Does the National Health Service Corps Help Underserved Areas?

Sarah Van Alsten

Background and Aims

Medically underserved areas are communities with a high levels of unmet healthcare needs, often secondary to insufficient access to primary care physicians (Shi et al., 2005; Starfield, Shi, & Macinko, 2005). Formal designation of underserved areas began in the late 1980s and is based on the Index of Medical Underservice (IMU), a weighted score of census or county-level infant mortality rate, primary care provider to population ratio, percent of population older than 65, and percent of population below the federal poverty line. Although primary care recruitment to and subsequent health improvement in underserved areas are often cited targets for healthcare and physician workgroups, medically underserved areas are consistently faced with physician shortages. Determinants of shortages include lack of physician exposure to rural health in medical school, individual preference to work in urban centers, lower pay for primary care specialties, and hospital closures (Fagan et al., 2015; Parlier, Galvin, Thach, Kruidenier, & Fagan, 2018; Rabinowitz, Diamond, Markham, & Paynter, 2001).

To increase physician supply to underserved areas, several loan repayment and subsidized training programs for primary care practitioners have been instituted, including the National Health Service Corps (NHSC) through the U.S. Department of Health and Human Services (National Health Service Corps Scholarship Program, 2003). The NHSC grants primary care physicians up to $50,000 in medical school loan forgiveness in exchange for two years of service in a medically underserved area, although select states offer additional compensation. Early assessments of the NHSC have shown the program to be successful in recruiting physicians to areas of high need (Holmes, 2005); unfortunately, less than half of participating physicians remain in their initial area of service after completion of the program, leading to high rates of turnover and lack of continuity in the care of underserved community members (Heisler, 2017; Pathman, Konrad, Dann, & Koch, 2004). Further, despite the purported goal of primary care recruitment programs in improving population health, few evaluations have explored the effects of the NHSC on population health indicators.

The aim of this analysis is to determine whether participation in the NHSC reduces county-level all-cause

and disease-specific mortality (herein, simply 'mortality').

Method

Data for this study come from three sources: one, the Area Health Resource Files (US Department of Health and Human Services, 2014) for the ratio of primary care physicians-to-population; two, the American Community Survey (ACS) (U.S. Census Bureau; American Community Survey, 2014) for population estimates, age demographics, and poverty and unemployment rates; three, the Centers for Disease Control Wide-ranging Online Data for Epidemiologic Research (CDC WONDER) (Centers for Disease Control and Prevention, 1995) for infant mortality rates, low birthweight rates, and crude all-cause, cancer, and heart disease mortality rates. Briefly, the Area Health Resource Files are a compilation of 50 separate health service and economic datasets provided by the American Medical Association, American Hospital Association, and Bureau of Health Workforce. Files are updated annually to monitor health service worker supply, regional demand for specific medical procedures (e.g. elective surgery, orthodontia), and costs of care. Among the provided indicators is a census of all physicians licensed in the U.S., disaggregated by county of practice and medical specialty. The ACS is an ongoing, nationally representative, multistage probability sample conducted by the U.S. Census Bureau as a supplement to the decennial census. Roughly 0.6% of the U.S. population is sampled for the ACS in any given year, and, because the Census Bureau contacts respondents multiple time to encourage participation, the response rate in 2014 was 96.7% (Torrieri, 2014). The ACS includes census tract-level demographics including estimates of population by age group and race/ethnicity, poverty, unemployment, and educational attainment. Finally, CDC WONDER is an online query system maintained by the National Center for Health Statistics that allows users to generate area- and period-level natality and mortality data from linkage to birth records and the National Death Index (> 90% of all deaths are accounted for in any given year) for all U.S. counties between 1999 and 2016. Area-level mortality rates disaggregated by age are provided for the 57 leading causes of death as determined by death certificate listed ICD-10 codes, although rates are suppressed for regions with fewer than 10 deaths from a specific cause.

Index of Medical Underservice Scoring

A county's eligibility to participate in the NHSC is determined by the index of medical underservice score, which quantifies an area's degree of physician shortage relative to healthcare needs. Scores are typically generated automatically by the Health Resource and Service Administration from Area Health

Resource Files, CDC WONDER, and Census data, with all scores of 62 or below eligible for NHSC participation. Although historic data on county eligibility are publicly available, actual index of medical underservice scores are only provided for counties currently eligible for the NHSC (i.e. in 2019; 2020 eligibility has not yet been determined). Thus, to obtain index of medical underservice scores for earlier years, I regenerate scores for each county in the U.S. using historic data and the Health Resource and Service Administration scoring guidelines (U.S. Department of Health and Human Services, 2015). Scores are calculated as the weighted sum of a county's primary care physician-to-population ratio, percentage of residents age 65 or older, percentage of population at or below the federal poverty level, and infant mortality rate or, in instances where infant mortality rates were suppressed due to small sample sizes, rates of low birthweight. All score subcomponents are based on five-year averages to prevent substantial annual variation in program eligibility for treated counties (i.e. the averages of 2009-2014 components were used to calculate 2014 scores), and counties with qualifying scores automatically retain eligibility for two years following designation, regardless of subsequent scores.[1] Though the NHSC has been in operation since the late 1980s, the current index of medical underservice scoring criteria were first implemented in 2014 such that all potential carry-overs from 2012/2013 to 2014 were disallowed. Thus, unlike for earlier periods, a county's NHSC participation in 2014 is based strictly on its 2014 score rather than any one of its 2012, 2013, or 2014 scores. Table 1 provides a more detailed description of scoring criteria and weighting procedures.

Treatment Assignment and Identification

Let Z_i be a dichotomous variable denoting the eligibility of county i to participate in the NHSC in 2014 (with 0 = ineligible, 1 = eligible). Also let X_i denote the index of medical underservice score in 2014, and D_i (1 = treated, 0 = control) denote actual participation in 2014. Z_i is a deterministic function of X_i such that:

$$E[Z_i|X_i] = \begin{cases} 1 \text{ if } X_i \leq 62 \\ 0 \text{ if } X_i > 62 \end{cases}$$

Due to budget constraints, not all eligible counties receive NHSC funding; however, no ineligible counties receive funding. Therefore, D_i is a probabilistic function of eligibility $f(Z_i)$ for counties below the cutoff,

---

[1] Although most of these components are not based on census-level data, it is not clear if or how the Department of Health and Human Services incorporates sampling error into score components. Thus, for generated scores, I use the point estimate in all analyses.

and deterministic for counties above such that:

$$E[D_i|Z_i] = \begin{cases} f(Z_i) & \text{if } Z_i = 1 \\ 0 & \text{if } Z_i = 0 \end{cases} \text{ where } f(Z_i) \neq 0$$

Because of this feature, monotonicity is guaranteed to hold, with the exception that state governors may grant exemptions for ineligible rural counties with demonstrated primary care shortages to receive funding for rural health clinics. However, the application process for exemption is long and requires extensive administrative oversight, with few requests ultimately being approved, so the population of defiers is both known and small (a total of five counties were exempted in 2014, all of which I exclude from further analyses). As shown in Table 2, derived treatment assignment according to simulated scores adhered to known participation and eligibility rules, with zero counties above the cutoff participating and approximately one-half of those below participating.

In addition to monotonicity, the ability to identify the RDD effect as the local average treatment effect for the subpopulation of compliers (i.e. counties which are not perpetually enrolled- or never enrolled- in the program) which have IMU scores near the cutoff hinges upon several additional assumptions, namely, continuity in potential outcomes and potential treatment near the threshold, the relevance of Z, and the exclusion restriction. While it is not necessarily possible to assess continuity in potential outcomes, I assess balance in covariates for counties just above and below the cutpoint and, to assess potential manipulation of the running variable X_i, I check for discontinuity in the density of observations just above and below the cutoff score (McCrary, 2008). I evaluate the relevance of Z using weak instrument tests. Finally, in regards to the exclusion restriction, eligiblity itself should not be linked to mortality other than through actual program participation, as NHSC funding cannot be used for other healthcare needs and does not impact eligibility for other population health programs. It is also expected that eligibility would not influence mortality through economic progress - at most, one to two physicians per county are recruited, which should neither appreciably increase county population nor substantially stimulate local economic conditions. Of greater threat to exogeneity are the IMU score components themselves, in particular poverty and age (infant mortality is excluded from other mortality estimates, while physician supply is the direct target of the program and is less apt to influence health other than through actual patient care; see Figure 4 in supplement). Assuming counties just above and below the cutoff are balanced on these components improves the capacity for identification, therefore, I test this explicitly in sensitivity analyses. However, population age distributions varied widely both between and within eligible and ineligible counties and, as the association between age and mortality is generally nonlinear and discontinuous (e.g. 64 year olds

4

have dramatically higher risk of death than do 18 year olds), these differences may lead to a violation of the exclusion restriction even with balance on raw percentages of elderly residents. To accomodate for this possibility, I used cause-specific crude rates and county-level population estimates for each 10-year age stratum to standardize cause-specific rates to the 2000 U.S. standard million (Anderson & Rosenberg, 1998), creating an additional outcome measure that was comparable across counties with different age distributions.

