University of Arkansas, Fayetteville
ScholarWorks@UARK

Theses and Dissertations

1-2019

Diversity and Health: Three Essays Exploring Social Context and Outcomes

Aaron Novotny
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

Part of the Behavioral Economics Commons, Civic and Community Engagement Commons, Community-Based Research Commons, Food Security Commons, Politics and Social Change Commons, and the Race and Ethnicity Commons

Recommended Citation

Novotny, Aaron, "Diversity and Health: Three Essays Exploring Social Context and Outcomes" (2019). *Theses and Dissertations*. 3218.
https://scholarworks.uark.edu/etd/3218

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact ccmiddle@uark.edu.

Diversity and Health: Three Essays Exploring Social Context and Outcomes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Economics

by

Aaron Novotny
University of Arkansas
Bachelor of Arts in French, 2009
University of Arkansas at Little Rock
Masters of Science in Mathematics, 2012
University of North Dakota
Masters of Arts in Economics, 2017

May 2019
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council:

Andrea Civelli, Ph.D.
Committee Chair

Michael Thomsen, Ph.D.
Committee Member

Arya Gaduh, Ph.D.
Committee Member

Abstract

As the world becomes more diverse and more integrated, examining how racial, political, and food diversity influences individuals' consumption, behaviors, and health becomes more paramount than ever before. The United States grows more racially diverse with large racial and ethnic shifts on the horizon regarding the proportion of the population. With the U.S. population expected to become more diverse, individuals' political affiliation becoming more prevalent to personal identity, and food security becoming more problematic; we examine how racial, political, and food diversity influences individuals' consumption and preferences with the intent to understand what changes in health and preferences may occur.

These essays contribute to the literature in a novel way by understanding how the local racial, political, and food environments impact individual consumption and behavior choices. Additionally, these papers novel approaches yield strong evidence that these different measures of diversity play an important and larger role in individuals' daily lives then realized. We utilize a cutting edge propensity score matching technique to understand the impact of the food access program WIC (Women, Infants, and Children Supplemental Health Program) in "WIC Participation and Relative Quality of Household Food Purchases: Evidence from FoodAPS" essay. In this essay, we show the health benefits participants in WIC receive when participating in the program. Second, we leverage a special type of survey called a List Experiment to understand respondents' social desirability bias given their political environment in the essay titled "Social Desirability Bias and Polling Errors in the 2016 Presidential Election." In this essay, we show that people respond differently to the question of whether they support a particular presidential candidate when they are given the opportunity to directly or indirectly express this support. Lastly, we explore how local racial diversity directly impacts an individual's healthy food consumption in the essay titled "Diversity and Health: Exploring Local Racial Diversity's Impact on Health Through Food Consumption". This final essay attempts to ascertain the associative effects of racial diversity on an individual's healthy

food consumption. This result shows the dramatic impact racial diversity has on healthy food consumption by improving individuals' food intake, but concentrated on certain races. Overall, this dissertation shows the dramatic influence social and environmental context has on individual outcomes via health and preferences.

Contents

List of Published Papers

1. Chapter 2: Social Desirability Bias and Polling Errors in the 2016 Presidential Election. Brownback and Novotny, 2017. Journal of Behavioral and Experimental Economics, Volume 74, 2018, Pages 38-56, ISSN 2214-8043, https://doi.org/10.1016/j.socec.2018.03.001.

2. Chapter 3: WIC Participation and Relative Quality of Household Food Purchases: Evidence from FoodAPS. Fang, Thomsen, Nayga Jr., Novotny, 2017. USDA grant no. 59-5000-5-0115 to the National Bureau of Economic Research entitled, Using FoodAPS for Research in Diet, Health, Nutrition, and Food Security.? Forthcoming April 2019 in Southern Economic Journal.

Introduction

As the world becomes more diverse and more integrated, examining how racial, political, and food diversity influences individuals' consumption, behaviors, and health becomes more paramount than ever before. The United States grows more racially diverse with large racial and ethnic shifts on the horizon regarding the proportion of the population. With the U.S. population expected to become more diverse, individuals' political affiliation becoming more prevalent to personal identity, and food security becoming more problematic; we examine how racial, political, and food diversity influences individuals' consumption and preferences with the intent to understand what changes in health and preferences may occur.

These essays contribute to the literature in a novel way by understanding how the local racial, political, and food environments impact individual consumption and behavior choices. Additionally, these papers novel approaches yield strong evidence that these different measures of diversity play an important and larger role in individuals' daily lives then realized. We explore how local racial diversity is related healthy food consumption in the essay titled "Diversity and Health: Exploring Local Racial Diversity's Impact on Health Through Food Consumption." This result shows the relationship between racial diversity and healthy food consumption is complex because numerous other factors contribute to healthy food outcomes such as income, race, population, and education. We use American Community Life Survey data to construct a localized diversity measure (by appropriating the Ethnoliguistic Fractionalization (ELF) measure pioneered in the development literature) about an EFNEP participant's Zip code to understand how variation in a participant's local diversity influences the healthfulness of food consumption by individuals. We find that individuals of White and Black races benefit from increased diversity by showing gains in healthy food consumption as measured by the Healthy Eating Index (HEI) score, but only when persons have less than a four year degree and live in an urban environment.

Second, we leverage a special type of survey called a List Experiment to understand respondents' social desirability bias given their political environment in the essay titled "Social Desirability Bias and Polling Errors in the 2016 Presidential Election." In this essay, we show that people respond differently to the question of whether they support a particular presidential candidate when they are given the opportunity to directly or indirectly express this support. Social scientists have observed that socially desirable responding (SDR) often biases unincentivized surveys. Nonetheless, media, campaigns, and markets all employ unincentivized polls to make predictions about electoral outcomes. During the 2016 presidential campaign, we conducted three list experiments to test the effect SDR has on polls of agreement with presidential candidates. We elicit a subject's agreement with either Hillary Clinton or Donald Trump using explicit questioning or an implicit elicitation that allows subjects to conceal their individual responses. We find evidence that explicit polling overstates agreement with Clinton relative to Trump. Subgroup analysis by party identification shows that SDR significantly diminishes explicit statements of agreement with the opposing party's candidate driven largely by Democrats who are significantly less likely to explicitly state agreement with Trump. We measure economic policy preferences and find no evidence that ideological agreement drives SDR. We find suggestive evidence that local voting patterns predict SDR. This paper was coauthored with Andy Brownback from the Department of Economics at the University of Arkansas.

Lastly, We utilize a cutting edge propensity score matching technique to understand the impact of the food access program WIC (Women, Infants, and Children Supplemental Health Program) in the "WIC Participation and Relative Quality of Household Food Purchases: Evidence from FoodAPS" essay. In this essay, we show the health benefits participants in WIC receive when participating in the program. We examine the effect of the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on the quality of household food purchases using the National Household Food Acquisition and Purchase Survey (FoodAPS) and propensity score matching technique. A healthy purchasing index (HPI) is used to measure nutritional quality of household food purchases. WIC foods explain the improvement in quality of food purchases, not self-selection of more nutrition-conscious households into the program. The improvement in purchase quality was driven entirely by WIC participating households who redeemed WIC foods during the interview week. There was no significant difference between WIC-participants who did not redeem WIC foods and eligible non-participants. In this sample, there is no evidence that lack of access to clinics has adverse effects on participation nor is there evidence that HPI depends on supermarket access. A supervised machine learning process supports our main conclusion on the importance of WIC foods.This paper was written with Di Fang, Michael R. Thomsen, and Rodolfo M. Nayga, Jr.. Di Fang is Assistant Professor, Michael R. Thomsen is Professor, and Rodolfo M. Nayga Jr. is Distinguished Professor and Tyson Chair in Food Policy Economics all in the Department of Agricultural Economics and Agribusiness, 217 Agriculture Building, University of Arkansas, Fayetteville, Arkansas 72701, USA. Address correspondence to Di Fang, difang@uark.edu. This research was supported by USDA grant no. 59-5000-5-0115 to the National Bureau of Economic Research, entitled, "Using FoodAPS for Research in Diet, Health, Nutrition, and Food Security." This version was cleared by Xingyou Zhang, Ph.D..

Overall, this dissertation shows the dramatic influence social and environmental context has on individual outcomes via health and preferences whether we examine an individual's participation in a food assistance program such as WIC, recognizing the association between racial diversity and healthy food consumption, or the influence of peer effects when answering a political question displaying support for a particular presidential candidate.

Chapter 1: Diversity and Health: Exploring Local Racial Diversity's Impact on Health Through Food Consumption

Abstract

We use American Community Life Survey data to construct a localized diversity measure (by appropriating the Ethnoliguistic Fractionalization (ELF) measure pioneered in the development literature) about an EFNEP participant's Zip code to understand how variation in a participant's local diversity influences the healthfulness of food consumption by individuals. We find associative gains in healthy eating consumption from increases in diversity for some individuals, but not all. We observe associative gains for only white and mixed race individuals in urban environments with some education beyond high school, but the effect is not present for individuals with a four year college degree.

Introduction

Diversity can illicit many different reactions from people because diversity can either be viewed as a good or a bad depending on the context and the individual. It can lead to conflict ranging from differing preferences, racism, and prejudice where we observe suboptimal outcomes or oppression of ethnic groups (Alesina and Ferrara, 2005). On the other hand, diversity can bring a variety of skills, abilities, and experiences which can lead to increases in creativity and growth (Alesina and Ferrara, 2005). Obviously, tradeoffs exist as diversity increases but we wish to examine how diversity influences health and what potential tradeoffs a more diverse populous may face?

Our study examines how changes in the local racial composition of an area can be associated with health changes due to changes in food consumption. Ideally, more heterogeneity among different races (and cultures) will spur retailers to carry more diverse foods. In turn, all individuals that visit this retailer will gain additional food choices which could then lead to a healthy diversification of food consumption which leads to healthier food consumption. We explore this simple mechanism by examining how racial heterogeneity may impact healthy food consumption.

As the world becomes more diverse and more integrated, examining how racial diversity influences individuals' consumption becomes more paramount than ever before. The United States grows more racially diverse with large racial and ethnic shifts on the horizon regarding the proportion of the population. With the non-hispanic White population's proportion of the population expected to shrink and the U.S. population expected to grow (Vespa et al., 2018), we examine how diversity influences individuals' food consumption with the intent to understand what changes in health may occur given an individual's local racial environment.

There is much work that suggests that racial and ethnic diversity can adversely affect individuals but many of these analyses are performed at the macroeconmic level. Some examples in this context are public goods provision, governance, and civil conflicts where we see macroeconomic outcomes given changes in (typically) ethnic diversity or ethnic concentration of a region or country. Additionally, there is limited work that shows positive effects from diversity at the macro or micro level.

This paper contributes to the diversity literature in a novel way by understanding how the local racial environment impacts individual consumption choices. Additionally, this paper's novel instrumentation approach yields strong evidence that diversity plays an important and larger role in individuals' daily lives then realized. Lastly, we measure actual health changes at the individual level due to local diversity. These results should help to yield better information for future policy regarding racial mixing decisions, implementation strategies of food assistance programs, and understanding how supply responds as the U.S. becomes more diverse.

Overall, this paper asks the simple (but difficult to answer) question, “How does local diversity influence ‘individual’ consumption choices?” We address this question by utilizing individual nutrition and food consumption measures for individuals that participate in the Expanded Food and Nutrition Education Program (EFNEP) across the United States. Next, we construct a localized Ethnolinguistic Fractionalization (ELF) measure about each individual at the ZIP code level to create each participants’ local racial environment. We use this ELF to determine whether the local racial environment is associated with healthy food consumption choices. A caveat to this approach is that individuals may move to (or out of) an area because of the racial makeup (i.e. individuals may want to avoid people of a different race and move closer to people of the same race) which suggests endogeneity in our analysis. This endogeneity limits us in two ways: 1) we make no causal claims in this paper since (as the previous literature has shown) it proves quite difficult to find a valid instrument and 2) these estimates will be biased.

Our hypothesis suggests that more diversity leads to broader food selection (in the local environment when controlling for income, size, etc..) which leads to healthier food consumption. So, the mechanism through which diversity impacts health is through added choice. Any endogeneity resulting from an individual’s ability to sort into more homogenous environment would lead to less food choices for individuals (i.e. the market would provide less diverse foods); therefore, we would observe no healthier food consumption suggesting this endogeneity attenuates our results. Moreover, if individuals self sort to more diverse locations and do not change their dietary habits, then this would also attenuate our results. Hence, we consider these estimates to be less in magnitude.

This paper is divided into five additional sections. Section II examines relevant literature addressing how diversity affects individuals and areas. Additionally, the literature examines the positive impact of diversity on market choice. We share literature that measures health using the HEI score, birth outcomes, or food security. This literature typically examines efficacy of welfare programs such as SNAP or WIC. Section III will describe the data and its limitations, data construction, and develops a simplistic model examining the relationship between healthy eating outcomes and local diversity. Section IV will present standard OLS results homogenized across different subsamples. Section V will be the conclusion and include possible future research ideas.

Literature

There has been much work that shows more diversity typically means less public goods provided on average especially regarding minority groups (Alesina and Ferrara, 2005; Alesina et al., 1999). In these cases, the majority racial or ethnic group does not wish to provide tax revenue or public donations which benefit a different racial (ethnic) group then their own. In this instance, the majority group has different preferences on what types of public goods are purchased with tax revenues than the minority groups and therefore this majority wishes to provide less tax revenue for public goods not preferred by this majority group. These instances obviously can be related to issues of racial segregation and racial animus. Moreover, Miguel and Gugerty (2005) show how diversity lessens contributions to school because there is no single social group which to pressure individuals to make contributions for public goods such as schools or water wells. Moreover, ethnic diversity can be attributed to more civil conflict, less growth and investment, and worse governance (Miguel and Gugerty, 2005; Esteban and Ray, 2008; Montalvo and Reynal-Querol, 2005; Easterly and Levine, 1997; Baldwin and Huber, 2010; La Porta et al., 1999) suggesting that as diversity increases we see that individuals become worse off in the aggregate. These papers can be summarized into three facets: 1) racial animus preventing contribution to public works that would increase overall welfare but not a particular group's welfare, 2) increased diversity leading to a lack of salience for goods that don't match preferences of a majority ethnic group, or 3) diversity leading to a lack of social enforcement among peer groups to ensure contribution to public works.

While we observe many adverse effects due to diversity, there are some positive impacts regarding market outcomes. Racial diversity leads to more diverse grocery purchases among individuals not in lower incomes (Blisard et al., 2002; Kinsey, 1994; Jekanowski et al., 2000). While these studies simply focus on differentiated brands or certain foods, the evidence suggests that a mechanism to achieve more diverse food consumption would be to offer more diverse food choices through the market which results from having different groups of individuals with different preferences. “A well off and ethnically diverse nation will demand variety.”(Blisard et al., 2002). The question remains: do more choices facilitate healthier individual outcomes?

At the individual health level, much of the literature focuses on food purchase patterns as it relates to characteristics of the household (Lo et al., 2012), how food insecurity affects health outcomes (Gundersen et al., 2011), or how participation in welfare programs such as SNAP or WIC improves food outcomes, food security, or birth outcomes (Bitler and Currie, 2005; Yen, 2010; Kreider et al., 2018; Bitler, 2014; Gundersen et al., 2017). Other studies examine food security as it pertains to the environment such as is an individual impacted by not living close to a grocery store (Ver Ploeg, 2010). These studies focus on environmental factors that dictate low levels of consumption versus high levels of consumption and examine whether households have access to healthy foods given certain environmental factors. Moreover, Fang et al. (2018) examine the efficacy of the WIC program utilizing a Healthy Purchase Index (HPI) based on the Healthy Eating Index score. They find that individuals that participate in WIC have healthier HPI scores than similar individuals that do not participate outlining gains in health through additional food purchases.

There has been limited work suggesting that diversity can improve public goods outcomes in experimental settings (Santos et al., 2008), but this result is undercut by favoritism of similar identity individuals in experimental games when identity is activated (Chen and Li, 2009). These works exemplify the complexity that diversity brings. Who does diversity impact and when? Additionally, why and how does diversity reduce individuals’ outcomes? Will diversity make everyone better off? Diversity presents to be a challenge worth understanding as the United States becomes more diverse.

Data and Model

This paper relies on multiple principal data sources: the Expanded Food and Nutrition Education Program (EFNEP), the American Community Life Survey (ACS), and OpenDataSoft's military base locations. The EFNEP data yields our controls and the outcome variable i.e. the Healthy Eating Index (HEI) score. We utilize the ACS data to construct local diversity measures at the ZIP code level which will constitute our variable of interest. Additionally, the ACS data yields more controls such as local median income levels, number of individuals that have health insurance, number of persons with military service, and food access measures.

Expanded Food and Nutrition Education Program

The EFNEP program began in 1969 with the intent of helping low socioeconomic individuals change their nutrition and physical activity behaviors to increase community welfare (USDA/NIFA 2019). EFNEP utilizes local "peer educators" which are trained by locally based staff or university personnel to facilitate a "nutrition education approach" with intent to treat diet quality, physical activity, food resource management, food safety, and food security. This program claims to reach half a million low-income individuals each year and is funded through the USDA. We will take the dataset set recorded in 2013/2014.

Aside from the typical control variables such as race, income, municipality, etc.., this dataset takes an "entry" HEI score before participants begin the weeks long training program and then records an "exit" HEI score upon completion of EFNEP. While EFNEP looks to measure efficacy of its federal program by examining the change in HEI, we are most interested in the entry score as this gives us a measure of an individual's health prior to commencement.

The EFNEP dataset includes over 120,000 individuals when we drop missing observations, but

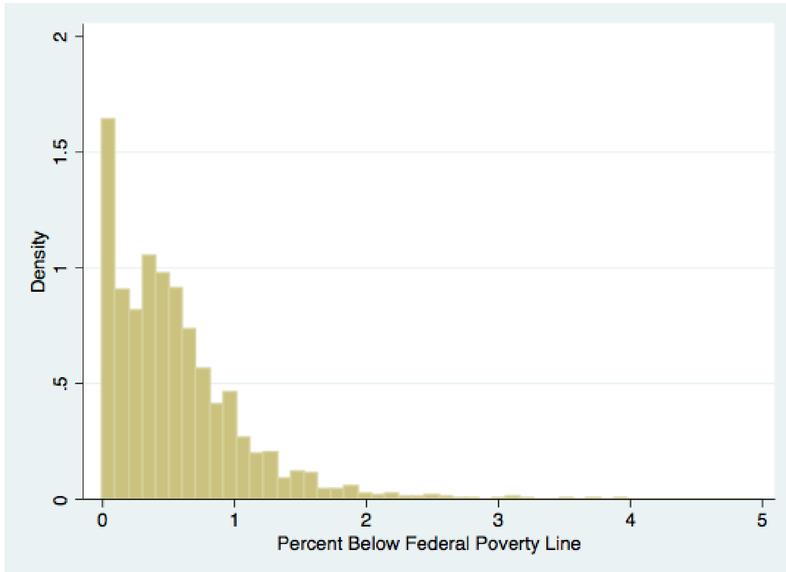
Table 1: EFNEP Participants Skew Majority Female in Data

Male and Female Participation in EFNEP			
Sex	Frequency	Percent	Cum.
Male	14,322	13.05	13.05
Female	95,456	86.95	100

there are some caveats which need to be addressed. First, we have 78,243 observations identified as female and 11,433 identified male participants. This dataset skews female and more represented by the Black race than the proportion in the U.S.. As the intent of EFNEP is to help nutrition education and security, this skewness comes as no surprise given women typically constitute the main food preparer in the home and more black persons are at or below poverty. There could be many reasons we observe such a discrepancy between male and female participation; but when running our models and isolating only male or female, we observe very similar outcomes associated with diversity except in a handful of cases which will be pointed out. Additionally, we hypothesize that more healthful purchases occur because of market variety or lessened racial animus due to increased diversity which should not affect males or females differently. Ergo, we do not make any attempt to understand why more women participate in the program. However, we do outline cases where the outcomes differ (in significance or magnitude) from increased diversity dependent on the sex.

