

Linking Potentially Misclassified Healthy Food Access to Diabetes Prevalence

A. E. Mullan¹, P. D. A. Nguyen², S.C. Lotspeich²

1: Vanderbilt University - Department of Biostatistics - Nashville, TN



2: Wake Forest University - Department of Statistical Sciences - Winston-Salem, NC



for more!

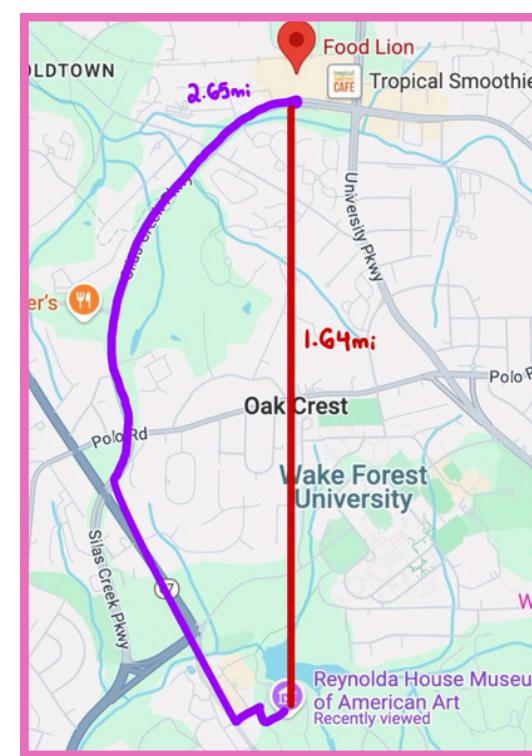


1. Motivation

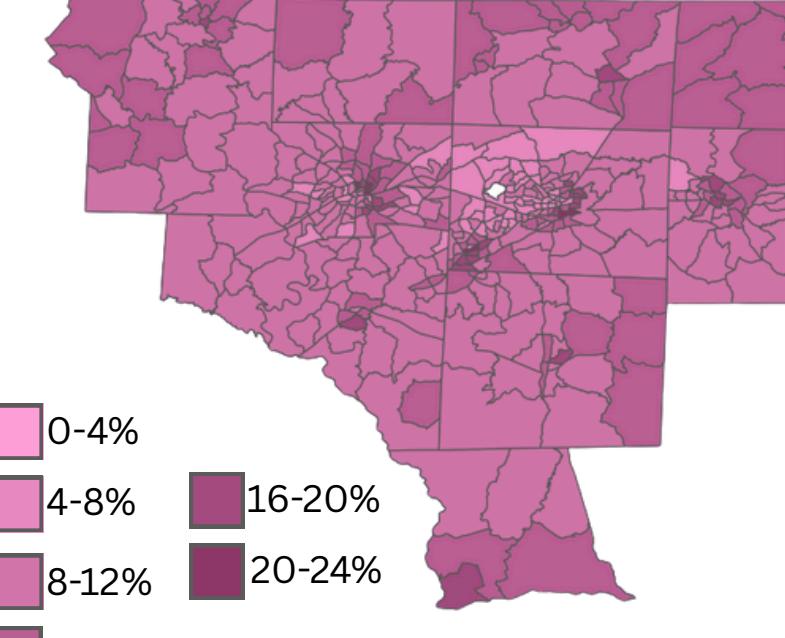
- A healthy diet increases the likelihood of good overall health and decreases risk of preventable illness (1).
- Maintaining a healthy diet requires consistent access to healthy food, which may be hindered by geography, income, or social factors such as structural racism.
- Review studies found high prevalence of diabetes in food-insecure households (2).