Model Specification and Estimation

To recover the regression discontinuity effect τ of county NHSC participation D_i on 2015 mortality rate Y_i, I use a fuzzy regression discontinuity design and estimate τ with a two-stage least squares approach where D is instrumented by Z. I allow for possible heterogeneity in the treatment effect as a function of X_i by including interaction terms between the zero-centered transformation of X_i ($\tilde{X}_i$) and eligibility Z_i. The models for the first stages are represented by equations 1 and 2, where $f_1(\tilde{X}_i)$ is a given polynomial transformation of the centered running variable.

$$D_i = \alpha_i + \gamma_0 Z_i + \gamma_1(f_1(\tilde{X}_i)) + \epsilon_i \ (1)$$

$$f_1(\tilde{X}_i)D_i = \alpha_i + \gamma_2 Z_i + \gamma_3(f_1(\tilde{X}_i))Z_i + \epsilon_i \ (2)$$

The second stage is estimated by equation 3, where $\hat{D}_i$ represents the fitted values of equation 1, and the various functions of $\tilde{X}_i\hat{D}_i$ represents fitted values from various functional forms of equation 2, with $f_1(\tilde{X}_i) = \tilde{X}_i^p$.

$$Y_i = \alpha + \tau\hat{D}_i + \beta_0\tilde{X}_i + \beta_1(\tilde{X}_i\hat{D}_i) + ...\beta_p(\tilde{X}_i^p\hat{D}_i) + \mu_i \ (3)$$

To determine the appropriate functional form of the regression, I estimate τ using both the full sample with first, second and third order polynomials (i.e. where a $f_1(\tilde{X}_i) = \tilde{X}_i; \tilde{X}_i^2; \tilde{X}_i^3$), and with local linear regression using a triangular kernel and observations within the robust bias-corrected (RBC) bandwidth (Calonico, Cattaneo, and Farrell, 2019).[2]

Sensitivity Analyses

I perform placebo tests using a range of false cutoff values of X_i to estimate local average treatment effects at different points along the running variable. To assess balance in covariates around the cutoff, I

---

[2]An aside regarding weights and clustering: Though counties do differ substantially in population, weighting for size would typically achieve an effect estimate appropriate for individual, rather than population inference. Additionally, although there is clear geographic clustering (e.g. counties within states), treatment (score) is not randomized at this level. Further, some clusters (namely, Delaware and Rhode Island) contain only two participating counties, leading not only to highly imbalanced cluster sizes and a substantial reduction in power.

estimate the LATE of program participation on 2014 percent poverty, percent of population over age 65, county-level median household income, percentage of residents who live in rural areas and who are White, Black, or Hispanic, and rates of unemployment, high school graduation, uninsurance, obesity, smoking, and insufficient physical activity using the same model specifications as above. As a final robustness check, I assess the effect of 2014 NHSC participation on placebo outcomes of 1999, 2005, and 2009 age-standardized mortality rates.

Results

Descriptive Statistics

A total of 851 counties were eligible for the NHSC in 2014, 406 (47.7%) of which ultimately received a physician recruit. In general, even restricting to counties within the median estimated RBC bandwidth, eligible counties differed significantly from ineligible counties by simple difference-in-means. On average, counties with IMU scores below the cutoff were less populous, had greater percentages of residents living in rural areas, lower percentages of non-Hispanic White residents, and performed worse on socioeconomic and health indicators, although in most instances differences were slight (Table 2). The density of observations just above and below the cutoff was continuous, showing no evidence of manipulation of the running variable (Figure 1; $p\_density = 0.68$). Table 3 confirms that Z and interactions between Z and X were strong instruments (all $F > 340$, $p < 0.001$).

Regression Discontinuity Effect

Table 4 depicts the estimated effects of participating in the NHSC program on county-level age-standardized and crude mortality rates. Virtually all estimated coefficients were positive, indicating that participation in the NHSC increased mortality rates, and, in general, estimates derived using the full sample indicated a stronger effect of program participation than did local estimations. Similarly, the absolute effect of program participation was larger for all-cause mortality than for disease-specific mortality, with the largest estimated effect being an absolute [i.e. crude] increase of 359 deaths per 10,000 person-years (95% CI = 317.53 - 401.39) in treated counties near the cutoff versus similarly scored untreated counties. However, particularly for local estimators with RBC confidence intervals, most estimates were imprecise and non-significant. Age-standardization improved precision somewhat, such that even after bias correction some estimates indicated statistically significant increases. With the exception of crude all-cause mortality rate, the magnitude and direction of effects were robust against alternate polynomial specifications.

Robustness to Various Bandwidth Selections

Based on graphical inspection of regression plots (Figures 6-8, available in Supplement), quadratic specifications provided a better fit to the data than did linear specifications, while cubic specifications did not provide substantial improvement over quadratic models. In balancing parsimony and correct specification, I used local models including quadratic transformations of X_i for further robustness checks.

For both age-standardized and crude all-cause mortality, the RBC estimate chosen at the optimal RBC bandwidth was contained in the 95% confidence intervals of estimated effects for both narrower and wider bandwidths (Figure 2). The magnitude of estimate effects was also relatively stable regardless of bandwidth selection, albeit with narrow bandwidths yielding an estimated effect approximately twice as large (an estimated excess 608 deaths per 10,000 person-years for NHSC participating counties) as that for the optimal bandwidth. Additionally, while the 95% confidence interval at the optimal bandwidth excluded the null for age-standardized rates, choosing a bandwith 1 point larger or smaller led to a widening of confidence intervals and inclusion of the null. Crude effects showed more sensitivity to choice of bandwidth, with point estimates including both positive and negative effects (Figure 2, bottom). In no case did crude RBC confidence intervals exclude the null.

Similar patterns occurred in both heart disease and cancer-specific mortality models (Figures 3 and 4). The positive estimated effects for age-standardized heart disease and cancer mortality were contained within the RBC 95% confidence interval for all alternative bandwidths excepting the narrowest (absolute distance of 4 points from the cutoff), after which estimated effects rose steeply. All regression discontinuity effects near the optimal bandwidth were nonsignificant but again indicated a higher mortality rate for treated counties. For crude rates, the effect of the NHSC on mortality were consistently null and small in magnitude.

Placebo Outcomes

The of placebo fuzzy regression discontinuities assessing covariate balance and program participation on pretreatment outcomes suggested that participation in the NHSC was not associated with placebo outcomes (Table 5). Encouragingly for the known possible exogeneity violation by IMU score components, NHSC participation was not associated with either percent county residents below the federal poverty line ($\tau = -2.37$, RBC 95% CI = -5.76 - 1.02) or percent of county residents age 65 or older ($\tau = 2.77$, RBC 95% CI = -1.36 - 6.91). Placebo treatment effects for all sociodemographic characteristics were non-significant, although there was some indication that participating counties had higher uninsured rates ($\tau = 0.04$, RBC 95% CI = - 0.01 - 0.08), smoking rates ($\tau = 0.06$, RBC 95% CI = -0.02 - 0.13), and lower percentages of non-Hispanic White residents ($\tau = -0.09$, RBC 95% CI = -0.28 - 0.10).

However, the local effects of 2014 NHSC participation on pretreatment mortality rates were significant in several instances, namely, for 1999 age-standardized all cause mortality ($\tau = 119.04$, RBC 95% CI = 5.29 - 232.8) and 2009 age-standardized cancer mortality rate ($\tau = 89.28$, 95% CI = 10.23 - 85.08). There are two possible explanations for this result. One, despite the change to eligbility criteria in 2014, some counties may have participated in the NHSC prior to 2014, meaning that pretreatment "effects" are a byproduct of past eligibility. This is likely more applicable to 2009 than 1999 mortality, given serial correlation of both mortality rate and eligibility determinants. Two, there may be reverse causality such that counties with higher historica mortality are preferentially chosen by the Department of Health and Human Services or by physicians themselves to receive funding.

Placebo Cutoffs

Using the sample of counties below the cutoff, there was consistently no local regression discontinuity effect on age-standardized or crude all cause mortality when setting placebo thresholds at IMU scores 10 to 20 points below the true threshold (i.e. for IMU scores of 42 - 52; Figure 9). With the exception of a placebo threshold 6 points below the true threshold, at which the estimated τ was larger than the 'true' RD effect (excess of 497 deaths versus 161 per 10,000 person-years) and significant, placebo thresholds closer to the true cutoff also showed null effects but were larger in magnitude and less stable than those observed further from the true threshold. At no placebo cutoff were estimated effects for crude heart disease mortality significant (Figure 10, bottom), but significant effects were observed at placebo cutoffs 17 and 6.8 points below the true threshold (Figure 10, top). As for all cause mortality, these estimated effects were substantially larger than the true discontinuity effect, indicating an excess of 166 and 302 versus 55 deaths from heart disease per 10,000 person-years, respectively. As depicted in Figure 11, placebo thresholds for both age-standardized and crude cancer mortality yielded similar null effects, again with the exception of a placebo threshold 5 points below the true cutoff for age-standardized rates ($\tau = 120.88$) and 4.3 points below the threshold for crude rates ($\tau = 1026.53$).

Given the relative consistency of identifying an effect at false cutoff scores in the range of 55 - 58 (4 to 7 points below an IMU score of 62), I performed additional exploration of treatement probability at various IMU scores (Figure 12). The anticipated probabilty of receiving an NHSC physician is approximatley equal for counties with IMU scores between 52 and 58, after which probability begins to drop, reaching zero [as expected] for scores greater than 62. There is no evidence of preferential sorting around these thresholds, and the only locally imbalanced covariate was percent of elderly residents (Table 6 in supplement). Thus, it is possible that the significant effects at these placebo thresholds are driven by imbalance in age. Alternatively, this may imply an implict threshold, perhaps due to administrative preference for specific

counties, remains an area for future investigation.