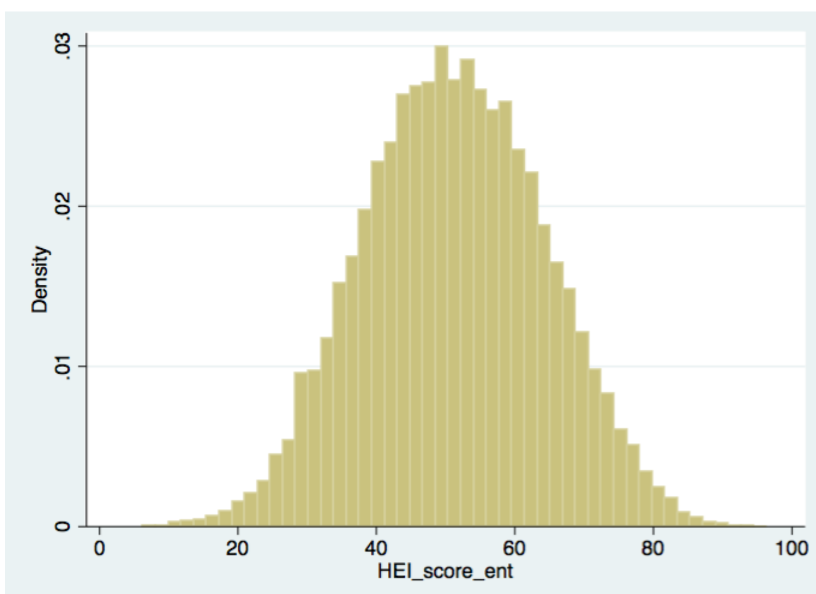
We recognize that participants are inherently individuals concerned about weight, health, or food security issues and they are not financially well off. Each suggests another possible source of endogeneity. We do not argue the selection issue as it pertains to food security, but suggest that this group of individuals at or below the poverty line benefit most from increased diversity. With 40 million individuals living in poverty (U.S. Census 2017) and many more living just above poverty, we believe these results reflect many average individual responses to diversity. Moreover, wealthy individuals have the resources to find healthy food options in any environment (Allcott et al., 2017), so we do not claim that wealthy individuals are influenced by the surrounding diversity through the market mechanism. As it pertains to health or weight, individuals with these motivations may be

Figure 1: Participants at or Below Poverty



more sensitive to food options, but one could argue no more than the 39.8% of the U.S. population that is obese (Hales et al., 2017) or participating in the $169 billion dollar weight loss industry (Reuters 2018).

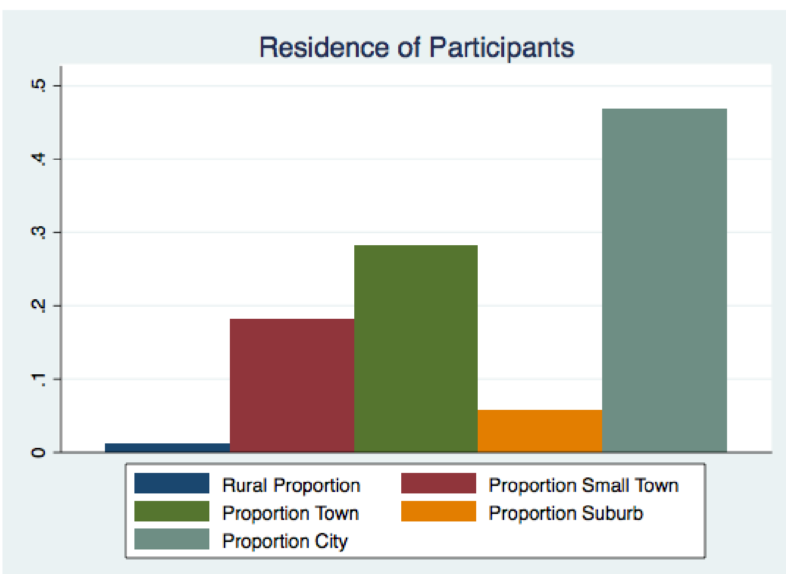
Figure 2: Dispersion of HEI Among Participants



Our participants show a normal distribution of entry HEI scores showing great diversity among all the participants in EFNEP. This stylized fact coupled with the inherent low SES of participants shows that healthy eating outcomes are being achieved in spite of income suggesting there exist mechanisms which individuals use to consume healthier food options such as food assistance programs or diverse food attainment. One might claim that it is the municipality of the participant i.e. individuals residing in an urban environment would have healthier outcomes versus small town residents, but this is not true. We observe a well defined bell shaped curve for all residential settings suggesting that participant's municipality suggesting even distribution accross all locales. A T-test does show differences in healthy food consumption on average , but these differences do not exceed 1.5 HEI points, so we do not attribute large scale healthy food consumption pattern effects to municipality.

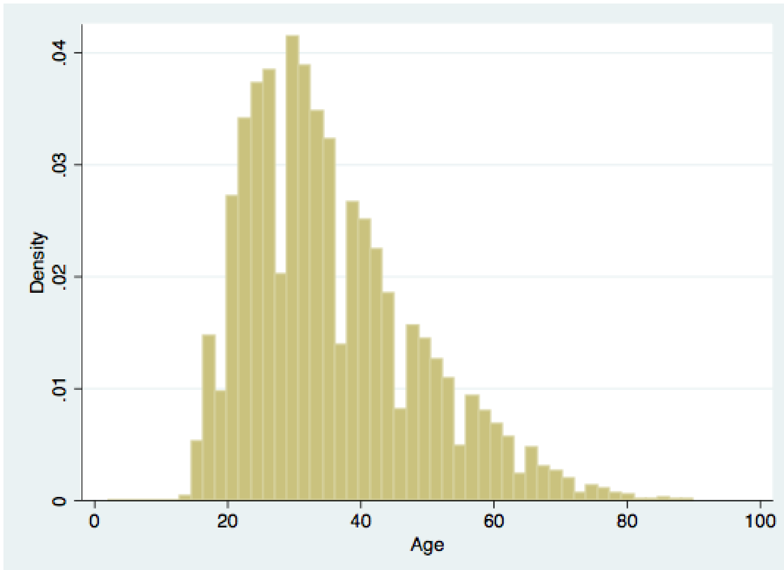
The EFNEP data categorizes participants into five municipalities: 1) city defined as having more than 50,000 residents, 2) Suburb defined as having 25,000-50,000 residents, 3) town defined as having 10,000-25,000 residents, 4) small town defined as having 1,000-10,000 residents, and 5) farm (or rural) defined as having less than 1,000 residents. 49% of participants reside in the city setting. We observe few individuals in the suburban setting (5%) and 28% live in a town setting. Another 18% of participants live in the small town and 1% live in a farm setting.

Figure 3: Municipality of Residents



Given the mission of EFNEP, we would suspect that we might observe certain age groups or low education among participants. We observe no age concentration among participants. In fact, we see a slightly skewed bell curve which makes sense considering this program is not intended for school age children.

Figure 4: Age of EFNEP Participants



However, participants are lower educated where we observe 85% of participants with less than a 4 year college degree. More than half, 56%, have completed high school and have some college. Education can contribute to healthy eating in multiple ways such as through increased income, managing food security issues, and making smarter consumption choices; so, this variable has important implications when parsing ELF's effects.

Lastly, program participants enter their race identification. Participants are asked what race they identify where possible responses include white, black, pacific, native, asian, or a combination of two or more. For simplicity, any individual that responds as a racial combination is counted as a mixed race person. We observe 48% of participants are white while 28% identify as black and 16% identify as mixed.

The EFNEP dataset shows to be a rich dataset, but it does come with limitations such as education, income, and residential settings. We acknowledge these limitations and limit interpretations to lower income, lower educated persons. Additionally, municipality will play a role in the analysis

Figure 5: Education Level of EFNEP Participants

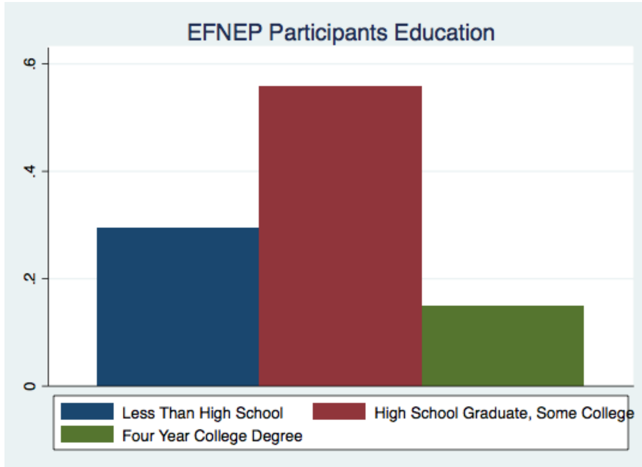
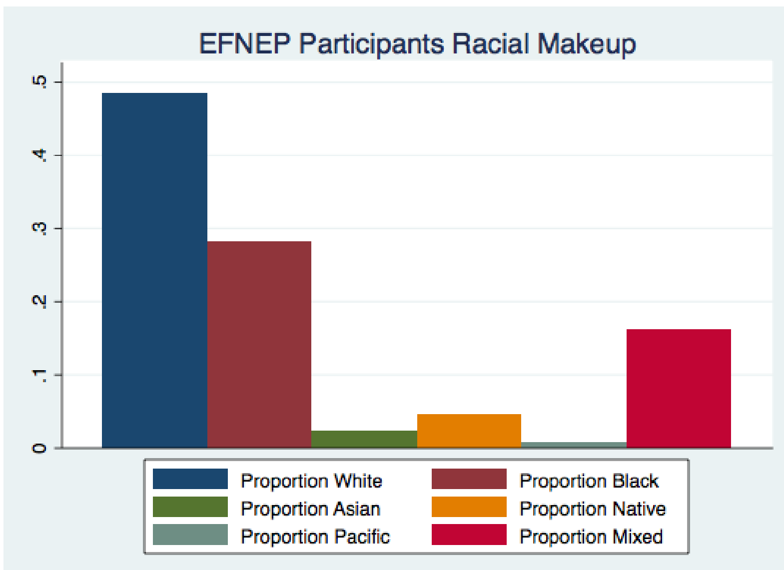


Figure 6: Race of EFNEP Participants



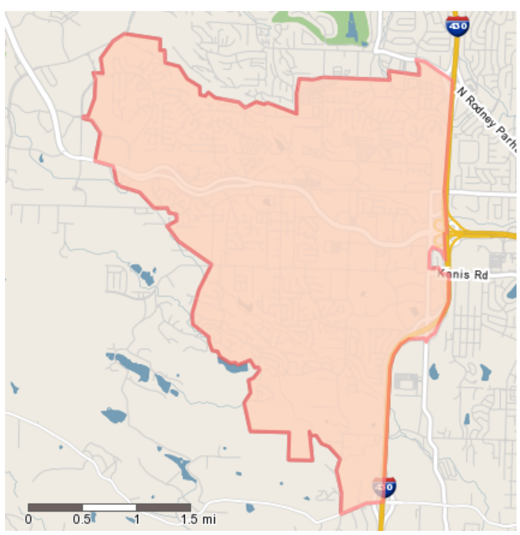
moving forward.

American Community Life Survey 2009-2013

We connect the EFNEP dataset to the ACS dataset by each participant's 5-digit ZIP code. The ACS dataset comes from the U.S. Census Bureau detailing population, demographics, housing, health insurance, military service, and food desert information. We use the ACS 5-year estimates ending in 2013 relating to race, health insurance, military service, low food access, and total population. The ACS dataset asks a similar question regarding what race an individual is with an exception that the ACS gives an option of "other" race.

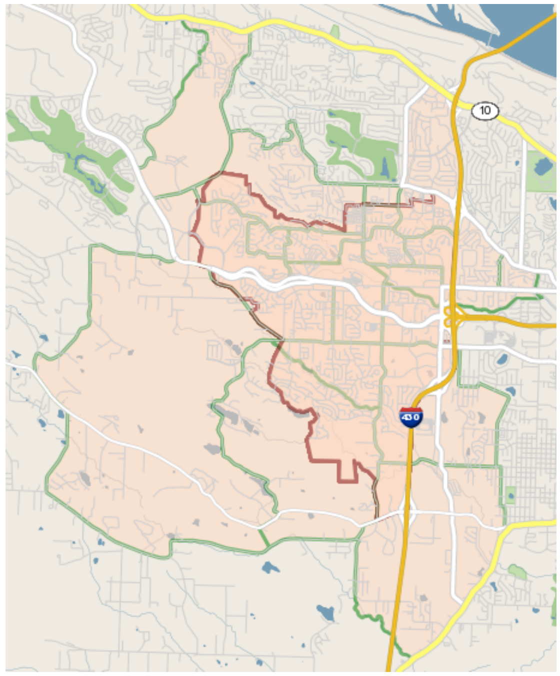
One issue with using these estimates is that the ACS does not provide data at the ZIP code level. Instead, the ACS provides block group racial census figures which we leverage in a unique way considering we have no way of knowing "where" within the ZIP code the individual may reside. In order for us to connect each person's ZIP to the block group, we identify all block group centroids within and adjacent to a ZIP code and include these areas as a person's local environment where we define adjacency as within the Zip code area or bordering the Zip. We tally all persons of the same race for all these block groups associated with each participant's ZIP. We perform a similar exercise to construct an localized area's average income, persons serving in the military, food access average, and individuals with health insurance.

Figure 7: Example ZIP Code of Participant



Associating adjacent and contained block groups of a ZIP, we can construct an individual's local racial environment; but we do acknowledge a couple of drawbacks. First, in less populated areas we will see that ZIP code area will cover more land area whereas the opposite will be true in more densely populated areas. Therefore, an individual's interaction with the racial makeup of their

Figure 8: Example ZIP Code of Participant With Associated Block Groups



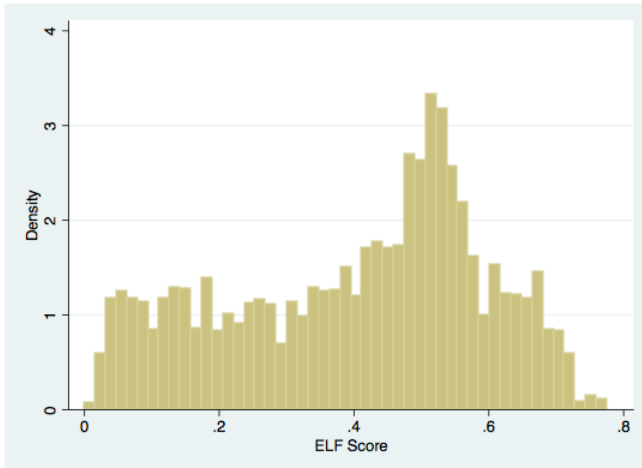
surroundings will differ by municipality due to the concentration of the population. This makes sense considering that a larger land area allows individuals to avoid others of a differing race or they must travel further to acquire food. Second, even though we consider the adjacent block groups to a ZIP there still could be instances where the participant can travel outside this area to avoid different individuals or find different food options. Both these caveats suggest attenuation bias as we would see no change in healthy eating consumption due to a change in diversity.

Lastly, we construct the ELF measure from our racial tallies for each land area. The ELF as defined in the development literature is

$$ELF = 1 - \sum_{i=1}^{n} p_i^2,$$

where p_i is the proportion of the population who identify as race i and n is the total number of identified races within our ZIP code area including adjacent and contained block groups. This measure is simply defined as the probability that two randomly chosen individuals in a given area will identify as different races in our context. While future analysis may use alternative measures, the ELF accomplishes our goal of beginning to understand the relationship between diversity and health.

Figure 9: ELF Measures for All Participants



We construct an ELF measure for each participant's ZIP code (including blocks groups within and adjacent to the Zip). A score of 0 indicates a perfectly homogenous population i.e. there exists only one racial group in the area while a score of 1 indicates perfect heterogeneity of races i.e. many racial groups of the same size. We observe a fairly uniform concentration of ELF measures across all individuals with a mean of .397 and an upper bound of .776. While we would prefer to see more heterogeneity in the sample, we cannot expect this outcome in any population center given the U.S. population (in general) skews majority white with much smaller racial subgroups those lowering ELF. There is a wide breath of ELF measures and we believe this measure is a good fit for our analysis given our population.

Model

Given the data and our intent, we present this model which we will employ in our analysis. The first model is given as:

$$HEI = \beta_0 + \beta_1 ELF + \beta_2 X_{Controls}$$

where $X_{Controls}$ includes income as measured as a percentage below or at the poverty threshold, average local area income, indicator for whether the participant exercises during the week, racial indicators, education indicators, municipality indicators, whether there are children in the household, the participant's age, average number of persons with health insurance in the area, average number of persons with military service in the area, and hispanic ethnicity indicator.

Aside from typical controls such as race, ethnicity, education, and municipality; we include the total number of individuals with health insurance in the local area to control for risk preferences of participants which can contribute to food consumption via alcohol consumption which negatively impacts HEI. Further, we include military service members in an area to control for local peer effects. Considering we will be using military bases as an instrument, it is important we note any effect these members may have on the EFNEP participants. Lastly, we include food access in the local area as a covariate to control for food acquisition problems which may arise in food insecure areas.

Overall, we hypothesize that as diversity increases we should observe increased healthy food consumption. The conduit is the market response to the racial heterogeneity in an area. Food purveyors will respond to increased diversity by adding more diverse food products, but the questions are at what point will they add supply to the market and does this lead change lead to healthier outcomes in general? For this supply change to account for the changing demographics we assume that income is an important factor. Without income in an area, supply cannot adjust to accommodate a more diverse population.

Results and Analysis of OLS Estimates

We start with the standard regression model without controls, with controls, and with clustered errors at the Zip code level. Here, we observe that without any controls our racial diversity (ELF) shows an increase of 6.40 HEI points when a local environment goes from perfectly homogenous (i.e. $ELF = 0$) to perfectly heterogeneous (i.e. $ELF = 1$). This is an unrealistic interpretation, so from hence forth we will show results for a 10% increase in ELF which in the control free model is an increase in HEI of .640 points. Additionally, all results will be clustered for standard errors at the Zip code level and will be reported as such.

When including controls in the model or clustering the standard errors, we see that an increase of 10% in ELF leads to an increase of .224 HEI points. One standard deviation of HEI is 13 points, so this increase is fairly small; but this effect is larger than many of the covariates including health insurance, military service numbers, or food access. Moreover, we expect attenuation bias suggesting that an instrumental approach may yield a larger impact on healthy food consumption.

There appears to be a discrepancy of ELF's association when homogenizing by Male and Female subgroups. Males display a .369 HEI point increase for a 10% increase in ELF while Females show a .19 increase. Given our Male sample is comprised of 12,459 observations (relative to the Female's 80,270 sample size), this difference suggests Males are associated more of an increase in diversity; but both groups benefit in a statistically significant and practically significant way at the 5% level for both males and females.

Some interesting sub-results show that the average number of food deserts (i.e. food access) does not contribute to healthy eating outcomes among our low SES sample group. Moreover, the number of military service individuals in an area and the number of individuals with health insurance in the local area do not contribute to food consumption health. The local area average income contributes positively to health outcomes in a statistically significant way, but there is essentially no practical significance to this result.

Further analysis using Oster (2017) method of stability analysis shows that the coefficient is stable and bound within .02 and a z-score=8.18. This stability analysis stays consistent throughout our subgroup analysis.