2. Distances and Data

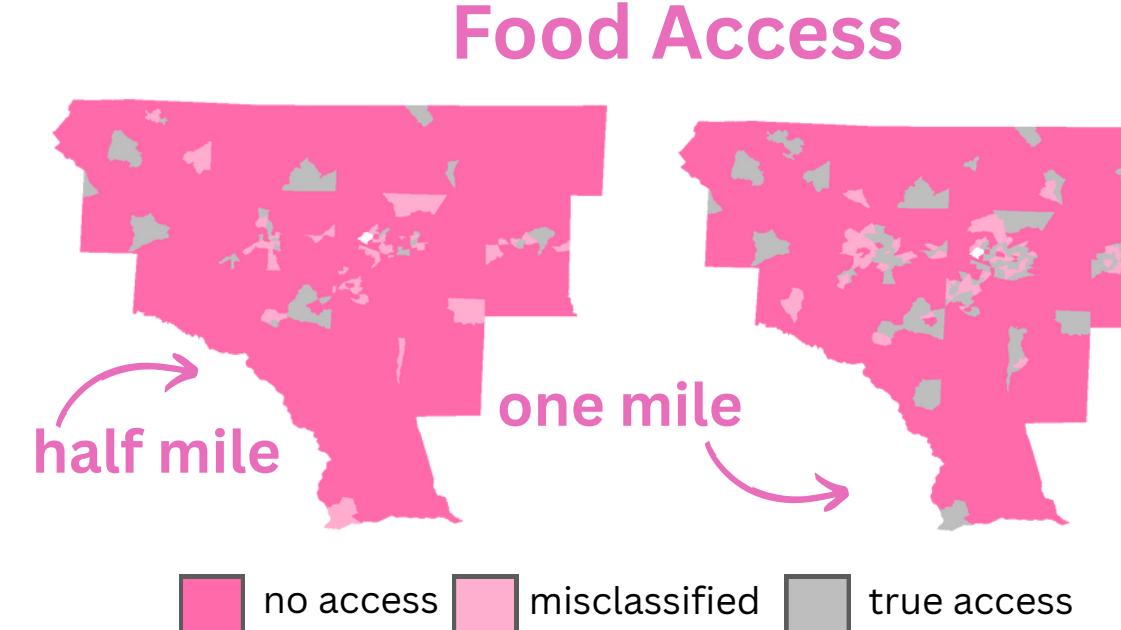
Distances



Diabetes Prevalence



Food Access



- We measure food access by constructing an **indicator variable** that evaluates to 1 if at least one healthy food retailer exists within a **given distance**.
→ (accurate but expensive)
- We measure distance in two ways, the **Haversine** distance and **route-based** driving distance.
→ (error-prone and always an underestimate)

3. Design and Analysis

- We use **two-phase design** to maximize the available food access information, measuring error-prone access X^* for all (N) neighborhoods and picking a **subset of n** neighborhoods to measure route-based access X .
- Factoring a **joint likelihood** sets us up to use the **EM** algorithm to find the **maximum likelihood estimator** for Poisson regression of **diabetes cases (Y)** on **food access and covariates (Z)**.

$$\ell(\beta, \eta) = \sum_{i=1}^N \left\{ Q_i \log P_\beta(Y | X, Z) P_\eta(X | X^*, Z) + (1 - Q_i) \log \sum_{x=0}^1 P_\beta(Y | X = x, Z) P_\eta(X = x | X^*, Z) \right\}$$

