eAppendix for **State and County Level Estimates of Healthcare Costs Associated with Food Insecurity**

Technical Appendix

NHIS and MEPS Data

To estimate the excess healthcare costs, if any, associated with food insecurity, we used linked data from the National Health Interview Survey (NHIS)[1] and the Medical Expenditure Panel Survey (MEPS)[2]. To do this, we needed to extract data on food insecurity status (from the NHIS Family file), and total healthcare expenditures (from MEPS). In addition, we needed to account for factors that could confound the association between food insecurity and healthcare costs. To identify these factors, we drew from a published conceptual model of the relationship between food insecurity and health.[3] Because our concern was to identify potential confounders, it was important to uses measurements of these confounders conducted at the same time as the assessment of food insecurity status. Therefore, we used variables from NHIS (from the person, sample adult, or sample child file as appropriate). These variables were age (from the person file), gender (from the person file), race/ethnicity (categorized as non-Hispanic white, non-Hispanic black, Hispanic, and Asian/other/multi-ethnic; from the person file), health insurance (categorized as private, Medicare, other public insurance [including Medicaid, Medicare-Medicaid 'dual eligibles', Children's Health Insurance Program, and Veterans Affairs], and uninsured; from the person file), income (expressed as a percentage of federal poverty level, which is adjusted annually for inflation and accounts for household size; from the family file), education (categorized as less than high school diploma, high school diploma, or greater than high school diploma; from the person file, not used in analysis of children), and census region (Northeast, South, Midwest, or West; from the person file).

Dartmouth Atlas of Healthcare

To estimate how a given county or state differed in healthcare spending (either based on prices or intensity of care) from the national average, we used data from the Dartmouth Atlas of Healthcare (http://www.dartmouthatlas.org/). Specifically, we calculated a 'cost factor' by taking the ratio of the average per beneficiary Medicare spending, adjusted for demographics, in a given county or state, divided by the national average. Because Medicare provides health insurance coverage in all US localities, and because it covers a standard set of services, demographically-adjusted Medicare spending is a good proxy for local variation in healthcare spending, as difference in spending reflect local differences in intensity of healthcare use and prices. This 'cost-factor' is greater than 1 for areas with higher than average costs and less than 1 for areas with lower than average costs. This correction allowed us to better adjust our estimates of what an 'average' person with food insecurity would cost in a given locality. To match the timeframe of MEPS data collection, we used data from 2012 and 2013, and created a 'cost-factor' averaged over those years.[5]