Racial and ethnicity show to be highly correlated with consumption outcomes in the controls model

Table 2: Prelimenary Model Without Controls, With Controls, With Controls Clustered at Zip Code, and Homogonized by Sex and Clustered at Zip Code Level

	No Control	Controls	Male	Female
	HEI	HEI	HEI	HEI
ELF_1	6.398***	2.235**	3.687*	1.933*
	(9.22)	(2.87)	(2.46)	(2.31)
zipdesertavg		0.615	-0.914	0.897
		(0.82)	(-0.62)	(1.11)
bgavginc		0.0000528***	0.0000413**	0.0000550***
		(6.91)	(2.66)	(6.66)
activity		0.601	0.663	0.560
		(1.57)	(1.11)	(1.34)
male		-1.964***		
		(-7.54)		
mixed		-0.492	0.348	-0.590
		(-1.59)	(0.41)	(-1.84)
hispanic		5.166***	3.573***	5.352***
		(17.72)	(6.21)	(17.10)
black		-1.117***	-0.476	-1.204***
		(-4.12)	(-0.93)	(-4.01)
asian		4.465***	6.395***	4.155***
		(5.94)	(6.09)	(4.95)
native		1.464*	3.340*	1.135
		(2.47)	(2.37)	(1.94)
pacific		2.131**	3.950*	1.848*
		(2.96)	(2.35)	(2.31)
hs_2year		0.654**	0.304	0.720***
		(3.25)	(0.60)	(3.34)
fouryear		2.679***	2.189**	2.764***
		(7.90)	(3.22)	(7.62)
farm		-0.542	2.126	-1.861*
		(-0.42)	(0.92)	(-2.41)
burb		-0.0886	-0.463	-0.0469
		(-0.17)	(-0.40)	(-0.08)
small_town		-1.643***	-1.391	-1.690***
		(-4.64)	(-1.84)	(-4.44)
town		-1.104***	-0.468	-1.197***
		(-3.76)	(-0.74)	(-3.79)
kids		-0.650*	0.585	-0.953**
		(-2.12)	(1.01)	(-2.77)
Age		0.0812***	0.0746***	0.0816***
		(10.96)	(4.32)	(10.35)
_cons	48.60***	43.49***	40.82***	43.77***
	(160.69)	(54.91)	(27.66)	(50.88)
N	104478	92729	12459	80270

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

for all races. Hispanic, Asian, Pacific, and Native ethnicity and racial covariates are associated with statistically significant increases in HEI with the largest increase being a .445 HEI increase for Asian persons. Table 2 shows that Black persons have a negative and significant relationship pertaining to HEI compared to a baseline of White persons with a decrease of .112 HEI points.

As expected, education level shows a monotonic increase in healthy food consumption with participants with a four year degree exhibiting an increase in HEI of 2.7 HEI points followed by an increase of .65 HEI points for individuals that have completed high school and have some college education. Our baseline are individuals without a high school diploma. This baseline shows that persons with more education have healthier food consumption.

Male participants have an average HEI score 1.96 points below their female counterparts. Additionally, any municipality outside a city shows negative effects on healthy eating scores except in a suburb where we observe a null effect. Moreover, having children contributes to less healthy eating. Aside from the male indicator, these municipal and kid indicators suggest that income or access may impact an individual's ability to acquire healthy foods or enough food. Lastly, if a person is active they will increase their HEI score by .6 points suggests active persons may choose healthier foods to consume.

OLS Results Homogenized by Race

Racial diversity shows to be highly correlated with consumption outcomes, but for only two racial subgroups when we investigate clustered models isolated by race. Hispanic, Asian, Pacific, and Native ethnicity and racial covariates are associated with statistically null increases in HEI . We see the results are driven primarily by white and mixed racial groups with respective associations of .224 and .449 HEI points. This result may stem from the fact that racial groups with already diverse diets (such as Asian persons) show no association with ELF because their diets are healthier than the average. For instance, Asian and Pacific persons have an average HEI score of 55 white White

Table 3: ELF Model Homogenized by Each Race
Clustered at the Zip Code Level

	White HEI	Black HEI	Asian HEI	Pacific HEI	Native HEI	Mixed HEI
ELF_1	2.235*	0.202	5.022	3.001	3.366	4.484*
	(2.23)	(0.14)	(1.49)	(0.47)	(0.85)	(2.51)
zipdesertavg	1.367	-0.244	-8.301	10.34*	2.947	0.699
	(1.57)	(-0.18)	(-0.89)	(2.05)	(1.27)	(0.30)
bgavginc	0.0000627***	0.0000305*	0.0000624**	0.0000693*	0.0000324	0.0000670
	(5.54)	(2.28)	(2.82)	(2.00)	(1.21)	(4.69)
activity	-0.0179	1.257*	5.450*	-0.724	1.628	-0.285
	(-0.03)	(2.05)	(2.12)	(-0.45)	(1.35)	(-0.36) ***
male	-2.556***	-1.459***	-0.413	-0.171	-1.536	-2.393***
	(-7.59)	(-3.45)	(-0.40)	(-0.10)	(-1.75)	(-3.36)
hispanic	5.734***	3.244***	-1.182	-2.156	7.786***	3.427***
	(16.46)	(3.77)	(-0.72)	(-1.24)	(5.69)	(6.07)
hs_2year	0.705**	0.596	0.662	-2.627*	-0.0358	1.293*
	(2.61)	(1.53)	(0.82)	(-2.03)	(-0.05)	(2.55)
fouryear	2.767***	2.981***	-1.251	-0.156	2.401**	3.717***
	(5.90)	(5.27)	(-0.94)	(-0.10)	(2.70)	(4.11)
farm	-1.451	0.412	16.41***	-8.098***	11.36***	-1.586
	(-1.88)	(0.27)	(10.71)	(-3.67)	(5.23)	(-0.64)
burb	1.056	-1.069	-3.805	2.194	-0.597	-0.953
	(1.43)	(-1.22)	(-0.98)	(0.75)	(-0.39)	(-1.26)
small_town	-0.758	-2.951***	-1.700	-0.167	-0.862	-2.329*
	(-1.53)	(-5.52)	(-0.91)	(-0.09)	(-0.67)	(-2.44)
town	-0.697	-1.580***	-0.805	-2.450	2.098	-1.565*
	(-1.76)	(-3.41)	(-0.43)	(-1.36)	(1.04)	(-2.38)
kids	-1.558***	-0.191	3.491	-0.866	1.798	0.147
	(-3.73)	(-0.36)	(1.89)	(-0.41)	(1.64)	(0.21)
Age	0.0633***	0.0827***	0.121***	0.154***	0.149***	0.114***
	(5.77)	(7.09)	(3.96)	(3.52)	(4.89)	(5.10)
_cons	44.36***	43.87***	37.75***	43.98***	37.33***	42.14***
	(42.24)	(33.72)	(8.97)	(8.85)	(12.82)	(24.22)
N	48545	27325	2294	672	4607	9286

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

persons show an average score of 51 HEI points.

We, again, observe minimal to no effects from our insurance, military service, and block group average income covariates suggesting that no matter an individual's race these effects are extremely small or nonsignificant suggesting that these local measures play a very limited role in healthy food consumption. The food access measure shows an effect on Pacific and White persons with a significant positive 10.34 and 1.37 point effects on HEI, respectively.

Here, we note consistent covariates that are either statistically null or the same sign across all racial groups. Hispanic ethnicity positively contributes all races with a maximal contribution of .78 HEI points for Native persons, but does not show significance for Asian and Pacific persons. Similarly, persons with a four year degree display positive gains in HEI across all groups except Pacific and Asian persons (statistically null) with a maximum HEI increase of 3.72 points. The age covariate increases all races HEI with a maximum of .154 HEI points for Pacific persons. On the opposite side, being male decreases all persons HEI no matter race with the largest drop of 2.56 points for White persons. The activity indicator shows increases HEI for Black, Asian, and Native individuals.

Municipality, having children, and having some college show mixed results across races. For example, living in a farming community benefits Native person's HEI by a large 1.136 HEI points, but this setting negatively impacts Pacific persons by a loss of .81 HEI points. Moreover, the small town setting has a -.30 HEI influence on Black persons and and mixed persons suggesting certain barriers for these races. We see s similar result with participants with some college education where Pacific persons lose .26 HEI points while Mixed persons gain .13 HEI points. These fluctuations suggest certain covariates impact persons in very different ways, especially given the participant's setting and racial/cultural upbringing. For example, Pacific persons residing in farm areas may have trouble finding a principle diet component such as fish protein.

OLS Results Homogenized by Education

As expected, restricting the model to only participants with a four year degree shows null effects from changes in ELF, thus demonstrating that more educated individuals do not show associative benefits from increased HEI. This result may stem from higher educated persons already eating better diets on average. We find statistically significant results for persons that have completed high school and some college with an associative gain of .27 HEI points. Persons with less high school showed similar effects from changes in HEI, but the significance is at the 6% level. These results may suggest that some education may be helpful in persuading persons to change diet when given more options, but this a speculative conclusion. Lastly, persons with limited education show no gains from increases in ELF.

When we homogenize by education, there exist themes across all education levels. Males exhibit lower HEI scores on average, Hispanic persons have higher HEI scores, living in a small town or town setting decreases HEI scores on average which may be caused from a lack of availability of differentiated foods. Black participants show statistically significant decreases in HEI given high school completion and some college education relative to their White peers, but we see gains for Asian persons in education levels below a four year degree. Pacific persons show gains in the less high school model and the four year model. Mixed persons show a decrease in HEI, but only for the less high school model. Additionally, Age contributes positively to all education levels.

We note the limited to null impact on HEI by total number insured, whether there are increased numbers of military service persons, and average area income covariates. For example, area's income contributes in an extremely fractional way where a $1,000 increase in average area income contributes less than 1 point to an individual's HEI score.

The ELF's influence on HEI is positive and significant for persons with either some college or less than a high school diploma (at 6%) while a person with a four year degree sees no benefit from an increase in local ELF. This finding matches many studies that suggest educational attainment increase healthy outcomes such as Schillinger et al. (2006). This outcome makes sense because of individual's ability to assess what constitutes healthy food consumption such as reading food labels. Moreover, parsing education from income shows that education contributes to healthful food consumption. With this low income sample, our low education status seems to contribute to ELF's effect on health. The local diversity contributes an average of 2.6 HEI points which is a practically significant increase considering these are low SES persons with limited education and resources.

OLS Model Homogenized by Municipality

Lastly, we homogenize by municipality since this will capture the food acquisition environment. Ahern et al. (2011) show that the food environment contributes in a significant way to health outcomes so parsing the sample by municipality will show where the ELF has more pronounced effect. .

Local racial diversity makes a significant contribution to HEI in the City setting only with a .30 HEI point increase. These other areas lack of influence may result simply because our supply side argument requires larger populations in general, let alone racial diversity.

We observe decreases for male again except in the farm communities. Hispanic persons show an increase across all municipalities as shown in Table 5. Black persons exhibit decreases in HEI in suburban, town, and small town settings. Considering all other races show null or positive effects against their White peers, the result that Black persons show decreases in these areas is an interesting result. Could we observe some level of animus toward black persons?

We see the same consistency as the previous models concerning Race, Male, Hispanic, Four year degrees, and Age. Average area income, total insured, and total persons with military service

Table 4: ELF OLS Homogenized by Education Level and Clustered at Zip

	Some HS HEI	More HS HEI	Four Year HEI
ELF_1	2.580	2.673**	-0.321
	(1.95)	(2.92)	(-0.19)
zipdesertavg	0.253	0.371	2.597
	(0.22)	(0.43)	(1.08)
bgavginc	0.0000501***	0.0000515***	0.0000648**
	(4.62)	(5.14)	(3.06)
activity	0.868	0.154	1.941
	(1.42)	(0.34)	(1.52)
male	-1.393**	-2.039***	-2.667***
	(-2.87)	(-6.48)	(-4.46)
mixed	-1.354**	-0.0809	1.305
	(-2.76)	(-0.20)	(1.50)
hispanic	6.520***	4.591***	3.789***
	(13.55)	(13.25)	(5.65)
black	-0.403	-1.367***	-0.880
	(-0.88)	(-4.11)	(-1.33)
asian	6.416***	5.410***	1.484
	(7.14)	(8.11)	(0.81)
native	1.600	1.263	2.284***
	(1.67)	(1.60)	(3.36)
pacific	3.218**	1.440	3.288*
	(3.06)	(1.45)	(2.57)
farm	-0.868	-0.311	-1.727
	(-0.89)	(-0.17)	(-1.10)
burb	0.155	-0.0694	-0.936
	(0.17)	(-0.10)	(-0.84)
small_town	-0.644	-1.896***	-2.580*
	(-1.13)	(-4.67)	(-2.28)
town	-0.668	-1.192***	-1.710**
	(-1.32)	(-3.41)	(-2.83)
kids	0.130	-0.724	-1.714*
	(0.28)	(-1.75)	(-2.45)
Age	0.0860***	0.0733***	0.0752**
	(7.41)	(7.59)	(3.27)
_cons	41.05***	45.16***	47.11***
	(34.73)	(45.76)	(21.40)
N	28187	52158	12384

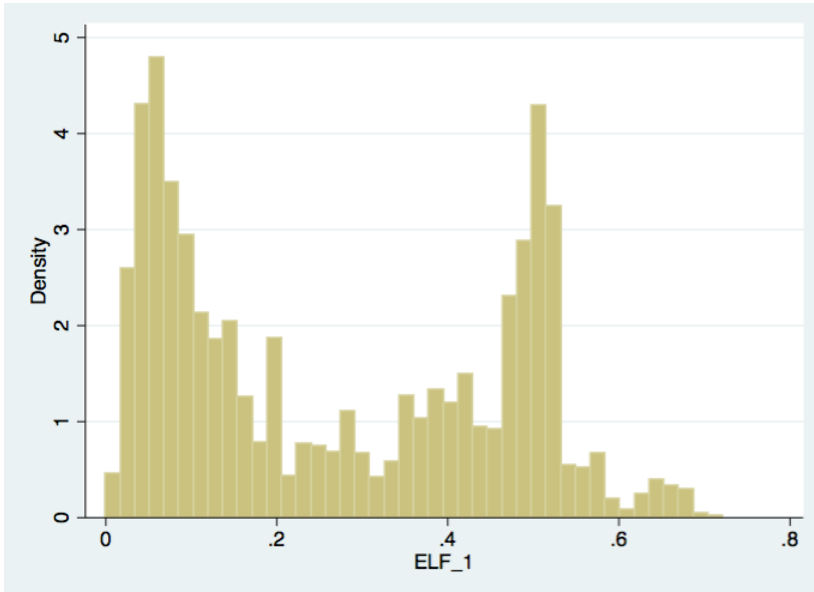
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: OLS Models Homogenized by Municipality Clustered at the Zip Code Level

	City HEI	Suburb HEI	Town HEI	Small Town HEI	Farm HEI
ELF_1	3.043**	3.130	2.374	0.378	6.396
	(2.58)	(1.05)	(1.62)	(0.28)	(1.66)
zipdesertavg	-1.085	0.285	1.403	1.574	-0.313
	(-0.51)	(0.05)	(1.00)	(1.63)	(-0.16)
bgavginc	0.0000413***	0.0000391*	0.0000733***	0.000103***	0.000172***
	(4.03)	(2.38)	(5.03)	(3.48)	(3.59)
activity	0.623	3.289	0.133	-0.276	-0.292
	(1.19)	(1.85)	(0.19)	(-0.42)	(-0.16)
male	-2.049***	-3.193**	-1.730***	-2.269***	-1.173
	(-5.52)	(-2.81)	(-3.91)	(-4.72)	(-1.01)
mixed	-0.116	-1.920*	-0.709	-1.171	-0.516
	(-0.27)	(-2.09)	(-1.34)	(-1.28)	(-0.20)
hispanic	5.028***	4.441***	5.335***	6.071***	5.782***
	(12.50)	(4.67)	(9.67)	(8.31)	(3.73)
black	-0.631	-3.299**	-1.076*	-1.242*	1.399
	(-1.58)	(-3.10)	(-2.12)	(-2.41)	(0.69)
asian	5.144***	-0.990	3.901*	3.550*	22.12***
	(8.56)	(-0.19)	(1.99)	(2.13)	(7.34)
native	1.163*	-0.864	3.245	0.111	12.42***
	(2.14)	(-0.60)	(1.30)	(0.14)	(4.94)
pacific	2.269	2.517	1.157	3.342**	-5.151*
	(1.83)	(0.97)	(0.88)	(2.69)	(-2.30)
hs_2year	0.765**	0.199	0.622	0.385	1.176
	(2.88)	(0.21)	(1.47)	(1.01)	(1.03)
fouryear	2.753***	1.937	2.601***	2.632*	3.825*
	(6.58)	(1.59)	(4.72)	(2.24)	(2.24)
kids	0.0185	-2.758	-1.963***	-0.261	-0.764
	(0.04)	(-1.82)	(-3.83)	(-0.52)	(-0.43)
Age	0.0860***	0.0866*	0.0818***	0.0636***	0.102*
	(7.79)	(2.54)	(5.83)	(4.83)	(2.32)
_cons	42.34***	46.01***	43.34***	41.57***	35.34***
	(36.68)	(13.27)	(34.28)	(27.00)	(9.81)
N	46920	6478	21822	16333	1176

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 10: ELF Distribution for Small Town Setting



demonstrate more significance when the sample is broken by municipality. While, we do see the military service become significant for the Suburb, and Town settings, these values are extremely small suggesting that these variables still have limited contribution to healthy food consumption. For example, in a Town setting (where we observe strongest significance) there would need to be an increase of 10,000 military persons to net a .34 increase in HEI which would essentially turn the municipality into a new categorization.

OLS Results Homogenized by Food Access

Lastly, we homogenize relative to the local area's food acquisition measure we assigned. We perform a similar exercise as before when computing our food desert measure by taking the USDA's food acquisition measures and filtering them by our Zip code and averaging the measures to create a composite food desert score for the Zip code area. In this case, the measure closer to zero represents a strong food acquisition score and a score of one represents a composite food desert score i.e. food acquisition is difficult.

We observe many individuals in our sample with access to food. In fact, only 1% of our sample resides in an area that is considered a complete food desert. Even though this may be the case, we ensure that food acquisition is not conflated with local racial diversity, so we homogenize by a composite score of .05 because .05 is the mean of the sample.

We can immediately see that food acquisition is not extremely problematic for our participants, but ELF's effect is significant in areas with access to food and shows no significant effect for person's with limited food access showing a .29 HEI increase. This result makes sense considering diversity contributes to the supply of more foods. Ergo, limited supply of food should imply that local diversity cannot matter as there is no conduit for participants to choose healthier food consumption.

As shown, standard OLS results continue to show positive effects from racial diversity in an area. In fact, the only subgroups that show no effects from an increase in ELF are the individuals with a four year degree and persons that reside in a small town setting where both of these results should be expected considering the literature.

Conclusion

The relationship between diversity and health is as complicated as the literature suggests. We see that the increases in local racial diversity are associated with better health outcomes among individuals regarding actual food consumption choices, but only for Black and White persons. Additionally, we find strong associations of ELF on HEI with individuals have lower education. Moreover, city residents show the only effect of ELF on HEI. These results show that diversity is associated with gains in health, but it appears this association occurs through limited channels such as less four year education, population size, and a baseline dietary habits.

We might not observe food consumption changes given a supply increase as Allcott et al. (2017)

Table 6: OLS Results Homogenized by Food Security

	Low Food Acq	High Food Acq
	HEI	HEI
ELF_1	-1.609	2.924***
	(-0.76)	(3.56)
bgavginc	0.0000819*	0.0000529***
	(2.29)	(6.72)
activity	0.653	0.599
	(0.70)	(1.45)
male	-2.315***	-1.900***
	(-3.88)	(-6.79)
mixed	-1.212	-0.408
	(-1.23)	(-1.25)
hispanic	5.592***	5.034***
	(6.25)	(16.45)
black	-0.341	-1.155***
	(-0.52)	(-3.88)
asian	2.502	4.381***
	(1.02)	(5.68)
native	2.389	1.276*
	(0.95)	(2.44)
pacific	2.548	1.889*
	(1.27)	(2.41)
hs_2year	0.187	0.761***
	(0.33)	(3.61)
fouryear	2.411***	2.745***
	(3.38)	(7.43)
farm	-1.363	0.339
	(-1.08)	(0.20)
burb	2.553	-0.224
	(1.54)	(-0.41)
small_town	-0.630	-1.550***
	(-0.70)	(-3.89)
town	0.355	-1.197***
	(0.43)	(-3.79)
Age	0.0890***	0.0798***
	(5.44)	(9.66)
_cons	41.70***	43.30***
	(14.79)	(51.66)
N	13840	78889

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

argues, but there may be additional factors, such as racial animus, preventing certain racial groups (Black persons) from increasing healthful consumption through a supply increase when compared to their White peers. Our paper suggests that there does not exist a blanket effect of racial diversity. As we show, local racial diversity can account for variation in health outcomes via food consumption, but by varying levels and individuals given their locale, education, and race.

Future studies in this vein may examine the efficacy of EFNEP or food assistance take-up within the context of local diversity to understand whether we observe heighten or lessened outcomes due to the racial environment. Polarization would be an interesting inclusion to future analysis as it was not presented here because of data problems. As the U.S. becomes more diverse, understanding individuals' responses and needs will be very important when setting policy, so fundamentally understanding the varying capacities ELF has on different individuals will become more important.

References

Agranov, M., J. K. Goeree, J. Romero, and L. Yariv (2017). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association*.

