segregation, Poverty, and Urbanicity and Its Impact on Food Store Availability in the United States, *Preventive Medicine*, 58, 33–39.

[6] Bennion N, Redelfs AH, Spruance L, Benally S and Sloan-Aagard C. (2022) Driving Distance and Food Accessibility: a Geospatial Analysis of the Food Environment in the Navajo Nation and Border Towns, *Frontiers in Nutrition*, 9: 904119.

[7] Burke MP, Jones SJ, Frongillo EA, Fram MS, Blake CE and Freedman DA. (2018). Severity of Household Food Insecurity and Lifetime Racial Discrimination Among African-American Households in South Carolina. *Ethnicity & Health*, *23*(3), 276–292.

[8] Odoms-Young A. and Bruce MA. (2018). Examining the Impact of Structural Racism on Food Insecurity: Implications for Addressing Racial/Ethnic Disparities. *Family & community health*, *41 Suppl 2 Suppl, Food Insecurity and Obesity*(Suppl 2 FOOD INSECURITY AND OBESITY), S3–S6.

[9] Ahern M., Brown C. and Dukas S. (2011), A National Study of the Association Between Food Environments and County-Level Health Outcomes. *The Journal of Rural Health*, 27: 367-379.

[10] Gallegos D, Eivers A, Sondergeld P and Pattinson C. (2021). Food Insecurity and Child Development: A State-of-the-Art Review. *International Journal of Environmental Research and Public Health*, 18(17), 8990.

[11] Kanchi R, Lopez P, Rummo PE, Lee DC, Adhikari S, Schwartz MD, Avramovic S, Siegel KR, Rolka DB, Imperatore G, Elbel B and Thorpe LE. (2021). Longitudinal Analysis of Neighborhood Food Environment and Diabetes Risk in the Veterans Administration Diabetes Risk Cohort. *JAMA Network Open*, 4(10), e2130789.

[12] Li Y, Wang S, Cao G, Li D and Ng BP. (2021). Disentangling Racial/Ethnic and Income Disparities of Food Retail Environments: Impacts on Adult Obesity Prevalence. *Applied Geography*, 137, 102607.

[13] Rhone A, Ver Ploeg M, Williams R and Breneman V. Understanding Low-Income and Low-Access Census Tracts Across the Nation: Subnational and Subpopulation Estimates of Access to Healthy Food, EIB-209, U.S. Department of Agriculture, Economic Research Service, May 2019

[14] Kahle D and Wickham H. (2013) ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144–161.

[15] Cooley D (2023). *googleway: Accesses Google Maps APIs to Retrieve Data and Plot Maps*. R package version 2.7.7, https://CRAN.R-project.org/package=googleway.

[16] Tao R, **Lotspeich SC**, Amorim G, Shaw PA and Shepherd BE. (2021) Efficient Semiparametric Inference for Two-Phase Studies with Outcome and Covariate Measurement Errors. *Statistics in Medicine*. 40(3): 725–738.

[17] **Lotspeich SC**, Shepherd BE, Amorim GC, Shaw PA and Tao R. (2022) Efficient Odds Ratio Estimation under Two-Phase Sampling Using Error-Prone Data from a Multi-National HIV Research Cohort. *Biometrics*. 78(4): 1674–1685.

[18] Cole SR, Chu H, Greenland S. (2006) Multiple-Imputation for Measurement-Error Correction, *International Journal of Epidemiology*, 35(4): 1074–1081.

[19] Little RJA. (1992) Regression with Missing X’s: a Review. *Journal of the American Statistical Association*, 87(420), 1227–1237.

[20] U.S. Department of Agriculture, Economic Research Service. (n.d.). *Food Access Research Atlas*. Retrieved April 25, 2023, from https://www.ers.usda.gov/data-products/food-access-research-atlas/.

[21] **Lotspeich SC**, Jarrett RT, Epstein R, Schaffer A, Gracey K, Cull M and Raman R. (2020) Incidence of Children Placed in State Custody and Neighborhood-Level Risk Factors, *Child Abuse & Neglect,* 109, 104767.

[22] U. S. Bureau of the Census. (1994). *Geographic Areas Reference Manual*. U.S. Dept. of Commerce, Economics and Statistics Administration, Bureau of the Census.

[23] *PLACES*. Centers for Disease Control and Prevention. (2022). Public-use data file and documentation. https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Census-Tract-D/cwsq-ngmh.

[24] Hijmans R (2022). *geosphere: Spherical Trigonometry*. R package version 1.5-18, https://CRAN.R-project.org/package=geosphere.

[25] Duan R, Cao M, Wu Y, Huang J, Denny JC, Xu H and Chen Y. (2017). An Empirical Study for Impacts of Measurement Errors on EHR Based Association Studies. *AMIA Annual Symposium proceedings. AMIA Symposium*, 1764–1773.

[26] Amorim G, Tao R, **Lotspeich S**, Shaw PA, Lumley T, Patel,RC and Shepherd BE. (2022). Three-Phase Generalized Raking and Multiple Imputation Estimators to Address Error-Prone Data. *arXiv preprint arXiv:2205.01743*.

[27] **Lotspeich SC**, Richardson BD, Baldoni PL, Enders KP and Hudgens MG. (2023). Quantifying the HIV Reservoir with Dilution Assays and Deep Viral Sequencing. *arXiv preprint arXiv:2302.00516*.

[28] Amorim G, Tao R, **Lotspeich S**, Shaw PA, Lumley T and Shepherd BE. (2021) Two-Phase Sampling Designs for Data Validation in Settings with Covariate Measurement Error and Continuous Outcome, *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(4): 1368–1389.

[29] **Lotspeich SC**, Shepherd BE, Amorim GC, Shaw PA and Tao R. (2023) Optimal Multiwave Validation of Secondary Use Data with Outcome and Exposure Misclassification. *Canadian Journal of Statistics*. https://doi.org/10.1002/cjs.11772.

[30] **Lotspeich SC**, Grosser KF and Garcia TP. (2022). Correcting Conditional Mean Imputation for Censored Covariates and Improving Usability. *Biometrical Journal*, 64, 858–862.

[31] **Lotspeich SC** and Garcia TP. (2022). Escaping the Trap: Replacing the Trapezoidal Rule to Better Impute Censored Covariates with Their Conditional Means. *arXiv preprint arXiv:2209.04716*.

[32] Grosser KF, **Lotspeich SC** and Garcia TP. (2023). Mission Imputable: Correcting for Berkson Error When Imputing a Censored Covariate. *arXiv preprint arXiv:2303.01602*.

[33] Swayne MRE and Lowery Bryce. (2021). Integrating Transit Data and Travel Time into Food Security Analysis: A Case Study of San Diego, California. *Applied Geography*, 131: 102461.