Conclusion and Limitations

At best, results indicated no effect of participation in the NHSC on county-level mortality; at worst, participation in the NHSC may lead to increased mortality for counties with scores near the cutoff. Nonetheless, a number of potential threats to identification were present in the data, including associations with pretreatment outcomes and significant effects at points along the running variable other than the true cutpoint. I did not find threats to the exclusion restriction in the potential confounding covariates included, but there remains the possibility that other unmeasured variables violated this assumption. The estimand may also have limited generalizable- for instance, I exploit the anamoly in 2014 scoring in order to identify the treatment effect, but it may be that this same anamoly makes 2014 qualitatively different from treatment in other years. For instance, changes in administrative rules necessarily require additional time and investment, perhpas detracting from other program components in ways that adversely affect health. Moreover, the estimand is applicable only to counties near the cutoff, so highly underserved areas with the greatest need likely experience different effects.

Additionally, a number of limitations present in the data may have lead to bias and/or measurement error, the most prominent of which was the fact that scores did not come from the NHSC administering body itself. Still, I will note that scores were generated using the same data sources and criteria used by the Department of Health and Human Services, and that generated scores paralleled prior knowledge of participation and eligbility.

Table 1: Description of index of medical underservice score subcomponents and weighting

Infant Mortality		Percent Over 65		Primary Care Ratio		Percent in Poverty	
Value	Score	Value	Score	Value	Score	Value	Score
≤ 8	26.0	≤ 7	20.2	≤ 0.05	0.0	≤ 0.1	25.1
≤ 9	25.6	≤ 8	20.1	≤ 0.1	0.5	≤ 1	24.6
≤ 10	24.8	≤ 9	19.9	≤ 0.15	1.5	≤ 4	23.7
≤ 11	24.0	≤ 10	19.8	≤ 0.2	2.8	≤ 6	22.8
≤ 12	23.2	≤ 11	19.6	≤ 0.25	4.1	≤ 8	21.9
≤ 13	22.4	≤ 12	19.4	≤ 0.3	5.7	≤ 10	21.0
≤ 14	20.5	≤ 13	19.1	≤ 0.35	7.3	≤ 12	20.0
≤ 15	20.5	≤ 14	18.9	≤ 0.4	9.0	≤ 14	18.7
≤ 16	19.5	≤ 15	18.7	≤ 0.45	10.7	≤ 16	17.4
≤ 17	18.5	≤ 16	17.8	≤ 0.5	12.6	≤ 18	16.2
≤ 18	17.5	≤ 17	16.1	≤ 0.55	14.8	≤ 20	14.9
≤ 19	16.4	≤ 18	14.4	≤ 0.6	16.9	≤ 22	13.6
≤ 20	15.3	≤ 19	12.8	≤ 0.65	19.1	≤ 24	12.2
≤ 21	14.2	≤ 20	11.2	≤ 0.7	20.7	≤ 26	10.9
≤ 22	13.1	≤ 21	9.8	≤ 0.75	21.9	≤ 28	9.3
≤ 23	11.9	≤ 22	8.9	≤ 0.8	23.1	≤ 30	7.8
≤ 24	10.8	≤ 23	8.0	≤ 0.85	24.3	≤ 32	6.6
≤ 25	9.6	≤ 24	7.0	≤ 0.9	25.3	≤ 34	5.6
≤ 26	8.5	≤ 25	6.1	≤ 0.95	25.9	≤ 36	4.7
≤ 27	7.3	≤ 26	5.1	≤ 1	26.6	≤ 38	3.4
≤ 28	6.1	≤ 27	4.0	≤ 1.05	27.2	≤ 40	2.1
≤ 29	5.4	≤ 28	2.8	≤ 1.1	27.7	≤ 42	1.3
≤ 30	5.0	≤ 29	1.7	≤ 1.15	28.0	≤ 44	1.0
≤ 31	4.7	≤ 30	0.6	≤ 1.2	28.3	≤ 46	0.7
≤ 32	4.3	>30	0.0	≤ 1.25	28.6	≤ 48	0.4
≤ 33	4.0	–	–	>1.25	28.7	>48	0.1
≤ 34	3.6	–	–	–	–	–	–
≤ 35	3.3	–	–	–	–	–	–
≤ 36	3.0	–	–	–	–	–	–
≤ 37	2.6	–	–	–	–	–	–
≤ 39	2.0	–	–	–	–	–	–
≤ 41	1.4	–	–	–	–	–	–
≤ 43	0.8	–	–	–	–	–	–
>43	0.2	–	–	–	–	–	–

[1] Infant Mortality Rate = Infant Deaths per 1000 Live Births
[2] PCP Ratio = Total FTE Non-federal Primary Care Providers per 1000 Population
[3] Poverty = At or Below Federal Poverty Level

Table 2: Sample characteristics of counties with Index of Medical Underservice (IMU) scores within the median RBC bandwidth. All characteristics are based on 2014 data or, for IMU components, the five-year average of 2009-2014 data and are shown as Mean (SD). Except for total population, infant health index, and primary care ratio, all values represent proportions of county residents with given characteristic.

	IMU > 51.88 and <= 62	IMU > 62 and <= 72.11	P-value
N	620	690	–
Enrolled in NHSC	290 (47.90)	0 (0.00)	–
IMU Components			
Poverty[1]	0.17 (0.03)	0.14 (0.05)	< 0.001
Older Than 65 Years	0.13 (0.05)	0.14 (0.05)	< 0.001
Infant Health Index[2]	25.67 (1.70)	24.76 (0.74)	< 0.001
Primary Care Ratio[3]	0.81 (0.57)	1.36 (0.60)	< 0.001
County Characteristics			
Total Population	29,304 (55,339)	54,822 (129,421)	< 0.001
Rural	0.75 (0.24)	0.60 (0.26)	< 0.001
Non-Hispanic White	0.76 (0.20)	0.81 (0.18)	< 0.001
Non-Hispanic Black	0.12 (0.17)	0.07 (0.11)	< 0.001
Hispanic	0.08 (0.14)	0.09 (0.14)	0.001
Health Indicators			
Current Smokers	0.24 (0.06)	0.22 (0.06)	< 0.001
Obese	0.33 (0.04)	0.32 (0.04)	< 0.001
Physically Inactive	0.30 (0.05)	0.26 (0.04)	< 0.001
Uninsured	0.16 (0.04)	0.14 (0.04)	< 0.001

[1] Poverty = At or Below Federal Poverty Level

[2] Infant Mortality Rate = Infant Deaths per 1000 Live Births

[3] PCP Ratio = Total FTE Non-federal Primary Care Providers per 1000 Population

Table 3: Tests of Weak Instruments For IMU Scores and Eligibility

| Outcome | Polynomial Order | | | | | | |
| | 1st | | | 2nd | | 3rd | |
	Outcome	F-Statistic	P-value	F-Statistic	P-value	F-Statistic	P-value
1	All-Cause Mortality	801.7083	< 0.001	480.7064	< 0.001	343.1566	< 0.001
3	Heart Disease Mortality	801.7083	< 0.001	482.6809	< 0.001	344.6082	< 0.001
5	Cancer Mortality	788.5905	< 0.001	472.8127	< 0.001	337.6507	< 0.001

Note:

Test statistics differ slightly by outcome given missing data on heart disease and cancer mortality
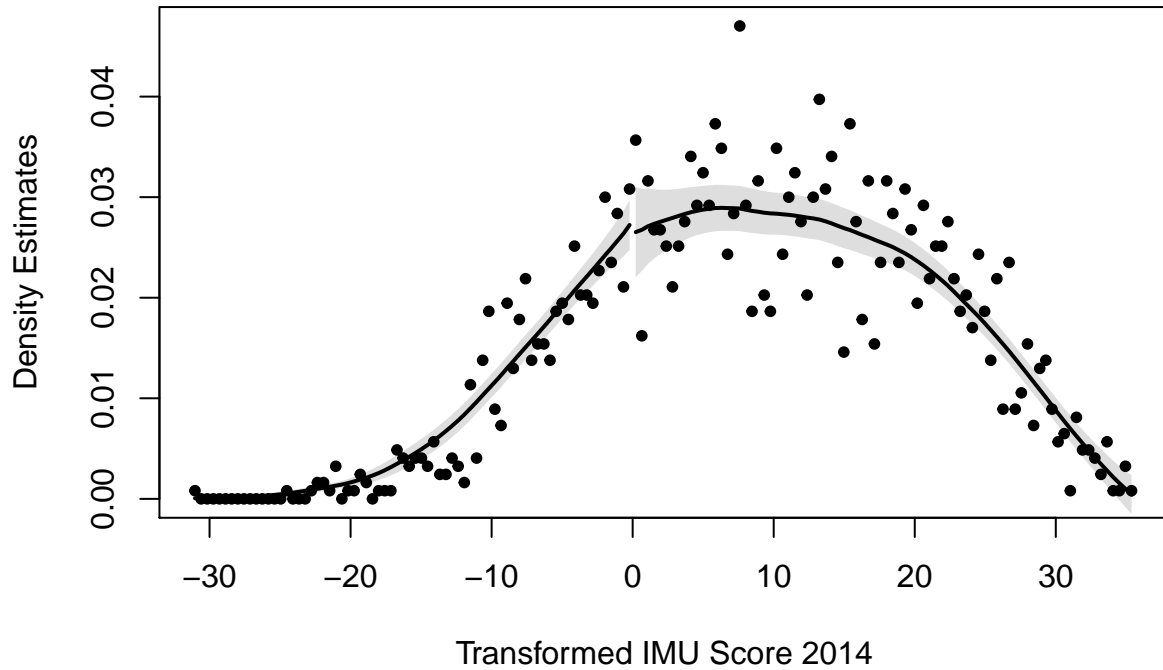
Figure 1: McCrary Density Test shows no evidence of sorting at the cutoff (p = 0.68)