Ahern, M., C. Brown, and S. Dukas (2011). A national study of the association between food environments and county-level health outcomes. *The Journal of Rural Health 27*(4), 367–379.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of economic literature 43*(3), 762–800.

Allcott, H., R. Diamond, J.-P. H. Dubé, et al. (2017). *The geography of poverty and nutrition: Food deserts and food choices across the United States*. National Bureau of Economic Research.

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Andris, C., D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden (2015). The rise of partisanship and super-cooperators in the us house of representatives. *PloS one 10*(4), e0123507.

Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, Volume 30. Springer Science & Business Media.

Baldwin, K. and J. D. Huber (2010). Economic versus cultural differences: Forms of ethnic diversity and public goods provision. *American Political Science Review 104*(4), 644–662.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Berinsky, A. J. (1999). The two faces of public opinion. *American Journal of Political Science*, 1209–1230.

Bishop, G. F. and B. S. Fisher (1995). Secret ballots and self-reports in an exit-poll experiment. *Public Opinion Quarterly 59*(4), 568–588.

Bitler, M. (2014). The health and nutrition effects of snap: Selection into the program and a review of the literature on its effects.

Bitler, M. P. and J. Currie (2005). Does wic work? the effects of wic on pregnancy and birth outcomes. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management 24*(1), 73–91.

Blair, G. and K. Imai (2012). Statistical analysis of list experiments. *Political Analysis 20*(1), 47–77.

Blisard, N., B.-H. Lin, J. Cromartie, and N. Ballenger (2002). America's changing appetite: Food consumption and spending to 2020. *FOOD REVIEW-WASHINGTON DC- 25*(1), 2–9.

Brown-Iannuzzi, J. L., M. B. Najle, and W. M. Gervais (2017). The illusion of political tolerance: Social desirability and self-reported voting preferences. *Journal of Experimental Social Psychology*.

Bursztyn, L., D. Cantoni, P. Funk, and N. Yuchtman (2017). Polls, the press, and political participation: The effects of anticipated election closeness on voter turnout.

Charness, G. and U. Gneezy (2008). What's in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization 68*(1), 29–35.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Claassen, R. L. and J. B. Ryan (2016). Social desirability, hidden biases, and support for hillary clinton. *PS: Political Science & Politics 49*(4), 730–735.

Coffman, K. B., L. C. Coffman, and K. M. M. Ericson (2016). The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*.

Connors, E., S. Klar, and Y. Krupnikov (2016, November). There may have been shy trump supporters after all. *The Washington Post*.

Coppock, A. (forthcoming). Did shy trump supporters bias the 2016 polls? evidence from a nationally-representative list experiment. *Statistics, Politics, and Policy*.

Corstange, D. (2008). Sensitive questions, truthful answers? modeling the list experiment with listit. *Political Analysis 17*(1), 45–63.

Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

Droitcour, J., R. A. Caspar, M. L. Hubbard, T. L. Parsley, W. Visscher, and T. M. Ezzati (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 185–210.

Dropp, K. (2015, December). Why does donald trump perform better in online versus live telephone polling? *Morning Consult*.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics 112*(4), 1203–1250.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research.

Enten, H. (2016a, May). Americans? distaste for both trump and clinton is record-breaking. *FiveThirtyEight*.

Enten, H. (2016b, November). 'shy' voters probably arent why the polls missed trump. *FiveThirtyEight*.

Esteban, J. and D. Ray (2008). Polarization, fractionalization and conflict. *Journal of peace Research 45*(2), 163–182.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *Review of Economic Studies 83*(2), 587–628.

Fang, D., M. R. Thomsen, R. M. Nayga Jr, and A. M. Novotny (2018). Wic participation and relative quality of household food purchases: Evidence from foodaps. Technical report, National Bureau of Economic Research.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research 20*(2), 303–315.

Fox News (2016, January). See which candidates qualified for the fox news-google gop debates.

Goffman, E. (1959). *The presentation of self in everyday life*. Harmondsworth.

Gundersen, C., B. Kreider, and J. Pepper (2011). The economics of food insecurity in the united states. *Applied Economic Perspectives and Policy 33*(3), 281–303.

Gundersen, C., B. Kreider, and J. V. Pepper (2017). Partial identification methods for evaluating food assistance programs: A case study of the causal impact of snap on food insecurity. *American Journal of Agricultural Economics 99*(4), 875–893.

Hales, C. M., M. D. Carroll, C. D. Fryar, and C. L. Ogden (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. US Department of Health and Human Services, Centers for Disease Control and ?

Halpin, J. and K. Agne (2009). State of american political ideology, 2009. *Center for American Progress*.

Heerwig, J. A. and B. J. McCabe (2009). Education and social desirability bias: The case of a black presidential candidate. *Social Science Quarterly 90*(3), 674–686.

Hoffman, E., K. McCabe, and V. L. Smith (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review 86*(3), 653–660.

Hopkins, D. J. (2009). No more wilder effect, never a whitman effect: When and why polls mislead about black and female candidates. *The Journal of Politics 71*(3), 769–781.

Huang, J. and C. Low (2017). Trumping norms: Lab evidence on aggressive communication before and after the 2016 us presidential election. *American Economic Review 107*(5), 120–24.

Janus, A. L. (2010). The influence of social desirability pressures on expressed immigration attitudes. *Social Science Quarterly 91*, 928–946.

Jekanowski, M. D., J. K. Binkley, et al. (2000). Food purchase diversity across us markets. *Agribusiness 16*(4), 417–433.

Kane, J. G., S. C. Craig, and K. D. Wald (2004). Religion and presidential politics in florida: A list experiment. *Social Science Quarterly 85*(2), 281–293.

Kantchev, G. and C. Whittall (2017, June). After another surprise election result, investors grow skeptical of pollsters.

Karlan, D. S. and J. Zinman (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics 98*(1), 71–75.

Keeter, S. and N. Samaranayake (2007). Can you trust what polls say about obama's electoral prospects. *Pew Research Center*.

Kinsey, J. D. (1994). Food and families' socioeconomic status. *The Journal of nutrition 124*(suppl_9), 1878S–1885S.

Kreider, B., J. V. Pepper, and M. Roy (2018). Does the women, infants, and children program (wic) improve infant health outcomes?

Krysan, M. (1998). Privacy and the expression of white racial attitudes: A comparison across three contexts. *Public Opinion Quarterly*, 506–544.

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. Vishny (1999). The quality of government. *The Journal of Law, Economics, and Organization 15*(1), 222–279.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social forces 13*(2), 230–237.

Lax, J. R., J. H. Phillips, and A. F. Stollwerk (2016). Are survey respondents lying about their support for same-sex marriage? lessons from a list experiment. *Public Opinion Quarterly*, 510–533.

Lo, Y.-T., Y.-H. Chang, M.-S. Lee, and M. L. Wahlqvist (2012). Dietary diversity and food expenditure as indicators of food security in older taiwanese. *Appetite 58*(1), 180–187.

Maccoby, E. E. and N. Maccoby (1954). The interview: A tool of social science. *Handbook of social psychology 1*, 449–487.

Mas, A. and E. Moretti (2009). Racial bias in the 2008 presidential election. *The American Economic Review 99*(2), 323–329.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of public Economics 89*(11-12), 2325–2368.

Miller, J. D. (1984). A new survey technique for studying deviant behavior. *Unpublished doctoral dissertation, George Washington University, Washington, DC*.

Miller, J. M. and J. A. Krosnick (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, 291–330.

Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review 95*(3), 796–816.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology 46*(3), 598.

Pew Research Center (2012, May). Assessing the representativeness of public opinion surveys.

Pew Research Center (2016, June). Partisanship and political animosity in 2016.

Powell, R. J. (2013). Social desirability bias in polling on same-sex marriage ballot measures. *American Politics Research 41*(6), 1052–1070.

Raghavarao, D. and W. T. Federer (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40–45.

Redlawsk, D. P., C. J. Tolbert, and W. Franko (2010). Voters, emotions, and race in 2008: Obama as the first black president. *Political Research Quarterly*.

Rothschild, D. (2009). Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly 73*(5), 895–916.

Rudman, L. A. and S. E. Kilianski (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin 26*(11), 1315–1328.

Santos, F. C., M. D. Santos, and J. M. Pacheco (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature 454*(7201), 213.

Schillinger, D., L. R. Barton, A. J. Karter, F. Wang, and N. Adler (2006). Does literacy mediate the relationship between education and health outcomes? a study of a low-income population with diabetes. *Public health reports 121*(3), 245–254.

Shepard, S. (2016, October). Gop insiders: Polls don't capture secret trump vote. *Politico*.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics 118*, 26–40.

Streb, M. J., B. Burrell, B. Frederick, and M. A. Genovese (2008). Social desirability effects and support for a female american president. *Public Opinion Quarterly 72*, 76–89.

Traugott, M. W. and V. Price (1992). A review: exit polls in the 1989 virginia gubernatorial race: Where did they go wrong? *Public Opinion Quarterly 56*(2), 245–253.

Tsuchiya, T., Y. Hirai, and S. Ono (2007). A study of the properties of the item count technique. *Public Opinion Quarterly 71*(2), 253–272.

Ver Ploeg, M. (2010). *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to Congress*. Diane Publishing.

Vespa, J., D. M. Armstrong, and L. Medina (2018). Demographic turning points for the united states: population projections for 2020 to 2060. *Washington, DC: US Census Bureau 42*.

Wagner, A., R. J. Zeckhauser, and A. Ziegler (2017). Company stock reactions to the 2016 election shock: Trump, taxes and trade. Technical report, National Bureau of Economic Research.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association 60*(309), 63–69.

Wolfers, J. and E. Zitzewitz (2004). Prediction markets. *The Journal of Economic Perspectives 18*(2), 107–126.

Wolfers, J. and E. Zitzewitz (2016). What do financial markets think of the 2016 election. Technical report, Working paper.

Yen, S. T. (2010). The effects of snap and wic programs on nutrient intakes of children. *Food Policy 35*(6), 576–583.

Chapter 2: Social Desirability Bias and Polling Errors in the 2016 Presidential Election

Abstract

Social scientists have observed that socially desirable responding (SDR) often biases unincentivized surveys. Nonetheless, media, campaigns, and markets all employ unincentivized polls to make predictions about electoral outcomes. During the 2016 presidential campaign, we conducted three list experiments to test the effect SDR has on polls of agreement with presidential candidates. We elicit a subject's agreement with either Hillary Clinton or Donald Trump using explicit questioning or an implicit elicitation that allows subjects to conceal their individual responses. We find evidence that explicit polling overstates agreement with Clinton relative to Trump. Subgroup analysis by party identification shows that SDR significantly diminishes explicit statements of agreement with the opposing party's candidate driven largely by Democrats who are significantly less likely to explicitly state agreement with Trump. We measure economic policy preferences and find no evidence that ideological agreement drives SDR. We find suggestive evidence that local voting patterns predict SDR.

Introduction

Political polls generate sweeping economic and political consequences far in advance of election day. Polling numbers motivate changes in campaign spending, staff deployment, fundraising efforts, and even policy positions. Strong polling numbers, for example, motivated Hillary Clinton's 2016 campaign to forgo campaigning in certain states in the upper Midwest that her opponent, Donald Trump, subsequently won. Polls play a structural role in winnowing television debate participants (Fox News (2016)), help voters evaluate the viability of candidates, and influence electoral turnout (Bursztyn et al. (2017); Agranov et al. (2017)). Polls also have direct economic consequences by influencing forecasts about the future business environment (Kantchev and Whittall (2017)). As a result, market prices fluctuate in response to polling (Wolfers and Zitzewitz (2016)) and election results that polls suggested were unlikely "shock" prices in predictable ways (Wagner et al. (2017)). Proponents of prediction markets cite their decreased volatility as an advantage over traditional polling (Wolfers and Zitzewitz (2004); Rothschild (2009)). Nonetheless, these markets respond to new polling information, exposing them to the risk of similar surprises.

Since an incentive-compatible method of collecting voting preferences would be infeasible—and illegal in most cases—methods that rely on stated preference between candidates have been accepted as viable, second-best alternatives. Critics of polling typically point to its vulnerability to non-response bias and optimism bias (Pew Research Center (2012); Armstrong (2001)). But social science research offers several other reasons that the assumption of truthful revelation in poll responses may be dubious. Since Maccoby and Maccoby (1954) and Edwards (1957), social scientists have known that these stated preference surveys are subject to "socially desirable responding" (SDR, hereafter)—that is, respondents tend to conceal preferences that are not perceived to be socially desirable. Researchers have identified SDR in many social, political, and economic contexts.[1] For example, feelings toward African-American politicians (Heerwig and McCabe (2009); Redlawsk et al. (2010); Stephens-Davidowitz (2014)), female politicians (Streb et al. (2008)), and Jewish politicians (Kane et al. (2004)) are affected by SDR. Brown-Iannuzzi et al. (2017) found that respondents conceal discriminatory political preferences only when it is "socially inappropriate" to discriminate against the group in question. SDR also has been shown to influence the expression of sentiments surrounding immigration (Janus (2010)), same-sex marriage (Powell (2013); Lax et al.

[1] Paulhus (1984); Droitcour et al. (1991); Fisher (1993); Rudman and Kilianski (2000); Karlan and Zinman (2012)

(2016); Coffman et al. (2016)), and race (Krysan (1998)).

In contrast to other research that analyzes secondary data, our paper analyzes data collected with the express purpose of identifying the effect of SDR on candidate polling. We cover both telephone and online environments using a methodology designed specifically to test for SDR in responses to questions about agreement with Clinton and Trump.[2] Other researchers have addressed the role of SDR in the 2016 election in different ways. Claassen and Ryan (2016) use two forms of indirect questioning asking about the perceptions of support for each candidate to measure the influence of SDR on the 2016 election, finding little or no influence of SDR. Coppock (ming) performs a list experiment similar to ours and finds no evidence of SDR affecting Trump support. Coppock (ming) has key differences from our study: it is exclusively online, does not compare the effect of SDR across both candidates, and repeats a question about voting intentions that was asked earlier in the survey.[3] A Morning Consult study offers an in-depth analysis comparing responses to telephone and online polls to find a small but not statistically significant increase in support for Trump in online polls (Dropp (2015)). Other analyses have re-analyzed traditional polls to assert that SDR provided no significant threat to the validity of traditional polls.[4]

Our results contradict the conclusions of polling agencies and data journalists and show marginally significant evidence that SDR causes polling respondents to understate their agreement with Trump and overstate their agreement with Clinton. We decompose our sample by political party and find that SDR causes a large and significant drop in the willingness of voters to state agreement with the opposing party's candidate. Additionally, we find that, while the effect of SDR is closely related to party identification, it is unrelated to political ideology. That is, SDR is closely tied to the party a voter has chosen but is unrelated to policy preferences that may have driven him or her to that party.

[2]All analysis is run within a polling medium to control for medium-specific effects.

[3]We elicited "agreement" with candidates because the telephone poll already included a question about voting intentions. We believe that the desire for consistency may bias our design away from finding SDR in this case. 85% of our sample in the Arkansas Poll indicated plans to vote for the candidate they "agreed with," making this a strong instrument for voting behavior. We maintained this measure in our online replication.

[4]For example, Enten (2016b); Connors et al. (2016); Shepard (2016)

With historically high candidate unfavorable ratings (Enten (2016a)), the 2016 presidential election provides optimal conditions under which SDR could threaten the validity of political polls. Moreover, the voting bases of each party also report historically high levels of partisanship (Andris et al. (2015); Pew Research Center (2016)). This allows us to understand how SDR interacts with a divided electorate.[5]

We use three list experiments (a method sometimes called the “item count” or “unmatched count” technique) to estimate the effect of SDR on political polling. This method was developed by Miller (1984) to understand the ways in which respondents predictably misreported answers to unincentivized polling questions.[6] In a list experiment, subjects are presented with a list of statements and asked to report the *total number* with which they agree. Half of the subjects are assigned to the *Implicit* treatment in which their list features five statements, including a “sensitive” statement of interest.[7] The other half of the subjects are assigned to the *Explicit* treatment; this list consists of the same four non-sensitive statements in the same order and is followed by a direct “Yes” or “No” question about the sensitive statement.[8] Thus, all respondents face the same five statements, but the treatment assignment randomly varies the observability of an individual’s response to the sensitive statement. Blair and Imai (2012) and Corstange (2008) validate and formalize the analysis and methodology of list experiments. Critical to the validity of this methodology is the restriction that only socially undesirable responses be affected. Tsuchiya et al. (2007) and Coffman et al. (2016) use placebo tests to validate the methodology.

Figure 11 displays our Implicit and Explicit elicitations. The first two experiments measure the SDR associated with statements of agreement with presidential candidates. The final experiment tests for a differential effect of economic policy preferences on the SDR associated with each candi-

[5]In 1969, Richard Nixon referred to the “silent majority” of people who concealed their support for the Vietnam War. Similarly, the “Bradley effect” was a hypothesized reluctance among voters to reveal their votes against Tom Bradley were racially motivated. In Great Britain, a similar theory has been labeled the “Shy Tory Factor.”

[6]A similar method was proposed in Raghavarao and Federer (1979).

[7]We are choosing to use the terms “implicit” and “explicit” to indicate whether or not the respondent openly revealed preferences for candidates. These should not be confused with similar terms from psychological research. Indeed, our terms are more similar to “indirectly” and “directly” revealed preferences.

[8]Miller and Krosnick (1998) find that the ordering of candidates can influence voter behavior. Thus, we chose to hold all ordering constant to provide a valid *comparison* between the two treatments.

date. In all three experiments, subjects are randomly assigned to the Implicit or Explicit treatment and then are presented with a sensitive statement that asks about *agreement with* a presidential candidate. Experiment 1—a live telephone poll of 800 Arkansas residents—elicits responses to the statement, "I often find myself agreeing with Donald Trump." In Experiments 2 and 3—online surveys with approximately 1,000 eligible voters each—we randomly assign subjects to respond to either 1) "I often find myself agreeing with Hillary Clinton" *or* 2) "I often find myself agreeing with Donald Trump."

It is important to note that our sensitive statement does not ask which candidate respondents intend to vote for, but simply asks if subjects "often agree" with a randomly assigned candidate. This accomplishes two objectives: 1) It prevents us from repeating a question that was previously asked explicitly in the telephone poll and 2) It allows us to explore the psychological motivations behind revealing candidate preferences that are not as transparent as candidate choice.

Figure 11: Examples of both Explicit and Implicit elicitations of support for Donald Trump.

IMPLICIT:

Consider the following list of statements. Below, we will ask how many of the statements you agree with.

- I think small businesses are important for the economy.
- I agree with George H.W. Bush's foreign policy.
- I think the threat of global warming is exaggerated.
- I often find myself agreeing with Donald Trump.
- I prefer presidential candidates who oppose the NRA.

How many of the previous statements do you agree with? (0) (1) (2) (3) (4) (5)

EXPLICIT:

Consider the following list of statements. Below, we will ask how many of the statements you agree with.

- I think small businesses are important for the economy.
- I agree with George H.W. Bush's foreign policy.
- I think the threat of global warming is exaggerated.
- I prefer presidential candidates who oppose the NRA.

How many of the previous statements do you agree with? (0) (1) (2) (3) (4)

Do you often find yourself agreeing with Donald Trump?

YES NO

Comparing responses under the Implicit and Explicit treatments yields a clean comparison of otherwise identical environments where the only change is the psychological cost associated with reporting a socially undesirable response. For example, in the Implicit treatment, a response of

"3," does not reveal a position on the sensitive statement. Subjects in the Explicit treatment always reveal their position on the sensitive statement. If agreement with one candidate is more stigmatized than another, we expect to find that the polling numbers for the stigmatized candidate improve relative to the non-stigmatized candidate when elicited under the Implicit treatment. Our results indicate that a differential effect of SDR does exist between the two candidates. In particular, in our first Mechanical Turk study, we find that the Explicit treatment decreases statements of agreement with Trump by 0.10 and increases statements of agreement with Clinton by 0.09. Dividing respondents by party identification, SDR grows for the opposing party's candidate and shrinks for the own-party candidate.

To explore the origins of SDR associated with each candidate, we merge our dataset to county-level election results. This merged dataset allows us to address the possibility that these phenomena may derive from the political preferences of a subject's region. We do not find consistent evidence that county-level voting patterns predict whether support for a given candidate is affected by SDR, but our results provide suggestions for future research.