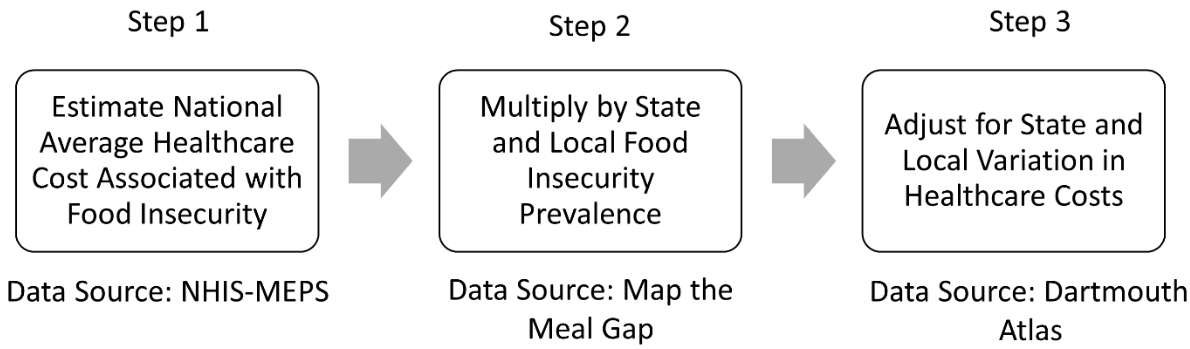
Estimating Healthcare Costs

To generate the cost estimates, we drew on prior work examining the association between food insecurity and healthcare costs.[6] This work used generalized linear modeling (GLM) for the estimates. However, healthcare cost data is notoriously difficult to analyze[7] and GLM relies on certain assumptions that may not always be met in practice. In particular, GLM requires assumptions about the distribution of healthcare costs that may not reflect real-world data, which often has a large number of $0 cost observations (point mass at 0) and a right-skewed distribution with a small number of individuals accruing large healthcare costs. To address these issues we extended the past work by applying a targeted maximum likelihood estimation approach (TMLE). TMLE is a doubly-robust analytic strategy that initially creates an estimate of the excess healthcare costs associated with food insecurity and then ‘updates’ that estimate using a sub-model that estimates the probability of being food insecure.[8] Additionally, TMLE allows for the incorporation, in its estimation procedures, of multiple types of analytic approaches, including both standard GLM approaches and non-parametric machine learning algorithms that do not make the same distributional assumptions as standard parametric modeling.[8] Taken together, these algorithms may be better able to accommodate the point mass at zero, skewed distributions, and extreme values common in analyses of healthcare data.[7] In this case, we employed an ensemble strategy which combined multiple candidate machine learning algorithms into a weighted average, where the weights of each individual algorithm were based on the accuracy of the modeling as indicated by mean squared error in a 10-fold cross-validation process. The algorithms included were a Bayesian GLM (gamma regression when modeling healthcare costs and logistic regression when modeling food insecurity), a GLM with a Tweedie distribution, a boosted trees algorithm, a generalized additive model, and a Bayesian additive regression trees algorithm. The combination of these algorithms, all of which make different assumptions, increases the chance that the model-based estimates will better reflect the ‘true’ difference in healthcare costs between otherwise similar individuals with and without food insecurity. TMLE provides an influence-curve based confidence interval for the estimates of excess healthcare costs, which we used in our analyses.[8]

Correlation Analysis

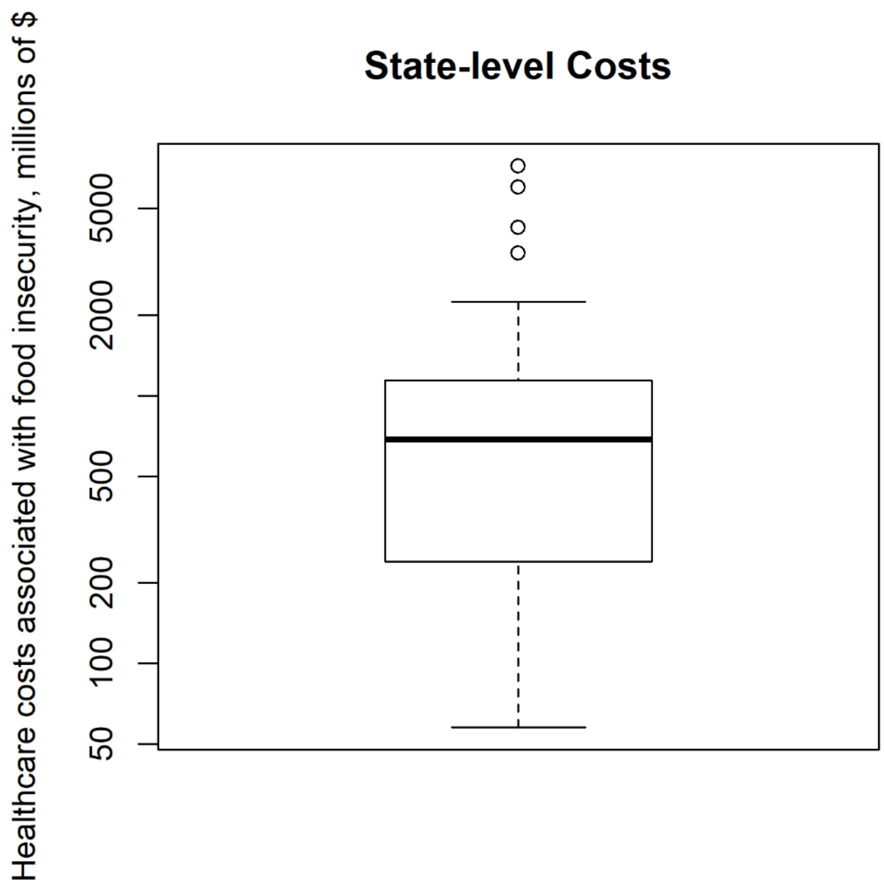
In order to examine which of two factors, food insecurity prevalence or local healthcare intensity and pricing are more closely associated with variation in healthcare costs associated with food insecurity, we conducted correlation analyses on the model-based estimates of food-insecurity associated healthcare costs. These analyses calculated R2 statistics relating the variation in food insecurity associated healthcare costs with variation in food insecurity prevalence or local cost factors.