Table 4: Effect of NHSC Participation on County-Level Mortality

		Full Sample			Local		
			95% CI				95% CI
Outcome	Order	Estimate	Conventional	BW	Estimate	Conventional	Robust
Age-Standardized Rate							
All Cause	1	**190.46**	**167.64 - 213.27**	9.39	171.75	**76.46 - 267.04**	**63.04 - 280.47**
	2	**190.40**	**167.59 - 213.22**	11.67	160.85	**37.51 - 284.18**	**22.05 - 299.65**
	3	**190.41**	**167.60 - 213.23**	10.24	79.32	-15.97 - 174.61	-29.39 - 188.04
Heart Disease	1	**74.93**	**65.19 - 84.67**	7.26	51.99	**4.64 - 99.33**	-4.17 - 108.14
	2	**74.91**	**65.17 - 84.65**	12.75	55.33	**2.03 - 108.64**	-4.82 - 115.49
	3	**74.92**	**65.18 - 84.66**	11.36	37.06	-10.28 - 84.41	-19.09 - 93.22
Cancer	1	**32.73**	**27.36 - 38.11**	8.50	39.78	**13.39 - 66.17**	**9.01 - 70.55**
	2	**32.74**	**27.37 - 38.12**	9.96	38.31	**3.34 - 73.29**	-1.22 - 77.85
	3	**32.74**	**27.37 - 38.12**	11.84	30.58	**4.18 - 56.97**	-0.19 - 61.35
Crude Rate							
All Cause	1	**359.46**	**317.53 - 401.39**	7.38	61.29	-99.44 - 222.01	-129.59 - 252.16
	2	**359.15**	**317.22 - 401.07**	8.36	-22.01	-230.84 - 186.82	-253.77 - 209.75
	3	**359.04**	**317.11 - 400.96**	12.55	-51.68	-212.41 - 109.04	-242.56 - 139.2