It is possible that the psychological motivation causing polling respondents to misreport their preferences persists into the voting booth. In this case, SDR would not present a threat to the predictive validity of polls. Voting mechanisms, however, provide incentives to truthfully reveal preferences that opinion polls do not, opening the door to the possibility of discrepancies between the polls and election results. In the discussion section, we suggest ways polling agencies could use results from our mechanism as a bias correction for their explicit polling responses.

Other studies have used other methods to test whether or not SDR presents a threat to the validity of political polling (Traugott and Price (1992); Bishop and Fisher (1995); Keeter and Samaranayake (2007); Hopkins (2009)). Berinsky (1999) shows that poll respondents more often claim to be undecided when answering a poll potentially subject to SDR. Recent elections, in particular, have motivated additional research on the role of SDR in politics. Mas and Moretti (2009) and Stephens-Davidowitz (2014) explore the role of race in the 2008 election using indirect methods as a measure of racial animus. Huang and Low (2017) measure the impact of the 2016 election on dissolving social norms, which drive SDR. Our paper expands on this tradition by collecting primary data to estimate the effect of SDR on political polls.

Research in psychology has long shown a desire to strategically present one's self (LaPiere (1934); Goffman (1959)). The list experiment methodology attempts to circumvent this psychological motivation by manipulating the observability of the respondent's answer much in the same way as the randomized response technique (Warner (1965)). Manipulating this observability allows respondents to reveal socially undesirable behavior without direct observation. The psychological motivation to strategically present one's self relates to results from economics suggesting that anonymity and the ability to excuse behavior affect choices in social interactions (Dana et al. (2007); Hoffman et al. (1996); Andreoni and Bernheim (2009); Charness and Gneezy (2008); Bénabou and Tirole (2011); Exley (2016)). In exposing how political polling can be influenced by these psychological motivations, we hope to shed light on new methods that could present a clearer picture of voter preferences.

Experimental Design

Our list experiment is designed to understand the effect of SDR on affirmations of agreement with a political candidate—either "I often find myself agreeing with Hillary Clinton" or "I often find myself agreeing with Donald Trump"—by varying the observability of that affirmation. To experimentally vary response observability, subjects are assigned to either the Implicit or Explicit treatment, where their responses are veiled or directly observed, respectively.

In the Implicit treatment, the affirmation of agreement is included in a list with four neutral statements. Subjects are asked to respond with the total number—zero to five—of statements they affirm from the list. Call this a subject's "total affirmations." These total affirmations do not reveal agreement with any one statement, thus a subject's response to the affirmation of agreement is concealed.

In the Explicit treatment, subjects see a list with only the four neutral statements and respond with the total number they agree with. They then respond to the affirmation of agreement directly with a "yes" or "no" answer. In this case, call the "total affirmations" the aggregate number of agreements from the list plus the directly elicited affirmation of agreement.

In Figure 11, we presented examples of the Implicit and Explicit elicitations for subjects assigned to evaluate their agreement with Trump. Subjects assigned to evaluate their agreement with Clinton saw an identical list except that the candidate name was changed. The four neutral statements are identical in all three experiments. Like the affirmation of agreement, they are political statements. But, since they are presented identically in both the Implicit and Explicit treatments, any influence they have on responses will be constant across treatments. We chose neutral statements that negatively covary—support for gun-control and skepticism about climate-change—to limit the number of responses of zero or five, which would transparently reveal the opinions of a subject in the Implicit treatment.[9]

Our study is comprised of the following three list experiments, each successively narrowing in on relevant psychological phenomena. The Arkansas Poll attempts to measure SDR with respect to telephone surveys about Donald Trump. The first Mechanical Turk study replicates these findings in a different medium and includes measures of SDR on Hillary Clinton. The final Mechanical Turk study measures political preferences in an attempt to uncover the role played by ideological alignment. Detailed demographics and balance tables for all three experiments can be found in the appendix. It is worth noting that, while neither the Arkansas Poll nor the Mechanical Turk samples are themselves representative of the population at large, they combine to form a more representative sample that crosses polling mediums. With consistent results between them, we can be confident that the results will hold in a larger, representative sample regardless of the polling

[9] We will repeat the analysis with these observable responses dropped from the Implicit treatment for robustness.

medium.

Experiment 1: Arkansas Poll

Our first experiment was included in the Arkansas Poll, a live-telephone survey of 800 Arkansas residents between October 18 and October 27, 2016. 60 percent of respondents answered using land-line telephones and 40 percent using cell phones. The cooperation rate was 29 percent and 25 percent for land-lines and cell phones, respectively. Poll workers continued calling residents until they achieved a sample of 800 valid responses. Respondents skewed toward Trump with 45 percent of the sample indicating plans to vote for him compared to 31 percent for Clinton. The sample was older than the national average with a median age of 63. None of the respondents indicated that they had already voted.

The Arkansas Poll consisted of approximately 50 questions with several possible follow-up questions. Question 9 asked respondents which presidential candidate they intended to vote for. Our experiment took the place of the 29th and 30th questions, depending on treatment assignment. Question 34 requested the party affiliation of the respondent (Republican, Independent, Democrat).

Due to space limitations, we only explored responses to one affirmation of agreement, "I often find myself agreeing with Donald Trump" and did not measure responses to affirmations of agreement with Clinton. Respondents were randomly assigned to the Implicit or Explicit treatment.

Experiment 2: M-Turk Poll 1

On November 1, 2016, we conducted a second list experiment online with 1,006 eligible American voters using Amazon's Mechanical Turk website. Respondents skewed toward Clinton. Of our sample, 56 percent indicated they intended to vote for Clinton compared to 23 percent for Trump. The sample was disproportionately young, with a median age of 31. We elicited the demographics and party affiliation of each respondent prior to asking about candidate support.[10]

We explored responses to affirmations of agreement with both candidates. Each subject was randomly assigned to the Implicit or Explicit treatment and then assigned to respond to either "I often find myself agreeing with Hillary Clinton" or "I often find myself agreeing with Donald Trump." This gives us a "two by two" randomization design.

Experiment 3: M-Turk Poll 2

Our final experiment took place on November 7 and 8, 2016.[11] We recruited 985 eligible American voters again using Amazon's Mechanical Turk website. Of our respondents, 57 percent indicated plans to vote for Clinton relative to 27 percent for Trump. The respondents' median age was 32.

We again elicited demographic information, party affiliation, and which candidate the subject intended to vote for. We used the same randomization design as Experiment 2: assigning subjects to respond about agreement with Clinton or Trump and assigning them to either the Implicit or Explicit treatment.

In this experiment we added six questions about economic policy preferences. Three of the questions indicated economic policy preferences more aligned with Donald Trump and three indicated preferences more aligned with Hillary Clinton.[12] This elicitation will allow us to perform subgroup analysis in treatment responses by ideology.

[10]Respondents were given the party affiliation options, "Democrat," "Lean Democrat," "Lean Republican," "Republican."

[11]November 8, 2016 was election day, so respondents could already have voted. Since our sensitive statement of interest asks about "agreement" with candidates, not voting preferences, we think this does not present a critical problem.

[12]The list of policy preference questions can be found in the appendix.

Results

Our outcome variable of interest will be Total Affirmations. Recall that, for the Implicit treatment, this measure captures the total number of statements from the list with which a subject agrees. For the Explicit treatment, Total Affirmations equals the number of statements from the list with which a subject agrees plus one if the subject also agrees with the affirmation of agreement with the assigned candidate. If the observability of the response is irrelevant to the subject—that is, if SDR is not a motivation—then Total Affirmations should be equal across the two treatments.

Since the number of Total Affirmations depends on the responses to five different questions, it can be thought of as the sum of five random variables. Thus, our experiment requires large sample sizes to find statistical differences between treatments. When possible, we will control for demographic characteristics to improve our statistical power. In the appendix, we will include a robustness check where we drop all responses of zero or five in the Implicit elicitation since they fully reveal preferences. An additional concern is that subjects in the Implicit elicitation may have *wanted* to state 0 or 5 but avoided it, knowing that this response would have revealed their agreement with the candidate. Figures 19, 21, and 23 plot the histograms under the Explicit elicitations and show very few responses of 0 and 4.[13] Thus, we would expect few subjects in the Implicit elicitation to face this concern after adding the question of candidate agreement.

The p-values reported in this section are drawn from the regression of Total Affirmations onto the treatment assignment (Explicit vs. Implicit) and demographic controls. These regressions have indicator variables for each combination of treatment and candidate and do not contain a constant term so that we can directly interpret coefficient values. P-values for our tests of differences in differences are drawn from tests of differences in coefficients in this regression.

[13] In the Arkansas Poll, 6.8% of respondents state a 0 or 4 in the Implicit elicitation. In the first and second Mechanical Turk studies the percentages are 2.8% and 3.3%, respectively.

Experiment 1: Arkansas Poll

Recall that in this experiment, we only elicited agreement with Trump. We present summary statistics and Total Affirmations across the two treatments in Table 7. Comparing responses across treatments tests for the effect of SDR on statements of agreement with Trump. We drop 70 subjects whose responses did not provide enough information to calculate their number of Total Affirmations.[14] In our sub-group analysis, we drop 55 subjects whose sub-group could not be determined.

The uncontrolled difference reveals that subjects in the Explicit treatment are 3 percentage points less likely to express agreement with Trump ($p = 0.72$). This estimated difference rises to 4.5 percentage points ($p = 0.59$) with the inclusion of demographic controls. As an alternative measure of SDR, compare the difference between the list responses across treatments—which implies 52.9% agreement with Trump—to the explicit statements of agreement—49.9%.

Table 7: Arkansas Poll: Total Affirmations

	Explicit	Implicit	Difference
Explicit Agreement	0.499		
	(0.03)		
List-Response	1.941	2.469	0.529
	(0.05)	(0.06)	(0.08)
Total Affirmations	2.439	2.469	0.030
	(0.06)	(0.06)	(0.08)
Total Affirmations	2.173	2.218	0.045
(with controls)	(0.42)	(0.42)	(0.08)
N	730	730	

Heteroskedasticity-robust standard errors.
Controls: gender, age, income, and education.

Table 8 interacts each treatment with party identification in order to explore heterogeneity in the effect of SDR.[15] Figure 12 displays these results graphically, showing that Democrats express

[14] This could indicate attrition from the survey or refusal to answer relevant questions.

[15] While the level of agreement with a candidate could drive party identification, we are exploring the *differential*

Table 8: Arkansas Poll: Total Affirmations by Party Identification.

	Total Affirmations	
Democrat	2.207	2.571
	(0.12)	(0.41)
Democrat $\times$ Explicit	-0.293**	-0.307**
	(0.14)	(0.14)
Republican	2.974	3.309
	(0.10)	(0.41)
Republican $\times$ Explicit	0.075	0.114
	(0.14)	(0.14)
Independent	2.320	2.666
	(0.09)	(0.39)
Independent $\times$ Explicit	0.042	-0.000
	(0.14)	(0.14)
Dem $\times$ Exp $-$ Rep $\times$ Exp	-0.368*	-0.421**
Controls	No	Yes
N	675	675

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.
Controls: gender, age, income, & education.

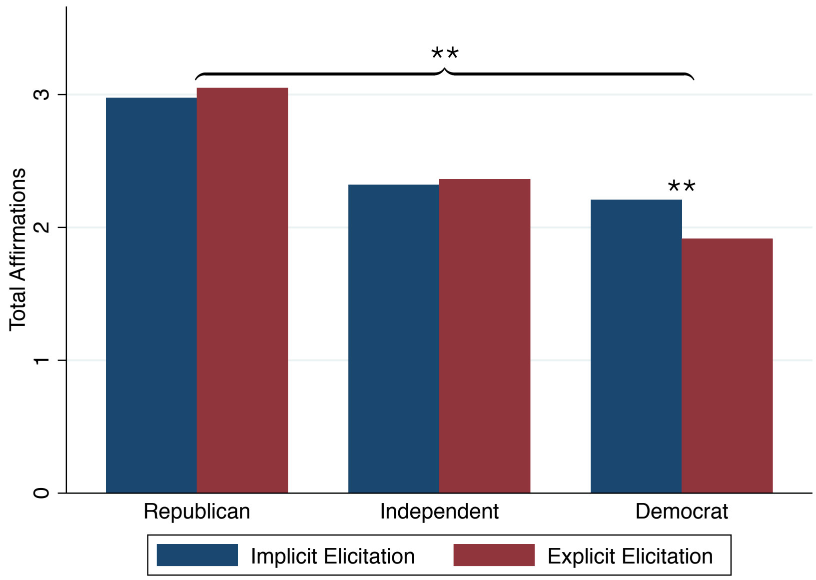
0.74 fewer Total Affirmations than Republicans, on average. When asked explicitly, Democrats' Total Affirmations drop by an additional 0.31 ($p = 0.03$) while Republicans increase their Total Affirmations by 0.11 ($p = 0.42$). This yields a difference in differences estimate of 0.421 ($p = 0.038$).

Experiment 2: Mechanical Turk Poll and Party Identification

In this experiment, we include affirmations of agreement with Clinton to compare the effect of SDR across candidates. In Table 9, we present the mean Total Affirmations for each combination of treatment and randomly assigned candidate. When questioned explicitly, subjects are relatively more likely to report agreement with Clinton and less likely to report agreement with Trump.

likelihood of expressing agreement explicitly. We believe this measure is sufficiently exogenous for use as a sub-group selection criteria.

Figure 12: Arkansas Poll: Mean of Total Affirmations split by treatment and party identification. Sensitive statement: "I often find myself agreeing with Donald Trump."



Individually, there is not a statistically significant effect of SDR on statements of agreement with either candidate, but when comparing the effect of SDR across the two candidates we estimate a marginally significant difference in differences of 0.193. This relatively greater effect of SDR on expressions of agreement with Trump can also be seen in the fact that his implied agreement (31.6%) outstrips his explicit agreement (22.9%), while the opposite is true of Clinton's implied (44.2%) and explicit (53.8%) agreement.

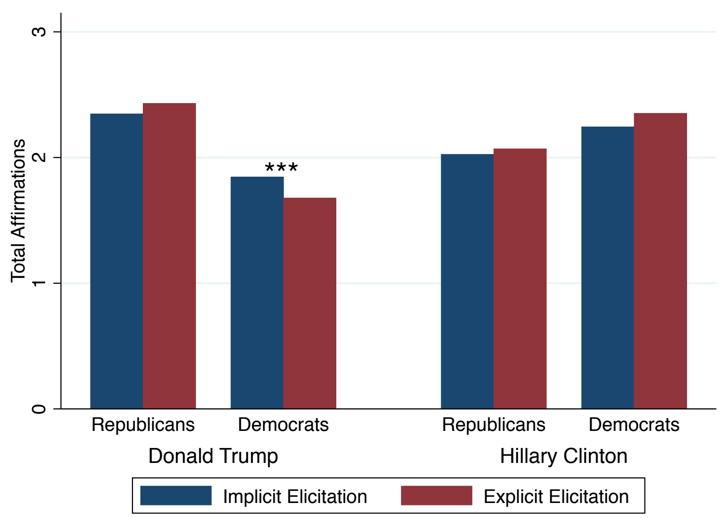
Table 9: Mechanical Turk Study 1: Total Affirmations

	Clinton Explicit	Clinton Implicit	Difference	Trump Explicit	Trump Implicit	Difference
Explicit Agreement	0.538			0.229		
	(0.03)			(0.03)		
List-Response	1.733	2.175	0.442	1.684	2.000	0.316
	(0.05)	(0.06)	(0.07)	(0.04)	(0.05)	(0.07)
Total Affirmations	2.271	2.175	0.096	1.913	2.000	-0.087
	(0.05)	(0.06)	(0.08)	(0.05)	(0.05)	(0.07)
Total Affirmations	2.125	2.036	0.090	1.754	1.857	-0.103
(with controls)	(0.70)	(0.69)	(0.08)	(0.70)	(0.70)	(0.07)
N	251	251		253	251	
Clinton Difference − Trump Difference: 0.193 ($p = 0.07$)						

Heteroskedasticity-robust standard errors. Controls for gender, age, & education.

We interact our treatment assignment with the subject's party identification to shed light on the origin of this differential effect of SDR. Table 10 shows the influence of SDR is greater on statements of agreement with the opposing party's candidate. Respondents in the Explicit treatment understate cross-party agreement relative to the Implicit treatment. Democrats, in particular, show a significant effect of SDR on their statements of agreement with Trump. Total Affirmations from Democrat respondents are 0.180 lower in the Explicit treatment ($p < 0.01$). SDR has little effect on statements of agreement with Clinton among Republicans. A separate subgroup analysis comparing SDR across own-party and opposing-party candidates shows that SDR increases significantly for statements of agreement with opposing-party candidates. The differential effect of the Explicit treatment on Total Affirmations is 0.213 larger for subjects assigned to their party's candidate instead of the opposing party's candidate ($p = 0.039$). This relies on the difference-in-differences approach—even though Republicans show no effect of SDR on statements of agreement with Clinton, there is a larger positive effect on explicit statements of agreement with Trump. Of course, the majority of this effect is driven by the impact of SDR on statements of agreement with Trump from Democrats.[16] Figure 13 repeats this analysis graphically.

Figure 13: Mechanical Turk Study 1: Subjects split by party identification and assigned candidate.



[16]In the appendix, we show that SDR is larger among the highly-educated. The difference-in-differences is 0.325 ($p = 0.010$).

Table 10: Mechanical Turk Study 1: Total Affirmations by Assigned Candidate and Party Identification.

	Total Affirmations	
Trump × Democrat	1.844	1.736
	(0.05)	(0.77)
Trump × Democrat × Explicit	-0.166**	-0.180***
	(0.07)	(0.07)
Trump × Republican	2.346	2.233
	(0.11)	(0.78)
Trump × Republican × Explicit	0.084	0.063
	(0.16)	(0.16)
Clinton × Democrat	2.244	2.146
	(0.07)	(0.77)
Clinton × Democrat × Explicit	0.108	0.102
	(0.09)	(0.09)
Clinton × Republican	2.025	1.903
	(0.09)	(0.78)
Clinton × Republican ×Explicit	0.044	0.034
	(0.14)	(0.14)
Own-Party × Exp − Out-Party × Exp	0.208**	0.212**
Controls	No	Yes
N	1,006	1,006

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors. Controls for gender, age, & education.

Experiment 3: Mechanical Turk Poll and Voter Ideology

In this study, we attempt to disentangle the effect of a subject's party identification from the ideology that might drive that identification. To do so, we use six questions about economic ideology adapted from Halpin and Agne (2009) to identify subgroups we will call "conservative" and "liberal."[17] Subjects are again assigned to evaluate their agreement with a randomly selected candidate.

[17]Questions were tailored to each candidate's policy positions. Subjects responded to each question on a 4-point scale. We label a subject conservative (liberal) if the sum of agreement with conservative (liberal) ideological statements exceeds the sum of the agreement with the liberal (conservative) statements. Our sample leans liberal: 72% identify more with liberal economic policy, 21% with conservative, and 7% are "moderate." Though, of course, much more than 7% of our respondents may consider themselves moderate. "Neutral" may be a more appropriate term.

Figure 14: Mechanical Turk Study 2: Respondents split by treatment, economic policy position, and randomly assigned candidate.