eFigure 1: Flowchart of Analytic Strategy

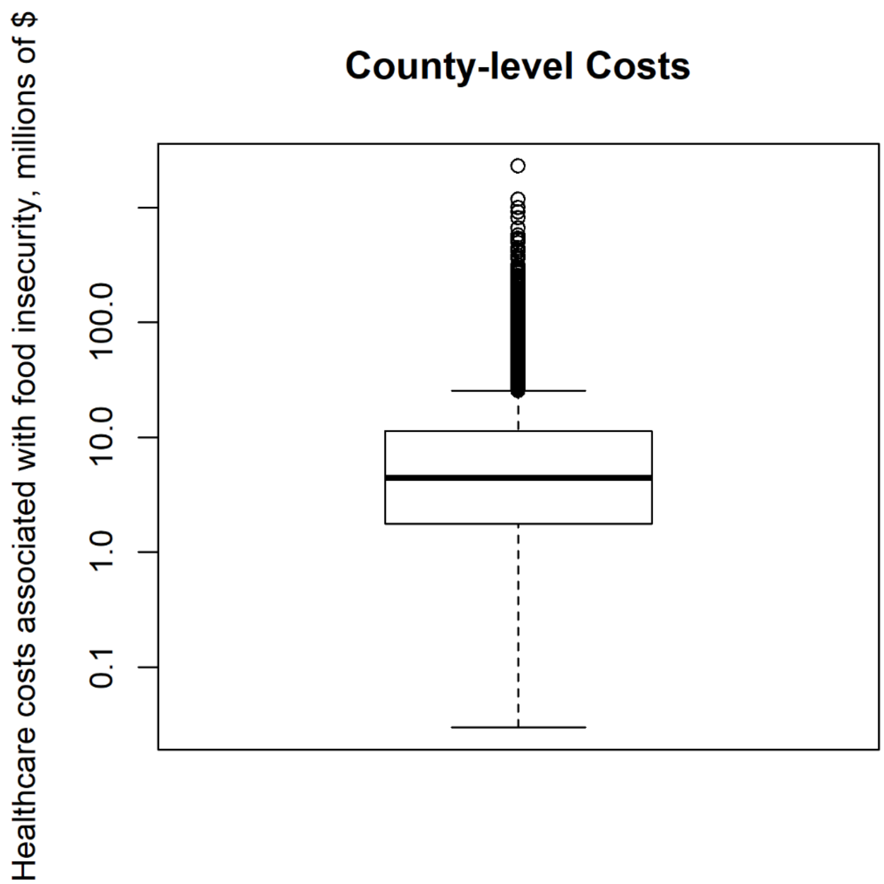


eFigure 1 Legend: Data for Step 1 Came from 2011 NHIS and 2012-2013 MEPS. For step 2, Map the Meal Gap uses state-level US Census and Bureau of Labor Statistics data from 2001-2016 to create models of food insecurity prevalence, which enable county-level food insecurity prevalence estimation for any time within that period. To express results for 2016 (the most recent available), we used 2016 5-Year American Community Survey data to estimate food insecurity prevalence for adults and children in 2016. Dartmouth Atlas cost data was used to adjust estimates for local healthcare use intensity and pricing. We used data covering the same time period as the MEPS data (2012-2013), as that was when cost data were collected. Because we were using 2016 food insecurity prevalence estimates, we express all dollar values in inflation-adjusted 2016 dollars.

eFigure 2: Box and Whisker Plot of Healthcare Expenditures by State



eFigure 3: Box and Whisker Plot of Healthcare Expenditures by State



1. NHIS - National Health Interview Survey Homepage. https://www.cdc.gov/nchs/nhis/index.htm. Published September 15, 2018. Accessed September 25, 2018.

2. Medical Expenditure Panel Survey Home. https://meps.ahrq.gov/mepsweb/. Accessed September 25, 2018.

3. Seligman HK, Schillinger D. Hunger and socioeconomic disparities in chronic disease. *N Engl J Med*. 2010;363(1):6-9. doi:10.1056/NEJMp1000072

4. Gundersen C, Engelhard E, Waxman E. Map the Meal Gap: Exploring Food Insecurity at the Local Level. *Appl Econ Perspect Policy*. 2014;36(3):373-386. doi:10.1093/aepp/ppu018

5. Downloads - Dartmouth Atlas of Health Care. http://www.dartmouthatlas.org/tools/downloads.aspx. Accessed September 25, 2018.

6. Berkowitz SA, Basu S, Meigs JB, Seligman HK. Food Insecurity and Health Care Expenditures in the United States, 2011-2013. *Health Serv Res*. June 2017. doi:10.1111/1475-6773.12730

7. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ*. 2001;20(4):461-494.

8. Schuler MS, Rose S. Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. *Am J Epidemiol*. 2017;185(1):65-73. doi:10.1093/aje/kww165