Heart Disease	1	**134.00**	**117.40 - 150.61**	9.75	23.94	-35.06 - 82.94	-46.21 - 94.09
	2	**133.93**	**117.33 - 150.54**	10.01	7.50	-76.17 - 91.18	-86.8 - 101.81
	3	**133.89**	**117.28 - 150.49**	14.92	9.81	-49.19 - 68.81	-60.34 - 79.96
Cancer	1	**80.04**	**69.65 - 90.44**	8.47	31.70	-11.25 - 74.65	-19.66 - 83.06
	2	**80.08**	**69.69 - 90.47**	11.34	31.50	-23.38 - 86.37	-30.67 - 93.67
	3	**80.01**	**69.62 - 90.40**	10.67	12.60	-30.34 - 55.55	-38.76 - 63.97

Note:
Estimates represent tau FRD estimated using either IV or a local regression with a triangular kernel
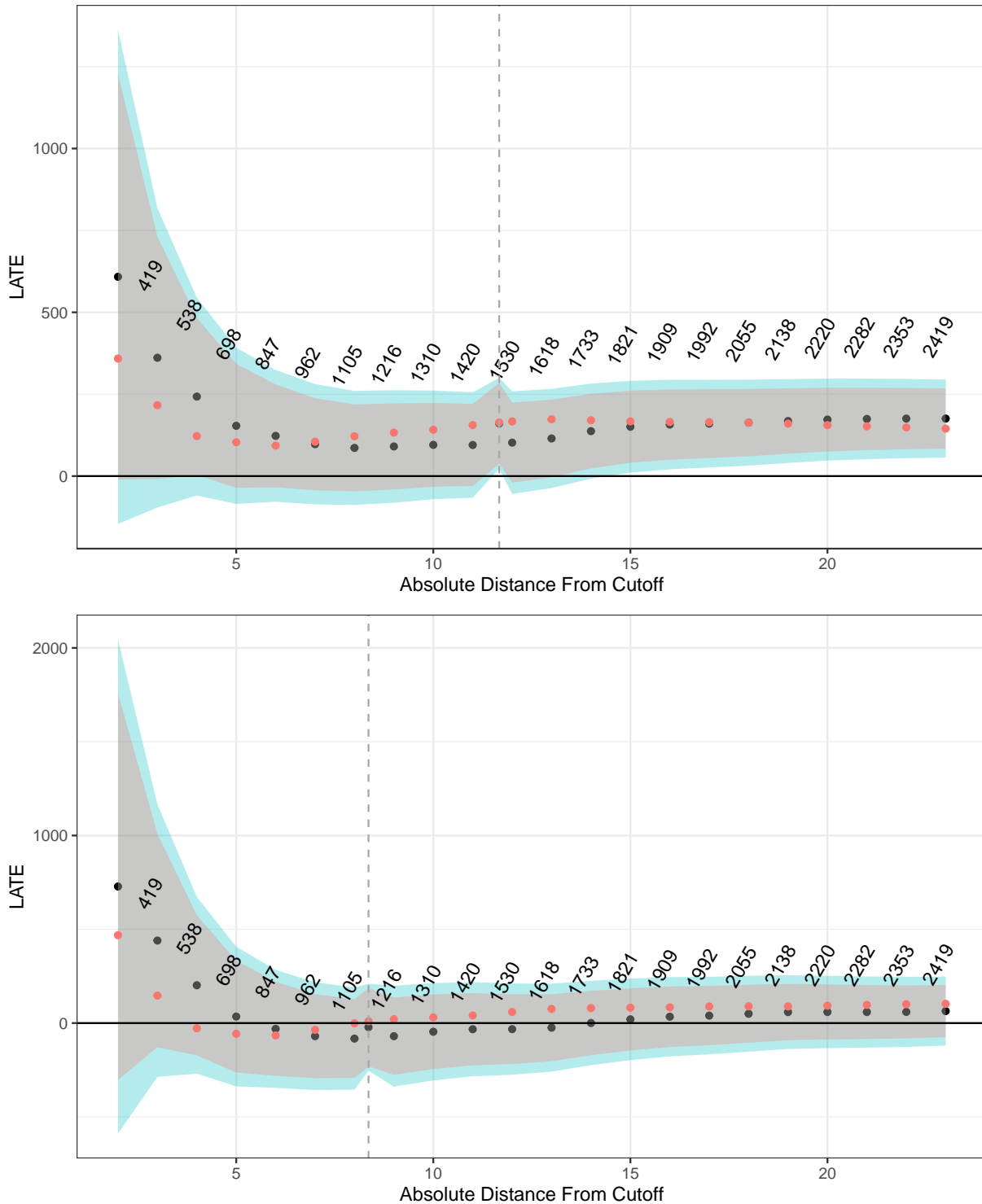[a] Bold face denotes statistical significance ($p < 0.05$)
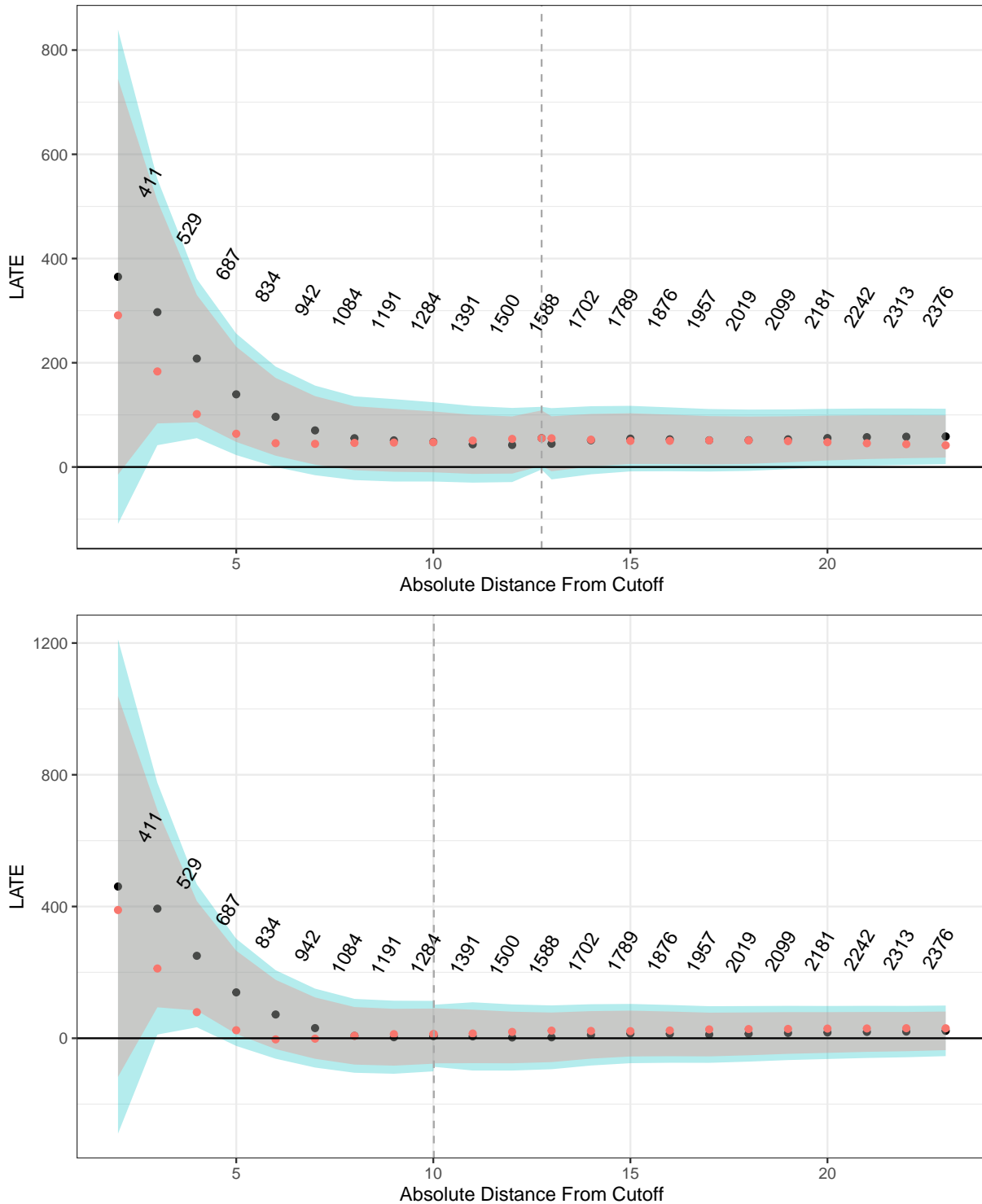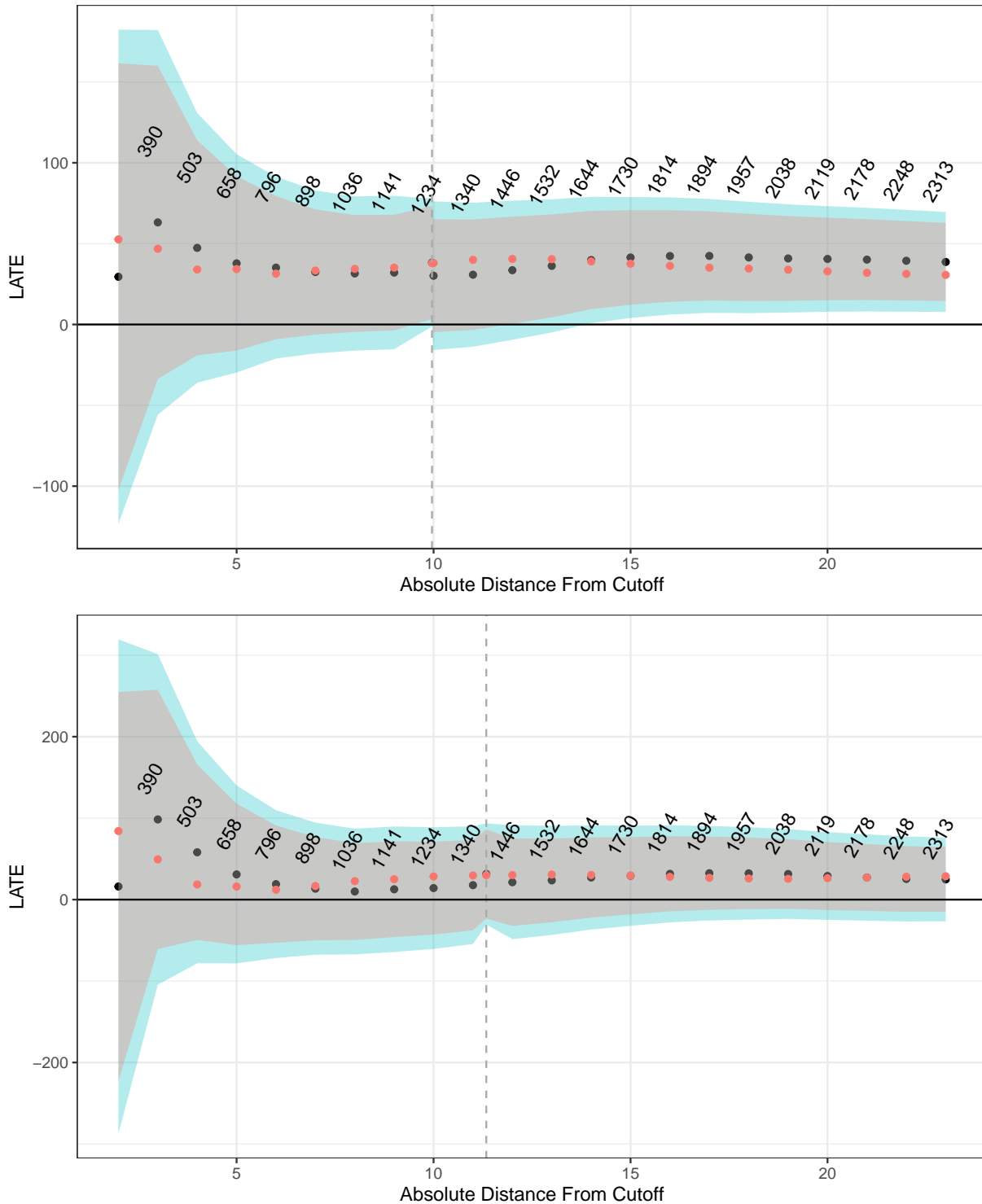
Figure 2: Age-standardized (top) and crude (bottom) LATE estimates for all-cause mortality are robust against choice of bandwidth. LATE is estimated as a local fuzzy RDD with a triangular kernel and second-order polynomial. Dashed line indicates optimal RBC bandwidth, numbers reflect number of observations within a given bandwidth, black dots and blue ribbon represents robust bias-corrected estimate and 95% CI, and pink dots and grey ribbon represents conventional estimate and 95% CI