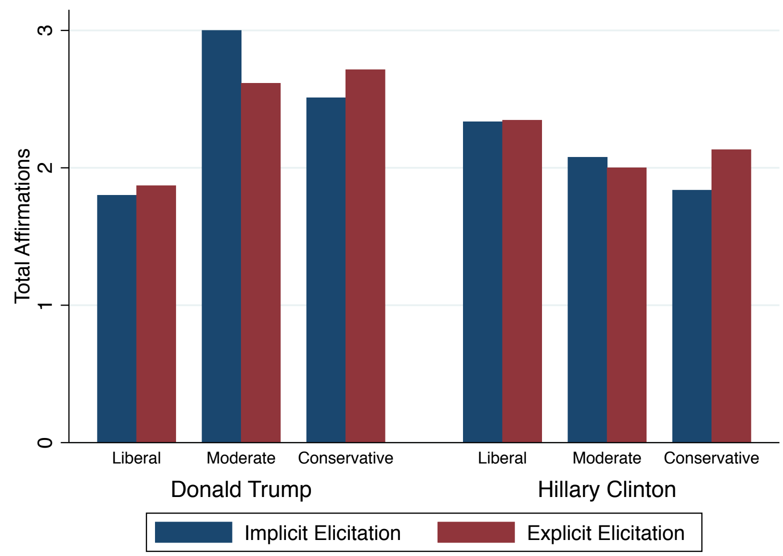


Table 11 interacts assigned candidate with the economic ideology of the respondent and plots Total Affirmations across the Implicit and Explicit treatments.[18] Ideological alignment does drive agreement with a given candidate; the more economically liberal (conservative) is a subject, the more Total Affirmations he or she reports when evaluating Clinton (Trump). However, there is no identifiable pattern to the differential effect of the explicit elicitation across ideologies or candidates. No individual comparisons are statistically significant at conventional levels. The lack of a systematic relationship is clear when looking at Figure 14. Thus, even though Experiment 2 clearly showed an association between party identification and SDR, these results suggest that this pattern is not driven by the underlying ideology that drove subjects to join their respective parties.

[18] Note that sample sizes are not balanced across the ideological bins.

Table 11: Mechanical Turk Study 2: Total Affirmations by Assigned Candidate and Economic Ideology.

	Total Affirmations	
Trump $\times$ Liberal	1.800	1.506
	(0.06)	(0.14)
Trump $\times$ Liberal $\times$ Explicit	0.070	0.078
	(0.08)	(0.08)
Trump $\times$ Moderate	3.000	2.703
	(0.34)	(0.36)
Trump $\times$ Moderate $\times$ Explicit	-0.385	-0.403
	(0.40)	(0.39)
Trump $\times$ Conservative	2.509	2.207
	(0.14)	(0.18)
Trump $\times$ Conservative $\times$ Explicit	0.205	0.180
	(0.20)	(0.19)
Clinton $\times$ Liberal	2.335	2.025
	(0.06)	(0.14)
Clinton $\times$ Liberal $\times$ Explicit	0.011	-0.002
	(0.09)	(0.09)
Clinton $\times$ Moderate	2.077	1.759
	(0.20)	(0.25)
Clinton $\times$ Moderate $\times$Explicit	-0.077	-0.029
	(0.34)	(0.34)
Clinton $\times$ Conservative	1.837	1.553
	(0.12)	(0.17)
Clinton $\times$ Conservative $\times$Explicit	0.295*	0.242
	(0.15)	(0.15)
Controls	No	Yes
N	985	985

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.
Controls: gender, age, and education.

Social-Signaling and County-Level Voting Data

One possible explanation for the origin of SDR is that local, in-person interactions create norms that evolve into socially desirable and undesirable behaviors. We use location data to explore the possibility that the influence of SDR on election polling may have geographic origins. Specifically, we will use a subject's geographic location as a proxy for his or her social setting. In our data, the county-level election winner was never determined by fewer than 77 votes, meaning that our subject's candidate preferences never influenced the outcome. Thus, we can use electoral outcome as an exogenous measure of a subject's social environment, yielding a clean estimate of the influence of environmental factors on SDR.

We use location data from each survey to merge our survey responses with county-level voting data.[19] We then test if a subject is relatively more likely to *explicitly* state agreement with the candidate who subsequently won the popular vote in the subject's county.[20]

Columns 1 and 2 of Table 12 interacts the treatment assignment with the candidate that won the respondent's county. In counties that Clinton won, the Explicit elicitation *increases* the likelihood of stating agreement with Trump, while in the counties that Trump won, the Explicit elicitation *decreases* that same likelihood.[21] Neither of these effects approach statistical significance. This could be a result of a relatively small sample—only 14 percent of our sample lives in Arkansas counties won by Clinton—but this pattern is the opposite of what would be expected if SDR were driven by county-level preferences. These results are graphically illustrated in Figure **??** in the appendix.

Columns 3 and 4 of Table 12 display results from the first Mechanical Turk poll. These results con-

[19]The Arkansas Poll collected each subject's county of residence; and our Mechanical Turk surveys collected geographic coordinates for each subject's IP address. IP addresses may not perfectly reflect the subject's place of residence, but should correlate with these, on average.

[20]These datasets were not collected simultaneously, since our experiments occurred before any voting took place. Thus, intervening events could weaken the connection between the two datasets.

[21]Clinton only won eight counties that appear in our data, making this a relatively low-powered test.

flict with the Arkansas Poll results and show that SDR toward Trump is manifest in the responses of subjects from counties that Clinton won. Subjects in these counties reveal 0.23 fewer Total Affirmations when asked about their agreement with Trump explicitly ($p = 0.017$). Comparing this effect to the same effect in counties that voted for Trump does not yield a significant difference in differences ($p = 0.277$). The results do not paint a clear picture, however, since explicit agreement with Clinton increases in counties that Trump won. Figures **??** and **??** repeat the Mechanical Turk analysis graphically in the appendix.[22]

Columns 5 and 6 of Table 12 repeat the analysis for the second Mechanical Turk poll. While no longer significant, these results are largely consistent with the results of the first Mechanical Turk poll. The Explicit treatment has an overall positive effect, but the effect is largest in both Clinton and Trump counties when they are explicitly stating agreement with their chosen candidate.

These interesting but conflicting results with respect to the geographic and social origins of SDR highlight a need for research designed specifically to address the role environment plays in developing social desirability. In particular, it will be useful for future studies to disentangle the local preferences for candidates from the local norms with regards to social desirability.

[22]Table 34 also repeats the analysis taking account of the margin of victory for each candidate. Under this specification, no coefficients reach conventional levels of statistical significance.

Table 12: Total Affirmations by County Voting Patterns

	Arkansas Poll		M-Turk 1		M-Turk 2	
Trump×Clinton-County	2.300	1.919	2.014	1.032	1.953	1.652
	(0.14)	(0.51)	(0.07)	(0.40)	(0.08)	(0.17)
Trump×Clinton-County×Exp	0.041	0.045	-0.217**	-0.235**	0.047	0.044
	(0.16)	(0.18)	(0.10)	(0.10)	(0.10)	(0.10)
Trump×Trump-County	2.492	2.142	1.988	0.993	2.127	1.764
	(0.07)	(0.48)	(0.09)	(0.40)	(0.12)	(0.19)
Trump×Trump-County×Exp	-0.046	-0.067	-0.042	-0.045	0.098	0.124
	(0.08)	(0.07)	(0.15)	(0.15)	(0.15)	(0.15)
Clinton×Clinton-County			2.193	1.220	2.148	1.825
			(0.07)	(0.39)	(0.08)	(0.18)
Clinton×Clinton-County×Exp			0.087	0.080	0.168	0.151
			(0.09)	(0.09)	(0.12)	(0.11)
Clinton×Trump-County			2.080	1.109	2.213	1.869
			(0.10)	(0.39)	(0.09)	(0.17)
Clinton×Trump-County×Exp			0.133	0.116	0.016	-0.003
			(0.13)	(0.13)	(0.12)	(0.12)
Controls	No	Yes	No	Yes	No	Yes
N	721	721	893	893	840	840

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors clustered by county where possible, otherwise by city.
Controls: age, gender, education. Income control added for Arkansas Poll.

Discussion and Conclusion

Our results reveal meaningful weaknesses in the current method of eliciting political preferences through explicit polling. In particular, we expose two mechanisms by which socially desirable responding can predictably cause electoral outcomes to deviate from predictions based on explicit poll numbers. Our paper largely agrees with Coppock (ming) in that the effect of SDR—if it exists—on the full population is limited. However, we find that SDR has a predictable *differential* impact on specific voters and specific candidates. In particular, it exaggerates differences in preferences between the two political parties. This exaggeration gives the false impression that Democrat and Republican voters have negligible overlap in their agreement with political candidates, leading to an underestimation of the likelihood of large swings in the electorate.

To determine the influence SDR might have on candidates' policy positions, we explore the connection between ideological agreement and SDR. We uncover a misalignment between the influence of SDR and a subject's ideological alignment with a candidate. Specifically, our data reject the claim that the influence of SDR decreases as ideological alignment increases. This presents a potential problem for the electorate: explicit statements of support for a politician are the primary means by which voters discipline policy choices between elections. So, when voters' willingness to explicitly reveal support for a candidate fails to respond to changes in the candidate's policy positions, their influence over policy evaporates.

Since we clearly identify a pattern of SDR associated with party affiliation, it is puzzling that ideology plays no important role in SDR, leaving open the question of what aspect of party affiliation drives SDR. Further research is required, but we believe that understanding party identity in the framework of identity economics (Akerlof and Kranton (2000)) may prove fruitful. The importance of cultural identity is underscored by our finding that SDR has the strongest influence on highly-educated respondents.

Finally, we use county-level voting data to look at the role environmental factors play in determining the social desirability of candidates. We find suggestive evidence that SDR might be more powerful when subjects are revealing agreement with the candidate who lost the popular vote in their county, though our results are conflicting. The geographic origins of SDR remain a compelling topic for future research, though follow-up studies will need more granular data—at the neighborhood level, for example—to explore a more nuanced concept of a voter's geographic region.

While our results cast doubt on the unbiased nature of explicit polling, our alternative methodology, implicit elicitation, is more complicated to administer and produces noisier estimates that require larger sample sizes. As such, there are clear limitations to when the implicit elicitation method can substitute for explicit polling. For instance, explicit polling will always be preferred when time or money are of particular concern. Instead of a complete replacement of explicit polling, we propose an alternative approach that respects the speed and simplicity of explicit polling while improving its accuracy through calibration using implicit polling. Specifically, we suggest that polling organizations conduct occasional, large-sample, implicit elicitations to detect bias in their polls. They then can recalibrate their explicit polling results according to the detected bias.

In developing a better understanding of the role SDR plays in political polling, this paper hopes to improve polling methodology and the reliability of forecasts derived from current and future polling methods. Moreover, we hope to provide evidence on the origins of SDR so that future social-science research can take into account respondent characteristics that make SDR an increasingly potent threat to the validity of poll results.

Appendix

Table 13: **Race**

Race	Number	Percent
White	653	82
Black	72	9
Hispanic	5	1
Asian	1	0
Native American	14	2
Multi-ethnic	21	3
Something else	10	1
Don't know	4	0
Refused	20	2
Total	800	100

*Source:*Arkansas Poll 2016

Table 14: **Political Affiliation**

Affiliation	Number	Percent
Republican	232	29
Democrat	199	25
Independent	295	37
Other	20	2
Don't Know	28	4
Refused	26	3
Total	800	100

*Source:*Arkansas Poll 2016

The effect of SDR is highest on statements of agreement with the candidate from the opposing party. Below, we repeat our primary regressions from our first Mechanical Turk survey. Here, we combine Democrats expressing agreement with Clinton and Republicans expressing agreement with Trump into "Own-Party" participants. The remaining participants are assigned to "Opposing-Party." The difference-in-differences effect of SDR on Own-Party candidates relative to Opposing-Party candidates shows that the effect of SDR is 0.21 points greater on statements of agreement with the Opposing-Party candidate ($p = 0.039$).

Figure 15: *Source:* Mechanical Turk: Nov. 1 Survey

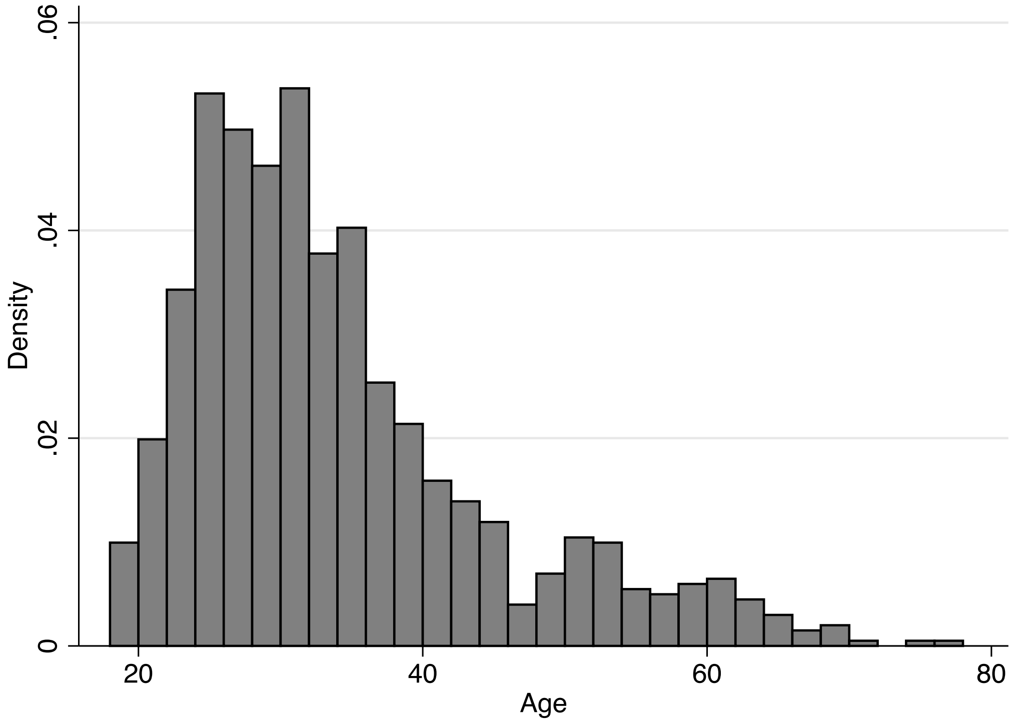


Figure 16: *Source:* Mechanical Turk: Nov. 8 Survey

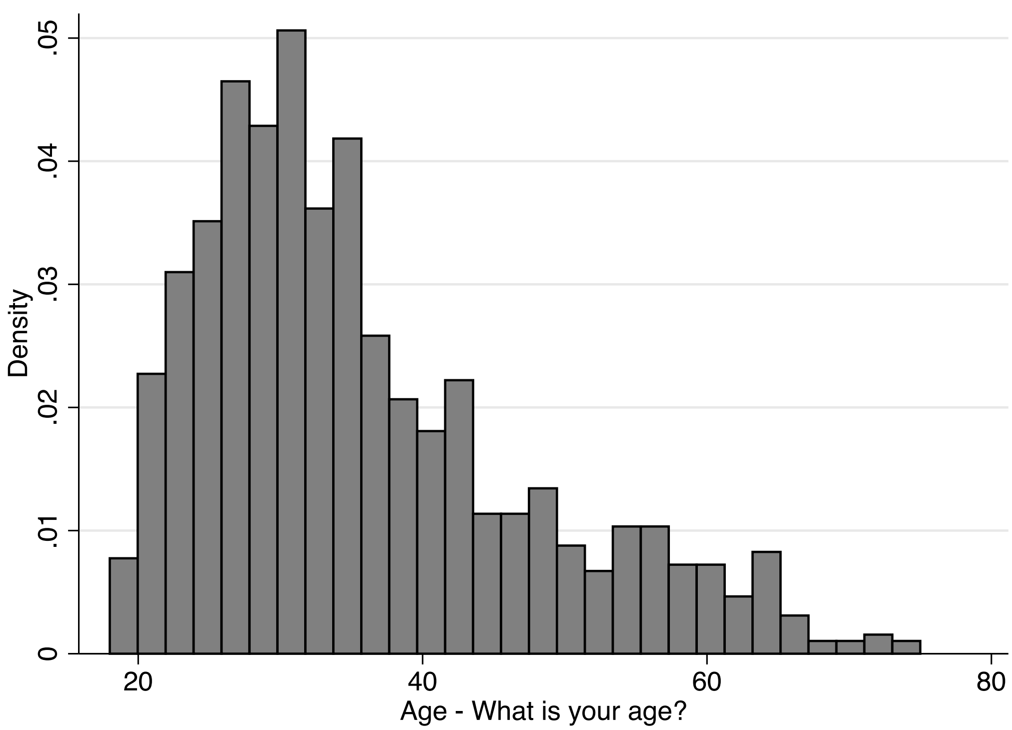


Table 15: **Gender**

Gender	Number	Percent
Male	357	45
Female	443	55
Total	800	100

*Source:*Arkansas Poll 2016

Table 16: **Education**

Education	Number	Percent
No High School	12	2
Some High School	75	9
High School Graduate	218	27
Some College Including Business or Trade School	193	24
College Graduate	150	19
Some Graduate School	28	4
Graduate or Professional Degree	104	13
Don’t Know	3	0
Refused	17	2
Total	800	100

Source: Arkansas Poll 2016

Table 17: **Income**

Income	Number	Percent
$7,500 or less	58	7
$7,501 to $15,000	57	7
$15,001 to $25,000	70	9
$25,001 to $35,000	76	10
$35,001 to $50,000	99	12
$50,001 to $75,000	94	12
$75,001 to $100,000	67	8
$100,001 or over	69	9
Don’t Know	51	6
Refused	159	20
Total	800	100

*Source:*Arkansas Poll 2016

Table 18: **Gender**

Gender	Number	Percent
Male	583	58
Female	423	42
Total	1,006	100

Source: Mechanical Turk: Nov. 1 Survey

Table 19: **Party**

Affiliation	Number	Percent
Democrat	371	37
Lean Democrat	327	33
Lean Republican	224	22
Republican	84	8
Total	1,006	100

Source: Mechanical Turk: Nov. 1 Survey

Table 20: **Education**

Education	Number	Percent
High School	3	0
Some College	88	9
College Degree	259	26
Some Graduate	99	10
Graduate or Professional Degree	413	41
Refused	144	14
Total	1,006	100

Source: Mechanical Turk Survey: Nov. 1

Table 21: Mechanical Turk 1 Balance Table

	Clinton × Explicit	Clinton × Implicit	Trump × Explicit	Trump × Implicit	(1) vs. (2)	(1) vs. (3)	(1) vs. (4)	(2) vs. (3)	(2) vs. (4)	(3) vs. (4)	Joint F-test
Male	0.594	0.549	0.608	0.566	0.331	0.755	0.544	0.206	0.714	0.365	0.578
	(0.033)	(0.033)	(0.034)	(0.033)							
Age	32.948	32.991	34.274	34.272	0.964	0.173	0.188	0.204	0.218	0.999	0.340
	(0.649)	(0.700)	(0.726)	(0.768)							
Education	4.310	4.226	4.311	4.184	0.480	0.991	0.269	0.485	0.726	0.275	0.629
	(0.081)	(0.088)	(0.085)	(0.079)							
Democrat	0.738	0.695	0.703	0.689	0.307	0.412	0.244	0.853	0.889	0.746	0.658
	(0.029)	(0.031)	(0.031)	(0.031)							

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22: **Gender**

Gender	Number	Percent
Male	544	55
Female	441	45
Total	985	100

Source: Mechanical Turk: Nov. 8 Survey

Table 23: **Education**

Education	Number	Percent
Some High School	11	1
High School Degree	108	11
Some College	280	29
Associate's Degree	101	10
Bachelor's Degree	383	39
Graduate Degree	102	10
Total	985	100

Source: Mechanical Turk Survey: Nov. 8

Table 24: **Party**

Affiliation	Number	Percent
Democrat	353	36
Lean Democrat	303	31
Lean Republican	211	21
Republican	118	12
Total	985	100

Source: Mechanical Turk: Nov. 8 Survey

Table 25: Mechanical Turk 2 Balance Table

	Clinton × Explicit	Clinton × Implicit	Trump × Explicit	Trump × Implicit	(1) vs. (2)	(1) vs. (3)	(1) vs. (4)	(2) vs. (3)	(2) vs. (4)	(3) vs. (4)	Joint F-test
Male	0.594	0.549	0.608	0.566	0.331	0.755	0.544	0.206	0.714	0.365	0.578
	(0.033)	(0.033)	(0.034)	(0.033)							
Age	32.948	32.991	34.274	34.272	0.964	0.173	0.188	0.204	0.218	0.999	0.340
	(0.649)	(0.700)	(0.726)	(0.768)							
Education	4.310	4.226	4.311	4.184	0.480	0.991	0.269	0.485	0.726	0.275	0.629
	(0.081)	(0.088)	(0.085)	(0.079)							
Democrat	0.738	0.695	0.703	0.689	0.307	0.412	0.244	0.853	0.889	0.746	0.658
	(0.029)	(0.031)	(0.031)	(0.031)							

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 26: Mechanical Turk Study 1: Total Affirmations by Own- or Opposing-Party Candidate

	Total Affirmations	
Own-Party	2.276	2.204
	(0.06)	(0.84)
Own-Party × Explicit	0.100	0.090
	(0.08)	(0.08)
Opposing-Party	1.901	1.820
	(0.04)	(0.85)
Opposing-Party × Explicit	-0.108*	-0.122*
	(0.07)	(0.06)
Trump × More Educated	1.963	0.993
	(0.06)	(0.26)
Trump × More Educated × Explicit	-0.102	-0.137
	(0.09)	(0.09)
Trump × Less Educated	2.067	1.427
	(0.09)	(0.18)
Trump × Less Educated × Explicit	-0.056	-0.033
	(0.13)	(0.13)
Clinton × More Educated	2.210	1.225
	(0.07)	(0.26)
Clinton × More Educated × Explicit	0.176*	0.188**
	(0.09)	(0.09)
Clinton × Less Educated	2.112	1.497
	(0.10)	(0.18)
Clinton × Less Educated ×Explicit	-0.065	-0.080
	(0.14)	(0.14)
Controls	No	Yes
N	1,006	1,006

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.