References

1. World Health Organization. Healthy diet, 2019. URL <https://iris.who.int/handle/10665/325828>

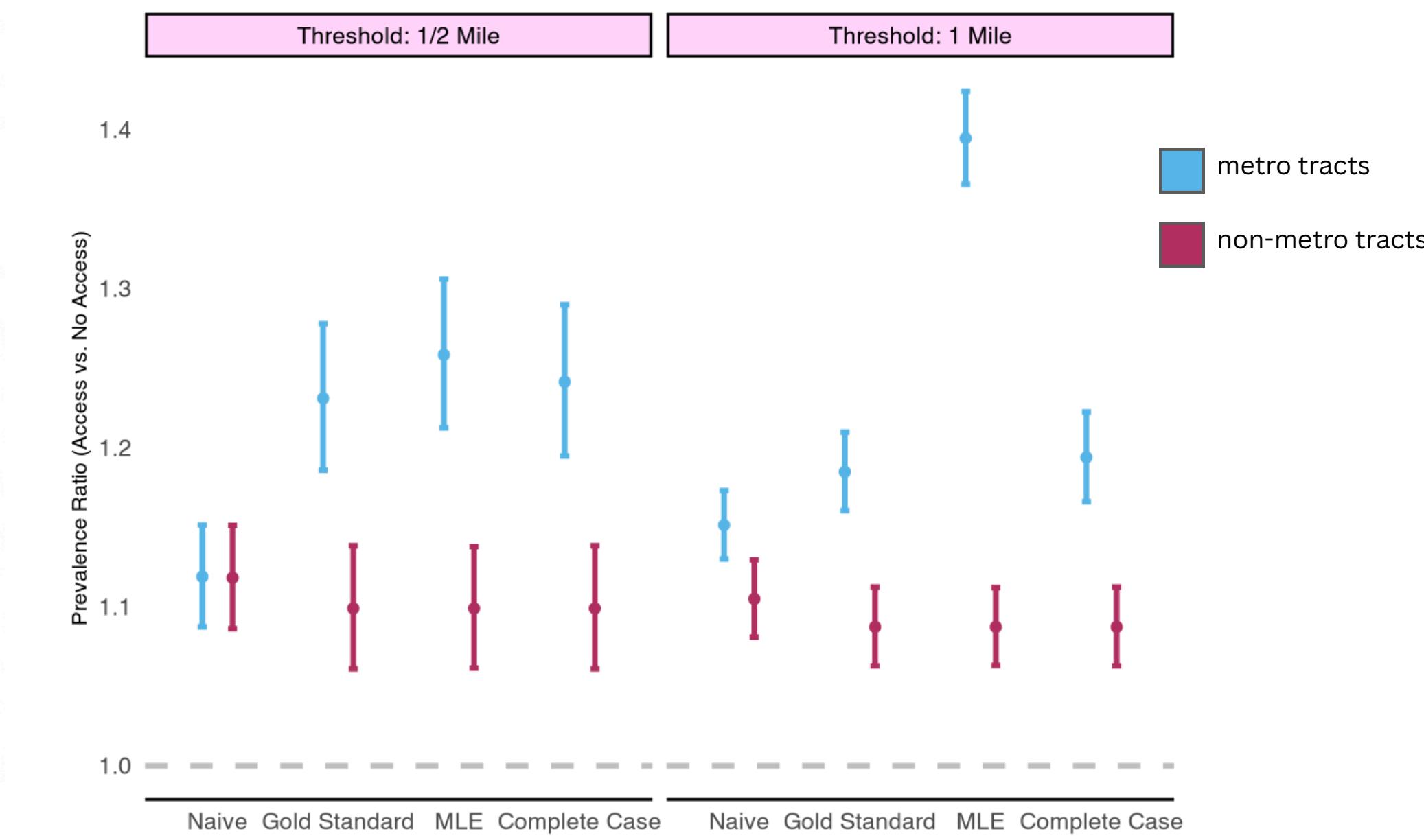
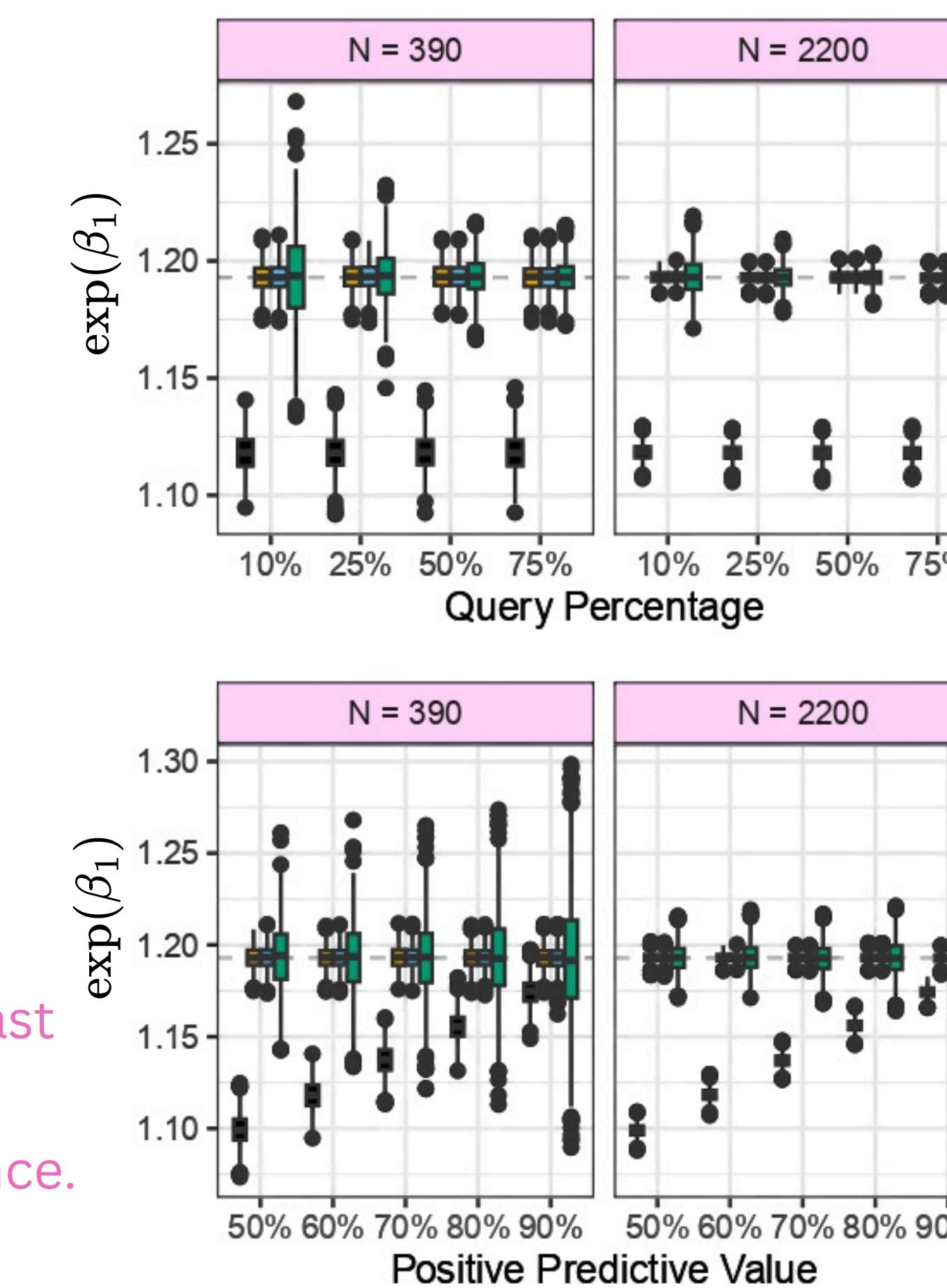
2. E. Gucciardi, M. Vahabi, N. Norris, J. P. Del Monte, and C. Farnum. The intersection between food insecurity and diabetes: a review. Current nutrition reports, 3:324–332, 2014

4.

Simulations and Case Study

Method █ naive █ gold standard █ MLE █ complete case

- The simulations explore various **query percentages** and **misclassification settings**.
- The **case study** models **diabetes prevalence** as a function of **food access** and metropolitan status.



5.

Takeaways

- The **MLE** retains the **unbiasedness** of the gold standard estimator and **recovers more efficiency** than the complete case in **simulations**.
- At both radii, we observe that tracts with **access to healthy food** tend to have **higher prevalences of diabetes** than tracts without it.

Acknowledgements



THE ANDREW SABIN FAMILY
CENTER FOR ENVIRONMENT
AND SUSTAINABILITY