Figure 3: Age-standardized (top) and crude(bottom) LATE estimates for heart disease mortality are relatively robust against choice of bandwidth. LATE is estimated as a local fuzzy RDD with a triangular kernel and second-order polynomial. Dashed line indicates optimal RBC bandwidth, numbers reflect number of observations within a given bandwidth, black dots and blue ribbon represents robust bias-corrected estimate and 95% CI, and pink dots and grey ribbon represents conventional estimate and 95% CI

Figure 4: Age-standardized (top) and crude (bottom) LATE estimates for cancer mortality are robust against choice of bandwidth. LATE is estimated as a local fuzzy RDD with a triangular kernel and second-order polynomial. Dashed line indicates optimal RBC bandwidth, numbers reflect number of observations within a given bandwidth, black dots and blue ribbon represents robust bias-corrected estimate and 95% CI, and pink dots and grey ribbon represents conventional estimate and 95% CI

Supplementary Material and Sensitivity Analyses
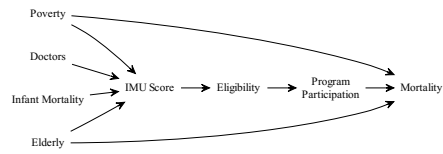
Figure 5: Directed Acyclic Graph depicting causal model of NHSC program and county-level mortality
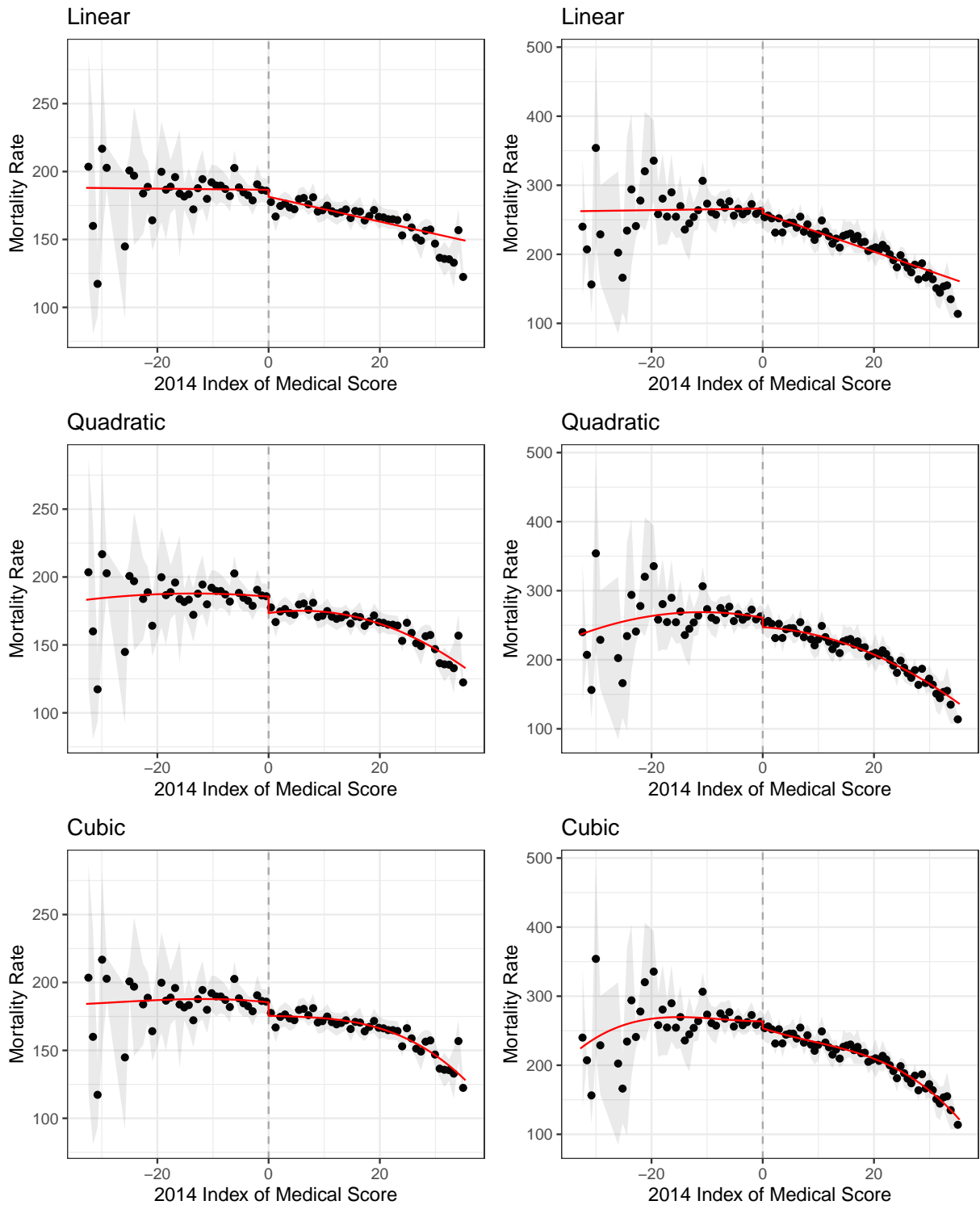
FRDD Specifications
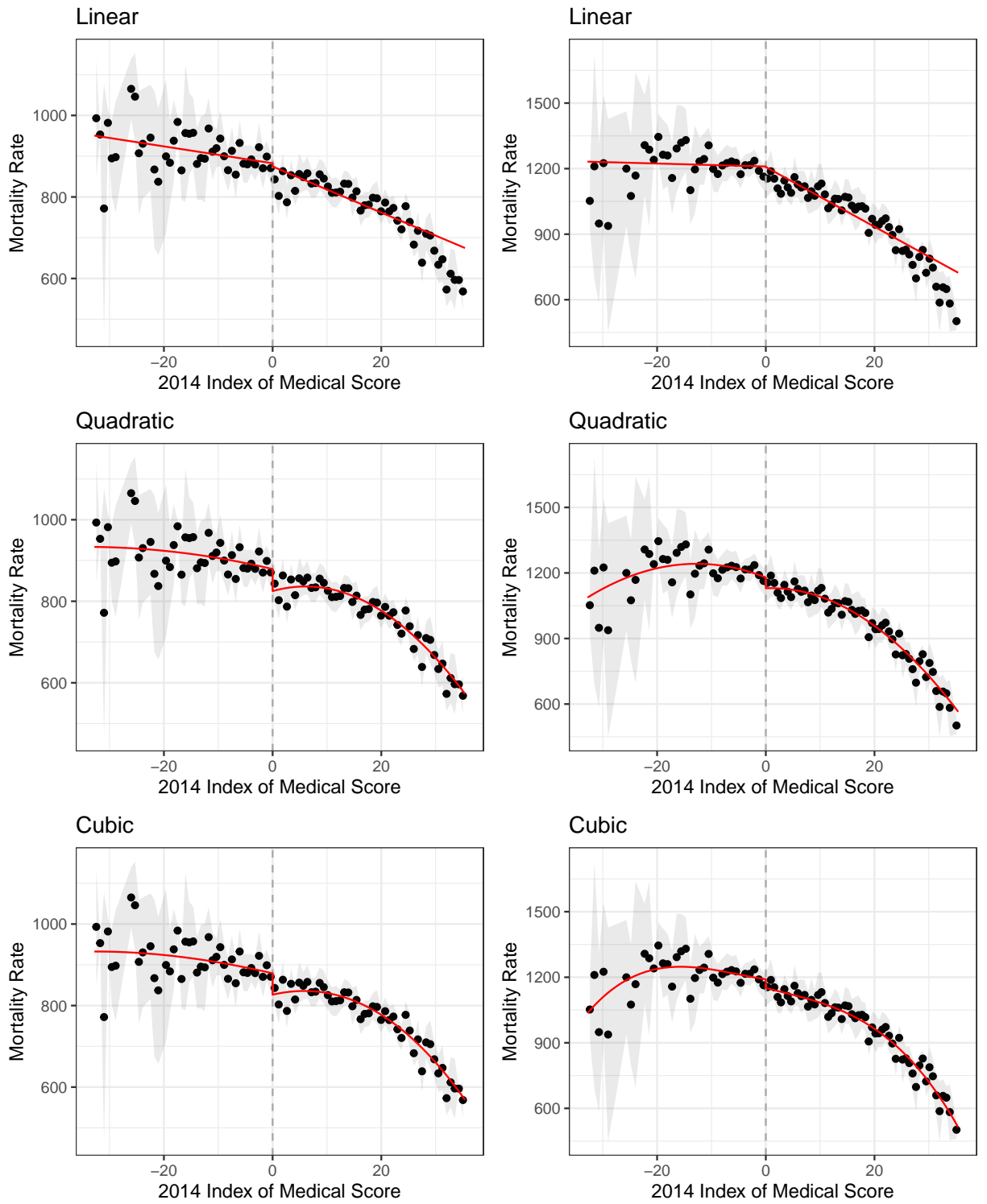
Lines represent predicted mortality rate for binned observations, and grey ribbons show 95% confidence intervals

Figure 6: Age-Standardized (Left) and Crude (Right) All Cause Mortality FRDD Fits
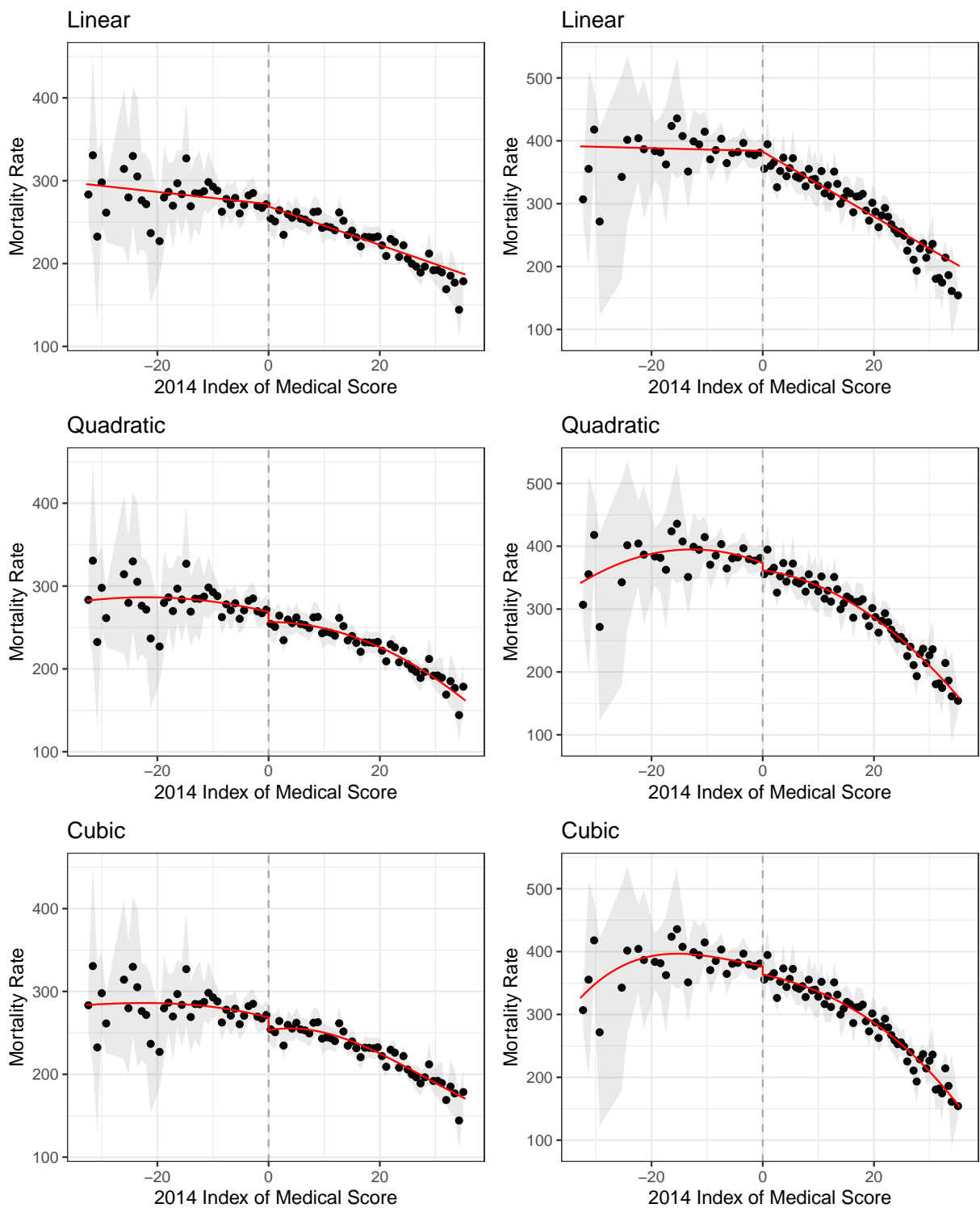
Lines represent predicted mortality rate for binned observations, and grey ribbons show 95% confidence intervals

Figure 7: Age-Standardized (Left) and Crude (Right) Heart Disease Mortality FRDD Fits

Lines represent predicted mortality rate for binned observations, and grey ribbons show 95% confidence intervals