The effect of the Explicit elicitation is largest on respondents with a 2-year degree or higher education level. More educated respondents express significantly more agreement with Hillary Clinton when asked explicitly. The effect of the Explicit elicitation is significantly different across the two candidates for the more-educated respondents with a difference-in-differences of 0.325 ($p = 0.010$). The same difference-in-differences for less-educated respondents is -0.047 ($p = 0.803$).

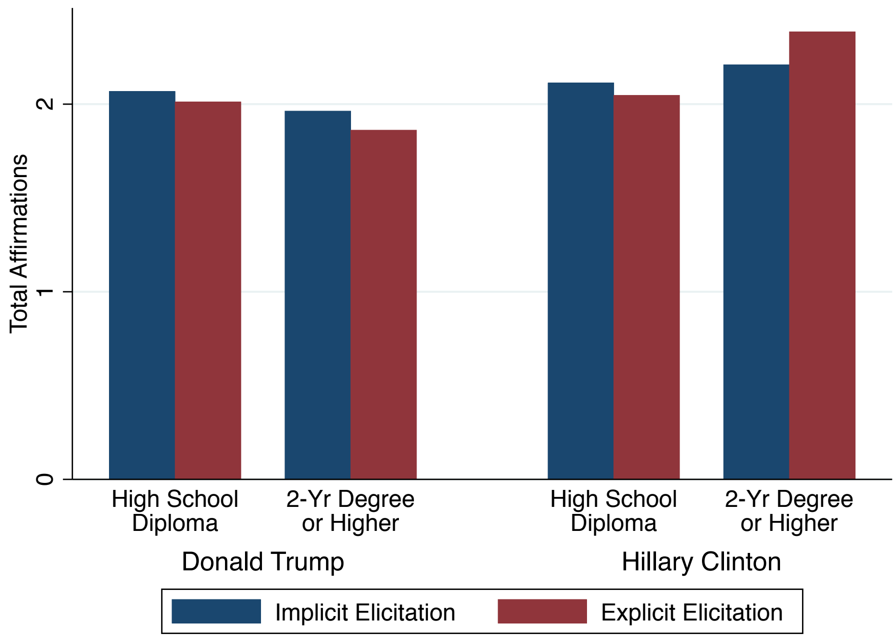
Table 27: Mechanical Turk Study 1: Total Affirmations by Education

	Total Affirmations	
Trump × More Educated	1.963	0.993
	(0.06)	(0.26)
Trump × More Educated × Explicit	-0.102	-0.137
	(0.09)	(0.09)
Trump × Less Educated	2.067	1.427
	(0.09)	(0.18)
Trump × Less Educated × Explicit	-0.056	-0.033
	(0.13)	(0.13)
Clinton × More Educated	2.210	1.225
	(0.07)	(0.26)
Clinton × More Educated × Explicit	0.176*	0.188**
	(0.09)	(0.09)
Clinton × Less Educated	2.112	1.497
	(0.10)	(0.18)
Clinton × Less Educated ×Explicit	-0.065	-0.080
	(0.14)	(0.14)
Controls	No	Yes
N	1,006	1,006

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.
Controls: gender, age, & education.

“Thank you. Let?s change things up again. I will read a list of five statements; I am interested in how many of the five you agree with. Rather than going item by item, please think about how many total statements you agree with and tell me that number when I?m finished, okay?” (INTERVIEWER: AVOID AN ITEM-BY-ITEM RESPONSE) (READ STATEMENTS) [RANDOMIZE ORDER OF STATEMENTS]

Figure 17: Subjects divided by party affiliation and assigned candidate.



1. I think small businesses are important for the economy.

2. I agree with George H.W. Bush's foreign policy.

3. I think the threat of global warming is exaggerated.

4. I often find myself agreeing with Donald Trump.

5. I prefer presidential candidates who oppose the NRA.

"How many of these statements do you agree with? "

- Questions about alignment with Hillary Clinton:
  - Government investments in education, infrastructure, and science are necessary to ensure America's long-term economic growth.
  - Government regulations are necessary to keep businesses in check and protect workers and consumers.
  - Rich people like to believe they have made it on their own, but in reality society has

Figure 18: AR Poll: Histogram of responses in Implicit treatment.

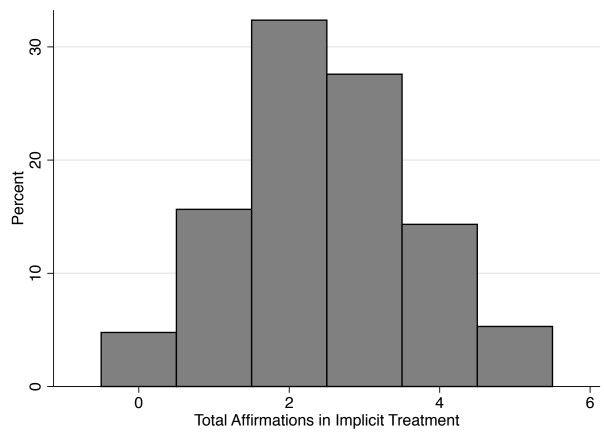


Figure 19: AR Poll: Histogram of responses from 4-item list in Explicit treatment.

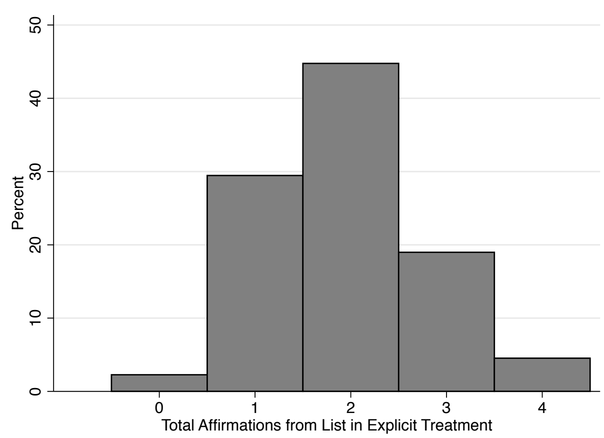


Figure 20: MTurk 1: Histogram of responses in Implicit treatment.

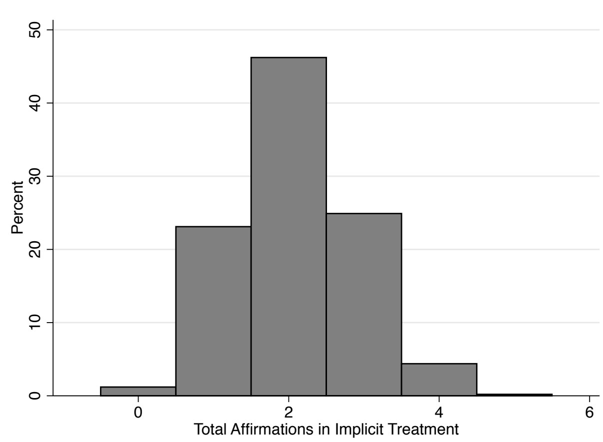


Figure 21: MTurk 1: Histogram of responses from 4-item list in Explicit treatment.

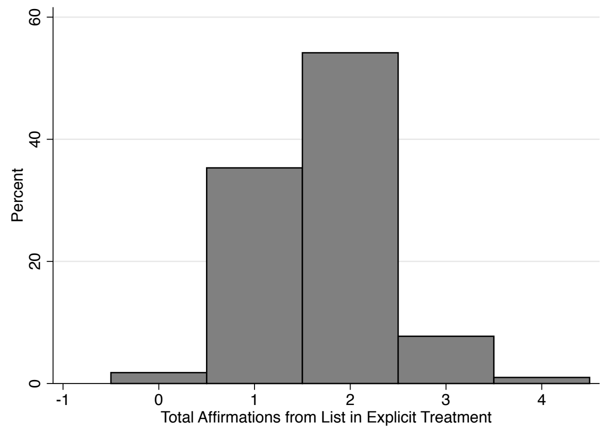


Figure 22: MTurk 2: Histogram of responses in Implicit treatment.

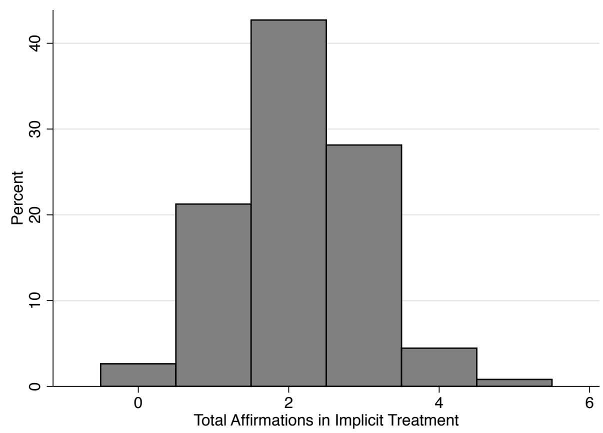


Figure 23: MTurk 2: Histogram of responses from 4-item list in Explicit treatment.

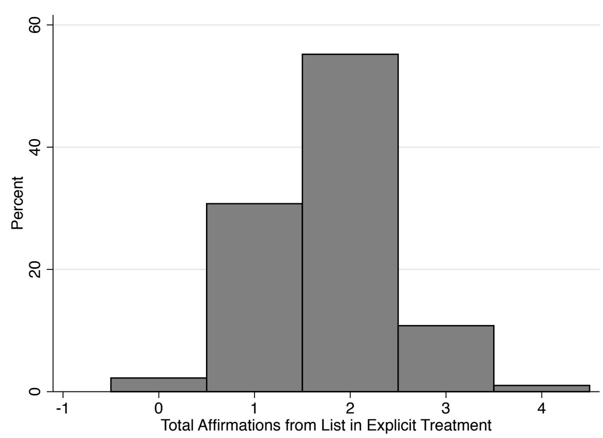


Table 28: Arkansas Poll Balance Table

	Implicit	Explicit	(1) vs. (2)
Male	0.470	0.440	0.430
	(0.027)	(0.027)	
Age	59.361	60.101	0.595
	(0.985)	(0.983)	
Educated	0.606	0.629	0.545
	(0.026)	(0.026)	
Democrat	0.231	0.265	0.303
	(0.022)	(0.024)	
Republican	0.318	0.301	0.615
	(0.025)	(0.025)	
Independent	0.400	0.360	0.281
	(0.026)	(0.026)	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 29: Arkansas Poll: Total Affirmations

	Total Affirmations
Explicit	2.439
	(0.06)
Implicit	2.451
	(0.05)
Controls	No
N	692

Heteroskedasticity-robust standard errors.

contributed greatly to their wealth.

- Questions about alignment with Donald Trump:
  - Government spending is almost always wasteful and inefficient.
  - Cutting taxes for individuals and businesses is the key to economic growth.
  - Immigrants today are a burden on our country because they take our jobs and abuse government benefits.

Table 30: Arkansas Poll: Total Affirmations by Party Affiliation

	Total Affirmations
Democrat	2.177
	(0.10)
Democrat × Explicit	-0.263**
	(0.13)
Republican	2.857
	(0.09)
Republican × Explicit	0.192
	(0.13)
Independent	2.304
	(0.08)
Independent × Explicit	0.058
	(0.13)
Controls	No
N	692

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.
Note: The difference in differences estimate is 0.455 ($p = 0.012$).

Table 31: Mechanical Turk Study 1: Total Affirmations

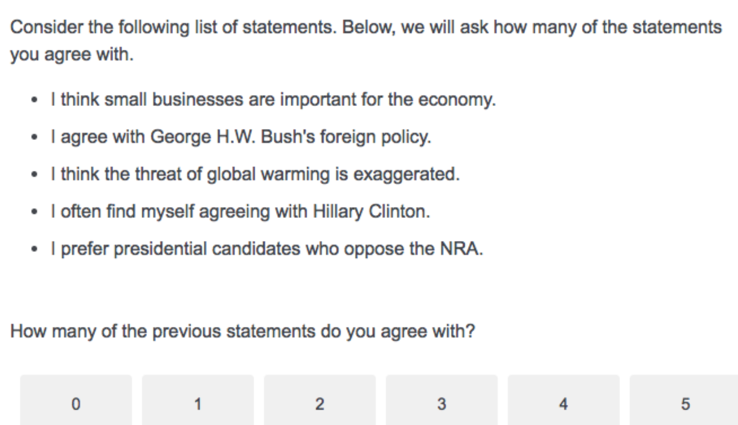
	Total Affirmations
Clinton	2.199
	(0.05)
Clinton × Explicit	0.072
	(0.07)
Trump	2.016
	(0.05)
Trump × Explicit	-0.103
	(0.07)
Controls	No
N	999

Heteroskedasticity-robust standard errors.
Note: The difference in differences estimate is 0.17 (p=0.097).

Table 32: Mechanical Turk Study 1: Total Affirmations by Assigned Candidate and Party Affiliation

	Total Affirmations
Trump $\times$ Democrat	1.865
	(0.05)
Trump $\times$ Democrat $\times$ Explicit	-0.187***
	(0.07)
Trump $\times$ Republican	2.346
	(0.11)
Trump $\times$ Republican $\times$ Explicit	0.084
	(0.16)
Clinton $\times$ Democrat	2.268
	(0.06)
Clinton $\times$ Democrat $\times$ Explicit	0.084
	(0.09)
Clinton $\times$ Republican	2.051
	(0.09)
Clinton $\times$ Republican $\times$Explicit	0.018
	(0.14)
Controls	No
N	999

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.


Consider the following list of statements. Below, we will ask how many of the statements you agree with.

- I think small businesses are important for the economy.
- I agree with George H.W. Bush's foreign policy.
- I think the threat of global warming is exaggerated.
- I often find myself agreeing with Hillary Clinton.
- I prefer presidential candidates who oppose the NRA.

How many of the previous statements do you agree with?

0 1 2 3 4 5

Figure 24: Screenshot from Qualtrics survey for Implicit treatment

Table 33: Mechanical Turk Study 2: Total Affirmations by Assigned Candidate and Economic Ideology

	Total Affirmations
Liberal×Clinton	2.360
	(0.0587)
Liberal×Clinton×Explicit	-0.0134
	(0.0858)
Moderate×Clinton	2.160
	(0.186)
Moderate×Clinton×Explicit	-0.160
	(0.337)
Conservative×Clinton	1.927
	(0.106)
Conservative×Clinton×Explicit	0.205
	(0.141)
Liberal×Trump	1.853
	(0.0520)
Liberal×Trump×Explicit	0.0171
	(0.0762)
Moderate×Trump	2.750
	(0.241)
Moderate×Trump×Explicit	-0.135
	(0.317)
Conservative×Trump	2.556
	(0.134)
Conservative×Trump×Explicit	0.159
	(0.196)
Controls	No
N	968

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Heteroskedasticity-robust standard errors.

Table 34: Total Affirmations and County Voting Margin

	Arkansas Poll	MTurk Study 1	MTurk Study 2
Trump	1.982	0.995	1.707
	(0.46)	(0.38)	(0.16)
Trump × Explicit	-0.051	-0.141	0.089
	(0.13)	(0.09)	(0.09)
Trump × Trump Vote-Margin	0.193	-0.138	0.032
	(0.27)	(0.15)	(0.18)
Trump × Trump Vote-Margin × Explicit	0.011	0.255	0.282
	(0.33)	(0.23)	(0.25)
Clinton		1.166	1.847
		(0.36)	(0.15)
Clinton × Explicit		0.094	0.073
		(0.08)	(0.08)
Clinton × Trump Vote-Margin		-0.074	-0.055
		(0.20)	(0.17)
Clinton × Trump Vote-Margin ×Explicit		0.109	-0.165
		(0.28)	(0.24)
Controls	Yes	Yes	Yes
N	721	890	851

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors clustered by county where possible, otherwise by city.
Controls: age, gender, education. Income control added for Arkansas Poll.

References

Agranov, M., J. K. Goeree, J. Romero, and L. Yariv (2017). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association*.

Ahern, M., C. Brown, and S. Dukas (2011). A national study of the association between food environments and county-level health outcomes. *The Journal of Rural Health 27*(4), 367–379.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of economic literature 43*(3), 762–800.

Allcott, H., R. Diamond, J.-P. H. Dubé, et al. (2017). *The geography of poverty and nutrition: Food deserts and food choices across the United States*. National Bureau of Economic Research.

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Andris, C., D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden (2015). The rise of partisanship and super-cooperators in the us house of representatives. *PloS one 10*(4), e0123507.

Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, Volume 30. Springer Science & Business Media.

Baldwin, K. and J. D. Huber (2010). Economic versus cultural differences: Forms of ethnic diversity and public goods provision. *American Political Science Review 104*(4), 644–662.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Berinsky, A. J. (1999). The two faces of public opinion. *American Journal of Political Science*, 1209–1230.

Bishop, G. F. and B. S. Fisher (1995). Secret ballots and self-reports in an exit-poll experiment. *Public Opinion Quarterly 59*(4), 568–588.

Bitler, M. (2014). The health and nutrition effects of snap: Selection into the program and a review of the literature on its effects.

Bitler, M. P. and J. Currie (2005). Does wic work? the effects of wic on pregnancy and birth outcomes. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management 24*(1), 73–91.

Blair, G. and K. Imai (2012). Statistical analysis of list experiments. *Political Analysis 20*(1), 47–77.

Blisard, N., B.-H. Lin, J. Cromartie, and N. Ballenger (2002). America's changing appetite: Food consumption and spending to 2020. *FOOD REVIEW-WASHINGTON DC- 25*(1), 2–9.

Brown-Iannuzzi, J. L., M. B. Najle, and W. M. Gervais (2017). The illusion of political tolerance: Social desirability and self-reported voting preferences. *Journal of Experimental Social Psychology*.

Bursztyn, L., D. Cantoni, P. Funk, and N. Yuchtman (2017). Polls, the press, and political participation: The effects of anticipated election closeness on voter turnout.

Charness, G. and U. Gneezy (2008). What's in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization 68*(1), 29–35.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Claassen, R. L. and J. B. Ryan (2016). Social desirability, hidden biases, and support for hillary clinton. *PS: Political Science & Politics 49*(4), 730–735.

Coffman, K. B., L. C. Coffman, and K. M. M. Ericson (2016). The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*.

Connors, E., S. Klar, and Y. Krupnikov (2016, November). There may have been shy trump supporters after all. *The Washington Post*.

Coppock, A. (forthcoming). Did shy trump supporters bias the 2016 polls? evidence from a nationally-representative list experiment. *Statistics, Politics, and Policy*.

Corstange, D. (2008). Sensitive questions, truthful answers? modeling the list experiment with listit. *Political Analysis 17*(1), 45–63.

Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

Droitcour, J., R. A. Caspar, M. L. Hubbard, T. L. Parsley, W. Visscher, and T. M. Ezzati (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 185–210.

Dropp, K. (2015, December). Why does donald trump perform better in online versus live telephone polling? *Morning Consult*.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics 112*(4), 1203–1250.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research.

Enten, H. (2016a, May). Americans? distaste for both trump and clinton is record-breaking. *FiveThirtyEight*.

Enten, H. (2016b, November). 'shy' voters probably arent why the polls missed trump. *FiveThirtyEight*.

Esteban, J. and D. Ray (2008). Polarization, fractionalization and conflict. *Journal of peace Research 45*(2), 163–182.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *Review of Economic Studies 83*(2), 587–628.