Figure 8: Age-Standardized (Left) and Crude (Right) Cancer Mortality FRDD Fits

Table 5: Placebo Effects of NHSC Participation At False Cutoff Score of 56

Outcome	Estimate	Robust CI
Percent Poverty	-2.37	-5.76 - 1.02
Percent Over 65	2.77	-1.36 - 6.91
Median HH Income	-3333.39	-9933.05 - 3266.28
Unemployment	-0.01	-0.03 - 0.01
High School Graduation	0	-0.09 - 0.09
Uninsured	0.04	-0.01 - 0.08
Adult Obesity	0.01	-0.03 - 0.05
Smoking	0.06	-0.02 - 0.13
Physical Inactivity	0.02	-0.02 - 0.06
Percent White	-0.09	-0.28 - 0.10
Percent Black	0.05	-0.08 - 0.18
Percent Hispanic	0.05	-0.10 - 0.21
Percent Rural	-0.03	-0.27 - 0.20
1999 All Cause Mortality	**119.04**	**5.29 - 232.8**
2004 All Cause Mortality	41.44	-92.28 - 175.15
2009 All Cause Mortality	89.28	-37.14 - 215.69
1999 Heart Disease Mortality	29.42	-36.77 - 95.61
2004 Heart Disease Mortality	16.81	-45.45 - 79.07
2009 Heart Disease Mortality	-3.2	-65.86 - 59.46
1999 Cancer Mortality	32.86	-0.97 - 66.7
2004 Cancer Mortality	5.2	-36.35 - 46.74
2009 Cancer Mortality	**89.28**	**10.23 - 85.08**

Note:

Estimates represent tau FRD using a local regression with a triangular kernel

[a] Bold face denotes statistical significance ($p < 0.05$), CI refers to 95% Confidence Interval

[a] All mortality rates are age-standardized

2015 age−standardized All−Cause Mortality
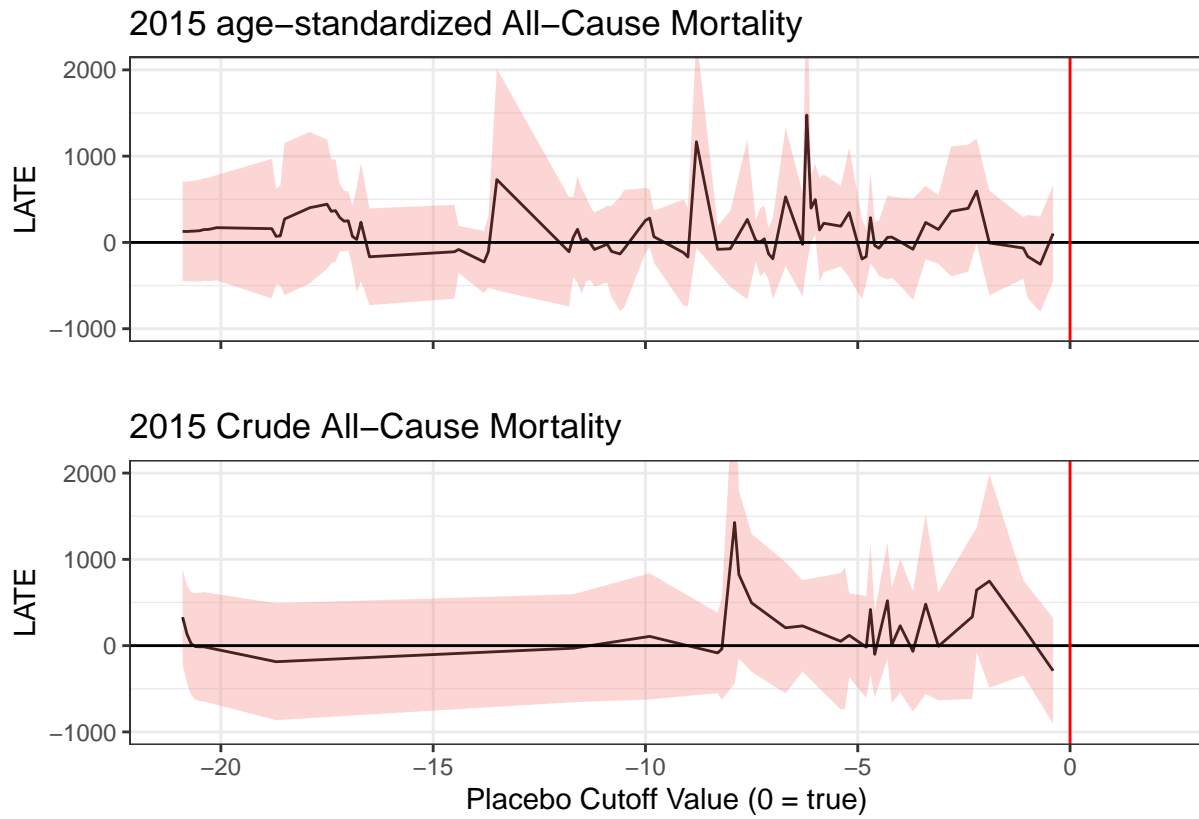

2015 Crude All−Cause Mortality

Figure 9: LATE at placebo cutoffs. Ribbon Shows RBC 95% Confidence Interval
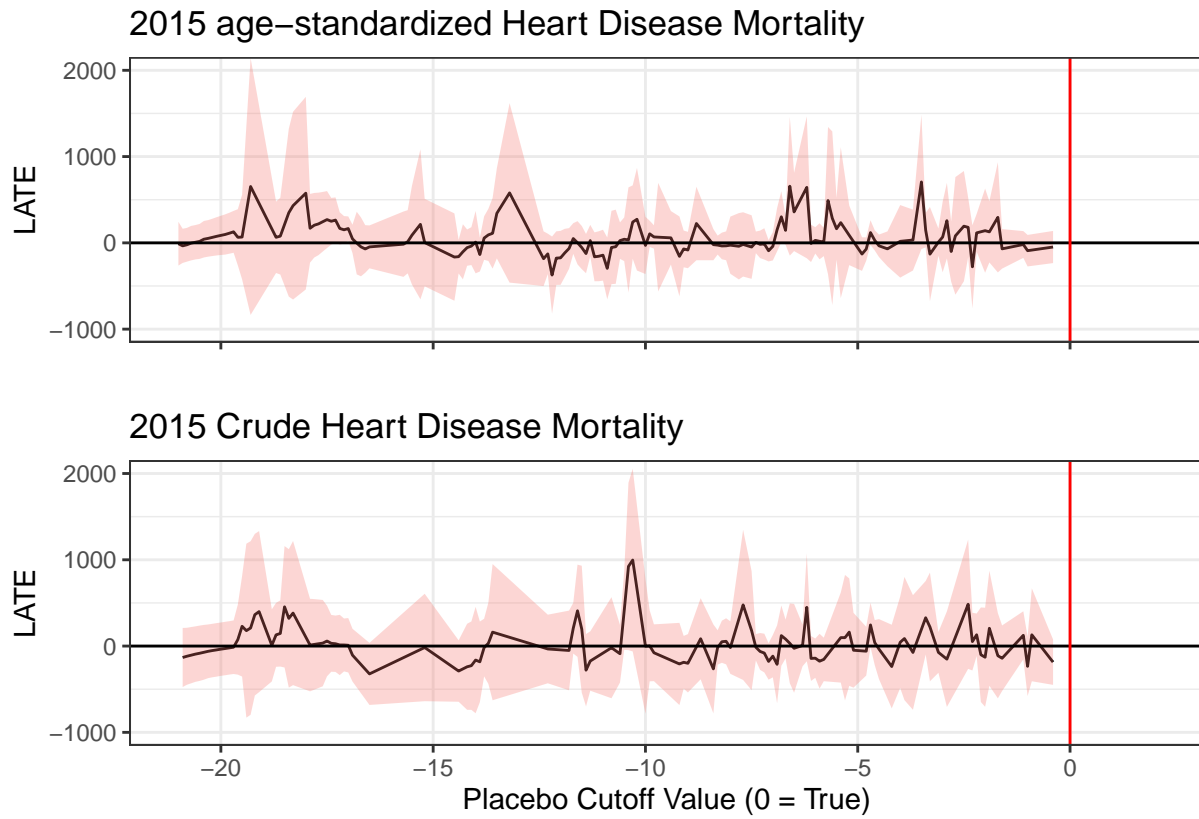
Figure 10: LATE at Placebo Thresholds. Ribbon Shows RBC 95% CI

2015 age–standardized Cancer Mortality

2015 age–standardized Cancer Mortality
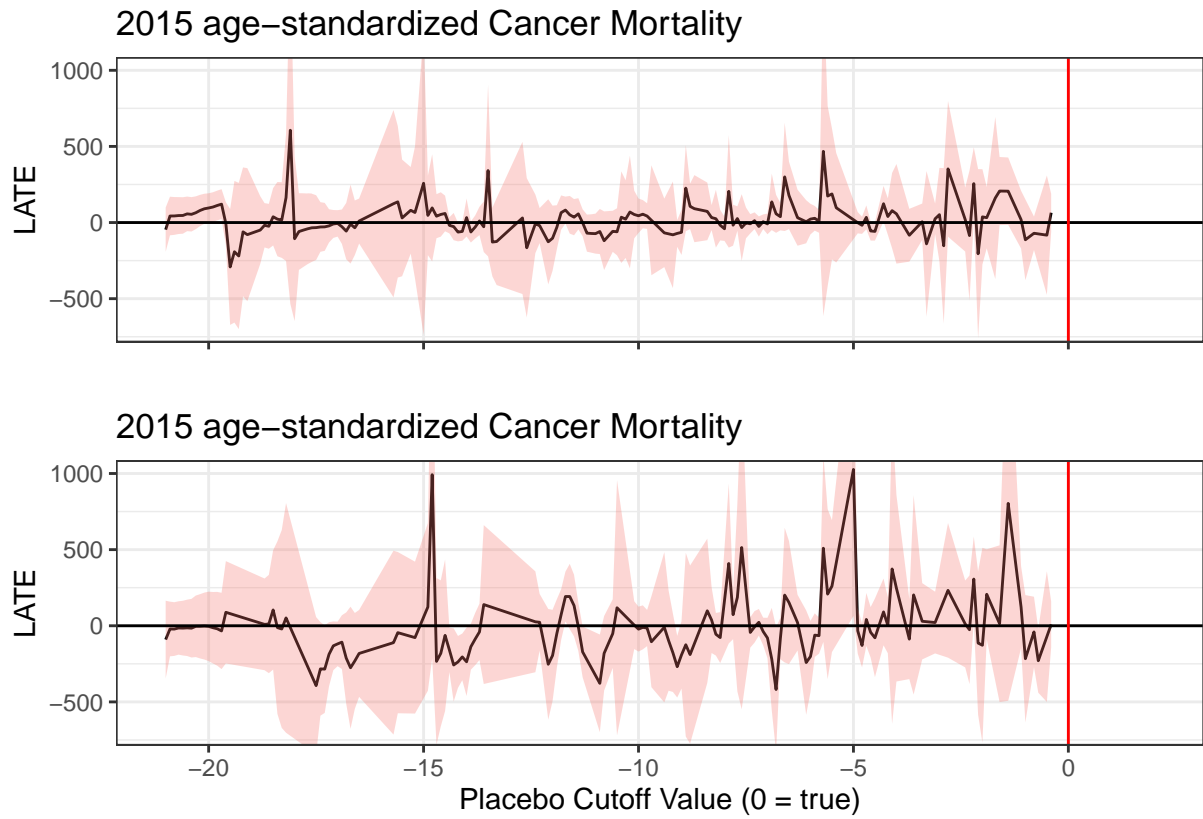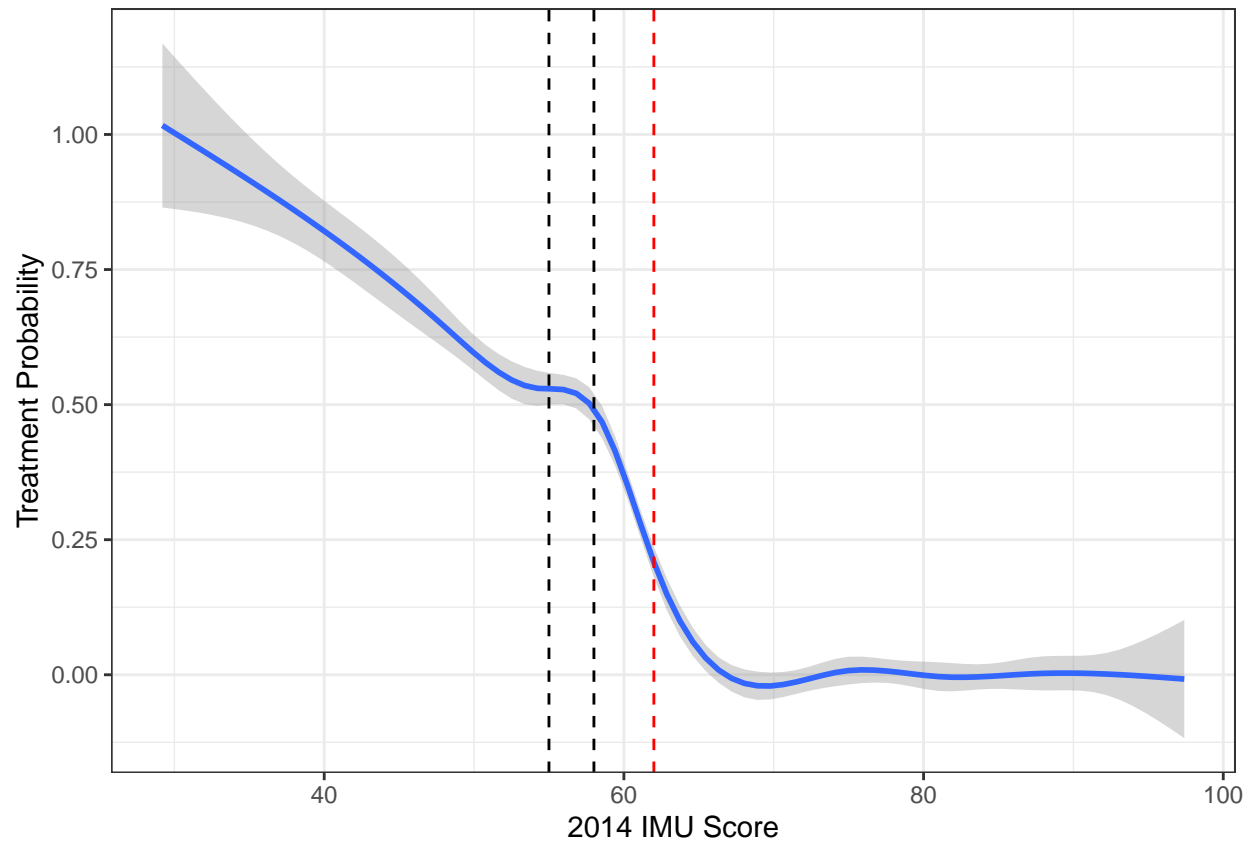
Figure 11: Treatment Probability by IMU Score. Dashed lines, from left to right, show scores of 55, 58, and 62 (the true cutoff)

Table 6: Placebo Effects of NHSC Participation At False Cutoff Score of 56

Outcome	Estimate	Robust CI
Percent Poverty	-14.77	-33.35 - 3.8
Percent Over 65	**21.13**	**0.81 - 41.44**
Median HH Income	-9843.63	-30940.5 - 11253.23
Unemployment	0.03	-0.02 - 0.08
High School Graduation	-0.53	-1.44 - 0.38
Uninsured	0.02	-0.09 - 0.12
Adult Obesity	0.14	-0.05 - 0.32
Smoking	0.02	-0.29 - 0.32
Physical Inactivity	0.05	-0.08 - 0.18
Percent White	-0.52	-1.25 - 0.21
Percent Black	0.67	-0.25 - 1.58
Percent Hispanic	-0.13	-0.42 - 0.16
Percent Rural	-0.27	-0.80 - 0.26

Note:

Estimates represent tau FRD using a local regression with a triangular kernel

[a] Bold face denotes statistical significance ($p < 0.05$), CI refers
to 95% Confidence Interval

References

Anderson, R. N., & Rosenberg, H. M. (1998). Age standardization of death rates: Implementation of the year 2000 standard. *National Vital Statistics Reports, 47*(3).

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2019). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *00*, 1–19. doi: 10.1093/ectj/utz022

Centers for Disease Control and Prevention. (1995). *CDC Wonder*. Retrieved from http://wonder.cdc. gov/

Fagan, E. B., Gibbons, C., Finnegan, S. C., Petterson, S., Peterson, L. E., Phillips Jr, R. L., & Bazemore, A. W. (2015). Family Medicine Graduate Proximity to Their Site of Training. *Family Medicine, 47*(2), 124–130.

Heisler, E. (2017). *The National Health Service Corps* (No. CRS Report R44970). Washington, DC: Congressional Research Service.

Holmes, G. M. (2005). Increasing physician supply in medically underserved areas. *Labour Economics*, 12(5), 697–725. doi: 10.1016/j.labeco.2004.02.003

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A

density test. *Journal of Econometrics, 142*(2), 698–714.

National Health Service Corps Scholarship Program. , 42 U.S.C. § 2541 (2003).

Parlier, A. B., Galvin, S. L., Thach, S., Kruidenier, D., & Fagan, E. B. (2018). The Road to Rural Primary Care: A Narrative Review of Factors That Help Develop, Recruit, and Retain Rural Primary Care Physicians. *Academic Medicine, 93*(1), 130–140. doi: 10.1097/ACM.0000000000001839

Pathman, D. E., Konrad, T. R., Dann, R., & Koch, G. (2004). Retention of Primary Care Physicians in Rural Health Professional Shortage Areas. *American Journal of Public Health, 94*(10), 1723–1729. doi: 10.2105/AJPH.94.10.1723

Rabinowitz, H. K., Diamond, J. J., Markham, F. W., & Paynter, N. P. (2001). Critical Factors for Designing Programs to Increase the Supply and Retention of Rural Primary Care Physicians. *JAMA, 286*(9), 1041–1048. doi: 10.1001/jama.286.9.1041

Shi, L., Macinko, J., Starfield, B., Politzer, R., Wulu, J., & Xu, J. (2005). Primary Care, Social Inequalities, and All-Cause, Heart Disease, and Cancer Mortality in US Counties, 1990. *American Journal of Public Health, 95*(4), 674–680. doi: 10.2105/AJPH.2003.031716

Starfield, B., Shi, L., & Macinko, J. (2005). Contribution of Primary Care to Health Systems and Health. *The Milbank Quarterly, 83*(3), 457–502. doi: 10.1111/j.1468-0009.2005.00409.x

Torrieri, N. (2014). American Community Survey Design and Methodology (January 2014). Available online at: http://Www.census.gov/Acs/Www/Downloads/Survey_methodology/Acs_design_methodology_report_2014.pdf.

U.S. Census Bureau; American Community Survey. (2014). 2014 American Community Survey 1-Year Estimates, Table S0103; generated by Sarah Van Alsten; using American FactFinder; Retrieved from http://factfinder.census.gov

US Department of Health and Human Services. (2014a). Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System Survey Questionnaire. Atlanta: CDC.

US Department of Health and Human Services. (2014b). Health Resources and Services Administration. Area Health Resources Files (AHRF).

U.S. Department of Health and Human Services. (2015, November). Shortage Designation Management System (SDMS): *Manual for Policies and Procedures.* Retrieved from http://contentmanager.med.uvm.edu/docs/sdms_manual_/ahec-documents/sdms_manual_.pdf?sfvrsn=10