Fang, D., M. R. Thomsen, R. M. Nayga Jr, and A. M. Novotny (2018). Wic participation and relative quality of household food purchases: Evidence from foodaps. Technical report, National Bureau of Economic Research.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research 20*(2), 303–315.

Fox News (2016, January). See which candidates qualified for the fox news-google gop debates.

Goffman, E. (1959). *The presentation of self in everyday life*. Harmondsworth.

Gundersen, C., B. Kreider, and J. Pepper (2011). The economics of food insecurity in the united states. *Applied Economic Perspectives and Policy 33*(3), 281–303.

Gundersen, C., B. Kreider, and J. V. Pepper (2017). Partial identification methods for evaluating food assistance programs: A case study of the causal impact of snap on food insecurity. *American Journal of Agricultural Economics 99*(4), 875–893.

Hales, C. M., M. D. Carroll, C. D. Fryar, and C. L. Ogden (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. US Department of Health and Human Services, Centers for Disease Control and ?

Halpin, J. and K. Agne (2009). State of american political ideology, 2009. *Center for American Progress*.

Heerwig, J. A. and B. J. McCabe (2009). Education and social desirability bias: The case of a black presidential candidate. *Social Science Quarterly 90*(3), 674–686.

Hoffman, E., K. McCabe, and V. L. Smith (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review 86*(3), 653–660.

Hopkins, D. J. (2009). No more wilder effect, never a whitman effect: When and why polls mislead about black and female candidates. *The Journal of Politics 71*(3), 769–781.

Huang, J. and C. Low (2017). Trumping norms: Lab evidence on aggressive communication before and after the 2016 us presidential election. *American Economic Review 107*(5), 120–24.

Janus, A. L. (2010). The influence of social desirability pressures on expressed immigration attitudes. *Social Science Quarterly 91*, 928–946.

Jekanowski, M. D., J. K. Binkley, et al. (2000). Food purchase diversity across us markets. *Agribusiness 16*(4), 417–433.

Kane, J. G., S. C. Craig, and K. D. Wald (2004). Religion and presidential politics in florida: A list experiment. *Social Science Quarterly 85*(2), 281–293.

Kantchev, G. and C. Whittall (2017, June). After another surprise election result, investors grow skeptical of pollsters.

Karlan, D. S. and J. Zinman (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics 98*(1), 71–75.

Keeter, S. and N. Samaranayake (2007). Can you trust what polls say about obama's electoral prospects. *Pew Research Center*.

Kinsey, J. D. (1994). Food and families' socioeconomic status. *The Journal of nutrition 124*(suppl_9), 1878S–1885S.

Kreider, B., J. V. Pepper, and M. Roy (2018). Does the women, infants, and children program (wic) improve infant health outcomes?

Krysan, M. (1998). Privacy and the expression of white racial attitudes: A comparison across three contexts. *Public Opinion Quarterly*, 506–544.

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. Vishny (1999). The quality of government. *The Journal of Law, Economics, and Organization 15*(1), 222–279.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social forces 13*(2), 230–237.

Lax, J. R., J. H. Phillips, and A. F. Stollwerk (2016). Are survey respondents lying about their support for same-sex marriage? lessons from a list experiment. *Public Opinion Quarterly*, 510–533.

Lo, Y.-T., Y.-H. Chang, M.-S. Lee, and M. L. Wahlqvist (2012). Dietary diversity and food expenditure as indicators of food security in older taiwanese. *Appetite 58*(1), 180–187.

Maccoby, E. E. and N. Maccoby (1954). The interview: A tool of social science. *Handbook of social psychology 1*, 449–487.

Mas, A. and E. Moretti (2009). Racial bias in the 2008 presidential election. *The American Economic Review 99*(2), 323–329.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of public Economics 89*(11-12), 2325–2368.

Miller, J. D. (1984). A new survey technique for studying deviant behavior. *Unpublished doctoral dissertation, George Washington University, Washington, DC*.

Miller, J. M. and J. A. Krosnick (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, 291–330.

Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review 95*(3), 796–816.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology 46*(3), 598.

Pew Research Center (2012, May). Assessing the representativeness of public opinion surveys.

Pew Research Center (2016, June). Partisanship and political animosity in 2016.

Powell, R. J. (2013). Social desirability bias in polling on same-sex marriage ballot measures. *American Politics Research 41*(6), 1052–1070.

Raghavarao, D. and W. T. Federer (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40–45.

Redlawsk, D. P., C. J. Tolbert, and W. Franko (2010). Voters, emotions, and race in 2008: Obama as the first black president. *Political Research Quarterly*.

Rothschild, D. (2009). Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly 73*(5), 895–916.

Rudman, L. A. and S. E. Kilianski (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin 26*(11), 1315–1328.

Santos, F. C., M. D. Santos, and J. M. Pacheco (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature 454*(7201), 213.

Schillinger, D., L. R. Barton, A. J. Karter, F. Wang, and N. Adler (2006). Does literacy mediate the relationship between education and health outcomes? a study of a low-income population with diabetes. *Public health reports 121*(3), 245–254.

Shepard, S. (2016, October). Gop insiders: Polls don't capture secret trump vote. *Politico*.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics 118*, 26–40.

Streb, M. J., B. Burrell, B. Frederick, and M. A. Genovese (2008). Social desirability effects and support for a female american president. *Public Opinion Quarterly 72*, 76–89.

Traugott, M. W. and V. Price (1992). A review: exit polls in the 1989 virginia gubernatorial race: Where did they go wrong? *Public Opinion Quarterly 56*(2), 245–253.

Tsuchiya, T., Y. Hirai, and S. Ono (2007). A study of the properties of the item count technique. *Public Opinion Quarterly 71*(2), 253–272.

Ver Ploeg, M. (2010). *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to Congress*. Diane Publishing.

Vespa, J., D. M. Armstrong, and L. Medina (2018). Demographic turning points for the united states: population projections for 2020 to 2060. *Washington, DC: US Census Bureau 42*.

Wagner, A., R. J. Zeckhauser, and A. Ziegler (2017). Company stock reactions to the 2016 election shock: Trump, taxes and trade. Technical report, National Bureau of Economic Research.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association 60*(309), 63–69.

Wolfers, J. and E. Zitzewitz (2004). Prediction markets. *The Journal of Economic Perspectives 18*(2), 107–126.

Wolfers, J. and E. Zitzewitz (2016). What do financial markets think of the 2016 election. Technical report, Working paper.

Yen, S. T. (2010). The effects of snap and wic programs on nutrient intakes of children. *Food Policy 35*(6), 576–583.

Chapter 3: WIC Participation and Relative Quality of Household Food Purchases: Evidence from FoodAPS

Abstract

We examine the effect of the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on the quality of household food purchases using the National Household Food Acquisition and Purchase Survey (FoodAPS) and propensity score matching. A healthy purchasing index (HPI) is used to measure nutritional quality of household food purchases. WIC foods explain the improvement in quality of food purchases, not self-selection of more nutrition-conscious households into the program. The improvement in purchase quality was driven entirely by WIC participating households who redeemed WIC foods during the interview week. There was no significant difference between WIC-participants who did not redeem WIC foods and eligible non-participants. In this sample, there is no evidence that lack of access to clinics has adverse effects on participation nor is there evidence that HPI depends on supermarket access. A supervised machine learning process supports our main conclusion on the importance of WIC foods. Di Fang, Michael R. Thomsen, and Rodolfo M. Nayga, Jr.. Di Fang is Assistant Professor, Michael R. Thomsen is Professor, and Rodolfo M. Nayga Jr. is Distinguished Professor and Tyson Chair in Food Policy Economics all in the Department of Agricultural Economics and Agribusiness, 217 Agriculture Building, University of Arkansas, Fayetteville, Arkansas 72701, USA. Address correspondence to Di Fang, difang@uark.edu. This research was supported by USDA grant no. 59-5000-5-0115 to the National Bureau of Economic Research, entitled, ”Using FoodAPS for Research in Diet,

Health, Nutrition, and Food Security." This version was cleared by Xingyou Zhang, Ph.D..

Introduction

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) is the nation?s third largest food assistance program (Morgan 2015) but focuses narrowly on pregnant, postpartum, or breastfeeding women; infants; and children up to five years of age. The program provides food assistance, nutrition education, breast feeding support, and referrals to health and other services (USDA Food and Nutrition Service (FNS) 2016). WIC foods are intended to be supplemental and address nutritional gaps in the recipients? diets[23]. Participation in the program involves frequent visits to a WIC clinic for other services. Typically, at least one clinic appointment is required within a three-month period.

The program began in 1974 and its reach has grown with time. Recent estimates indicate that WIC serves half of U.S. infants and close to 30 percent of children, pregnant women, and postpartum women (Oliveira and Frazo 2015). WIC targets lower income households, those with incomes below 185 percent of the federal poverty guidelines, but participants in some higher-income households are adjunctively eligible by having previously qualified for participation in another assistance program such as Medicaid (Thorn et al. 2015). That said, in 2014 only a small fraction, less than two percent, of participants were from households with income over 185 percent of the federal poverty level. The overwhelming majority, nearly three quarters, were in households with income below 100 percent of the poverty threshold (Thorn et al. 2015). That WIC is associated with improvements in birth, health, and nutritional outcomes has been documented in numerous studies, with birth outcomes receiving the most attention. Reviews of this literature include Owen and Owen (1997), Fox et al. (2004), Devaney (2010) and Black et al. (2012). That said, selection bias is

[23] 1.In 2012 and 2013, the average dollar value of WIC foods was 45.00*and*43.26 per recipient per month, respectively (USDA FNS 2018). Benefits by household would be higher because there can be more than one WIC beneficiary per household. The average household size of WIC participants in April 2014 was 4.1 persons (USDA FNS 2017).

a concern in this work because nearly all studies compare participants to eligible non-participants (Gordon and Nelson 1995; Besharov and Germanis 2001). If participation is more attractive to those who are concerned about nutrition and/or otherwise exhibit behaviors contributing to better health, then the beneficial effects of the program could be overstated or spurious. Conversely, there is evidence that participants in WIC are more likely to have characteristics associated with poor health outcomes relative to eligible nonparticipants, which could result in published findings that understate beneficial effects of the program (Bitler and Currie 2005a).

Nevertheless, the emerging evidence is that the effects of WIC on birth outcomes are robust to selection bias (Currie and Rossin-Slater 2015). Recent studies have examined birth outcomes using empirical strategies designed to address selection bias and continue to show that WIC is beneficial. Strategies include models with maternal fixed effects (Sonchak 2016; Currie and Rajani 2014) and instrumental variables models (Gai and Feng 2012). Others exploit the staggered deployment of WIC across counties during the early stages of the program (Hoynes, Page, and Stevens 2011) and compare outcomes from mothers transitioning into and out of the program over multiple births (Figlio, Hamersma, and Roth 2009). There has been less direct attention on the role of selection bias in estimating the effect of the program?s nutritional outcomes. Overwhelmingly, studies that have examined nutritional outcomes find that WIC participation is associated with improvements in dietary quality, nutrient intakes, and/or biochemical indicators of nutritional adequacy across a number of different datasets and time periods (see Fox et al. 2004; Black et al. 2012). The evidence of nutritional improvements is noteworthy because WIC foods are prescribed to the individual. Once the foods are in the home, other household members could consume them resulting in program leakage. It could also be that WIC households are healthier. Because of the selection issue, it is possible to argue that those attracted into the program would have purchased the supplemental WIC foods anyway, and that despite the large body of evidence showing a strong association between the program and improved diet, the program?s actual benefits in terms of meaningfully augmenting nutrition are limited.

Assessing how selection might explain the nutritional outcomes of WIC is especially important now. WIC foods account for about 70 percent of program costs (Oliveira and Frazo 2015) and WIC is one of several nutrition and health programs targeted for cuts in a recent White House budget proposal (Aisch and Parlapiano 2017). The supplemental nature of WIC foods has led some to question whether the WIC food packages are sufficiently meaningful to the overall diet of pregnant women to alter birth outcomes and even whether the birth outcomes being considered would be sensitive to the level of nutrition supplementation that WIC provides (Joyce, Gibson, and Coleman 2005). Others have conceded the possibility that the food packages may be an incentive that induces pregnant women to participate in the program and that it is other program features, such as education and health referrals, that could be responsible for mitigating the likelihood of poor outcomes (Bitler and Currie 2005b)[24]. Thus, evidence on whether WIC foods matter is important to understanding the mechanisms by which the program leads to better health. The question we address in this paper is whether participation in WIC meaningfully alters food choices in a way that would be conducive to improvements in diet. We address this question using USDA?s National Household Food Acquisition and Purchase Survey (FoodAPS). FoodAPS provides a relatively small but nationally representative sample of U.S. households and includes information about where households shop for food, the availability and types of food stores in the communities where these households are located, and household eligibility and participation for food assistance programs (USDA Economic Research Service (ERS) 2017a). Additionally, FoodAPS contains detailed information about factors that may influence food purchases such as socioeconomic characteristics of the household, number of individuals residing within the household, and racial composition. An advantage of the FoodAPS data is that they permit us to look at food purchases directly and see how they differ among participants who did and did not use WIC to purchase foods. Since the sample was not stratified by time of month or date of delivery of WIC benefits, the samples of WIC households using or not using their WIC benefits should be nearly

[24] Since this debate, WIC food packages have been updated. The revised food packages include new food categories (e.g., fruits and vegetables, whole-grains, and infant foods), adjusted maximum purchase quantities, and optional food substitution policies to accommodate dietary behaviors of ethnic groups (USDA FNS 2011). The revised food packages were implemented in 2009, several years before FoodAPS.

random.

To determine how WIC contributes to the nutritional quality of household food purchases, we use an adaption of the Healthy Eating Index (HEI), a broadly accepted measure for the overall quality of an individual?s diet and apply it to each household?s food purchases. Because we are assessing purchases and not dietary intake, we term this measure the healthy purchasing index (HPI). FoodAPS provides a detailed record for participating households during an interview period (typically a week). These purchase records include information about what foods were purchased, where they were purchased, and whether purchases were made using benefits from nutrition assistance programs such as WIC. Moreover, FoodAPS includes nutrient values for the food items contained in the household?s purchase record. This permits application of the HEI over these purchases to obtain the HPI.

Given the lack of data on a valid instrumental variable that could be used to help identify the effect of WIC participation, we use propensity scores to match WIC participating households to eligible non-participating households. We then estimate the average treatment effect on the treated (ATT) and show that WIC participation is associated with modest but statistically significant improvements in nutritional quality of household food purchases. This finding is consistent with earlier work showing an association between WIC participation and improved nutritional outcomes. Next, we compute the ATT among participants who redeemed WIC benefits for foods during the interview week and those who did not. We find that the improvement in nutritional quality of food purchases is driven entirely by households who redeemed WIC foods. We found no meaningful program effect on the purchases of WIC participating households who did not redeem WIC foods. In sum, we present evidence that WIC foods are the most plausible explanation for earlier findings of a positive association between WIC participation and nutritional outcomes. Moreover, we find no evidence to support the contention that this association can be explained by systematic differences among WIC and comparable non-WIC households.

We conduct several robustness checks on this finding. First, food retailers who accept WIC benefits as payment carry healthier foods. It is plausible that differences in shopping venue could explain this finding. Hence, we redo the analysis using a sample comprised only of households who shopped at a WIC approved retailer during the interview week and reach the same conclusion. WIC foods continued to explain the difference in purchase quality in this follow-up analysis.

Second, our finding could be due to a secondary selection issue wherein some WIC-participating households have characteristics that make them more likely to fully redeem WIC benefits. To assess this, we exclude food items procured on shopping trips where WIC accounted for a majority of the value of a household?s purchases, recompute the HPI, and redo the analysis. When the WIC shopping events are excluded, the ATT among those households who redeemed WIC foods is no longer significant. Again, this supports the conclusion that WIC foods explain the improvement in nutritional quality, not self-selection of more nutrition-conscious households into the program.

While propensity score matching is a well-established method, one weakness is that the researcher must specify a relationship between the likelihood of program participation and observed covariates. As a third robustness check, we use an ?honest? random forest that places no restrictions on model complexity but that does penalize overfitting to derive the ATTs (Athey, Tibshirani, and Wager 2017; Athey 2017; Wager and Athey 2018). This approach uses machine learning methods to identify proximity of participating and non-participating households and thereby sidesteps the requirement of a researcher-supplied functional specification.

A secondary aim of this paper is to use the rich information FoodAPS contains about the commercial food environment to understand whether geographic barriers impact WIC participation. Specifically, we look at whether inadequate access to WIC clinics limits participation and whether the HPI differs meaningfully for participants without access to supermarkets. We find little evidence from the FoodAPS data that these barriers pose significant hurdles to program participation and effectiveness.

Data and Methods

FoodAPS contains a total of 4,826 households who completed the survey between April 2012 and January 2013 (USDA ERS 2017a). Among the 4,826 households who took part in FoodAPS, there were 1,007 households with at least one member who was categorically eligible for WIC and who met other program requirements for income or adjunctive eligibility through participation in Medicaid or other qualifying assistance program. Our focus is on this subsample of FoodAPS households. Of these 1,007 eligible households, 461 households were participants in WIC. Households recorded purchases in food-at-home and food-away-from-home food diaries. We further restricted the WIC-eligible sample to those households with at least one food-at-home event during the interview period and those that constituted complete cases over all measures reported in Table 1.

As shown in Table 1, our analysis sample includes 928 households. Of these, 505 households were eligible for WIC but did not participate in the program and 423 participated. Of the 423 participating households, 152 used WIC benefits on one or more purchase occasions during the interview period and 271 did not. It is important to emphasize that household purchases are only observed during the week of the interview. Thus, an observation that a WIC household did not use WIC for purchases cannot be taken as evidence that the household failed to redeem its WIC benefits. All we can determine is that benefits were not used during the week in question.

FoodAPS contains information about household characteristics that may affect WIC participation and food choice. As shown in Table 1, these characteristics include educational attainment, monthly income, marital status, presence of different categories of WIC-eligible individuals, and household racial and ethnic composition. Most of these measures are based on the characteristics of individuals in the household, which are then aggregated up to the household level[25].

[25] Educational attainment reflects the highest attainment of anyone in the household. The number of WIC infants and WIC children is based on individuals aged 0 to 1 and 1 to 4, respectively. A binary measure is used to indicate the presence of a WIC eligible woman. This is set to one if there was a pregnant woman in the household, there was an

In addition, we are able to examine two geographical barriers: access to supermarkets and access to WIC clinics. Supermarket access could be one barrier that influences both participation in WIC and food choices in general. To designate households without easy access to supermarkets, we use USDA?s tract-level measure indicating limited access to supermarkets based on vehicle travel (USDA ERS 2017b). Second, as demonstrated by Rossin-Slater (2013), access to WIC clinics is another determinant of WIC take-up. Since the FoodAPS dataset does not provide information on clinic access, we assembled a list of WIC clinic locations. This involved collection of data on clinic locations across numerous state WIC agencies. To correspond temporally to the FoodAPS data collection period, we used 2012-2013 locations if available, but only the 2015-2016 locations were available for some states. We supplied the resulting geocodes for WIC clinics to USDA-ERS personnel who spatially joined the clinic locations to the FoodAPS households and provided a file with radial distances from each household to the nearest WIC clinic within that household?s state of residence. We then measured clinic access as a binary variable taking the value of one if the household was within one (ten) miles of a clinic and located in an urban (rural) tract[26].

Measuring the Healthy Purchasing Index (HPI)

As noted earlier, the HPI measure we use for nutritional quality is based on the HEI. Specifically, our measure is based on the HEI-2010, which reflects diet quality in terms of conformance to the 2010 Dietary Guidelines for Americans (USDA FNS 2010). The HPI measure we use differs primarily in that (a) it is computed over food purchases as opposed to food intake and (b) it is measured at the household as opposed to the individual level. Nevertheless, this measure still provides a baseline measure of a household?s ability to meet dietary guidelines from its food purchases.

infant being breastfed, or if there was a birth within the household within the last three months. There is the potential for underreporting WIC-eligible women since WIC provides postpartum benefits for six months but FoodAPS flags birth events within the past three months.

[26]Our use of one and ten mile radii for urban and rural tracts is analogous to the definition used to identify limited access to supermarkets for purposes of defining food deserts (see USDA ERS 2017b).

The HEI, and by extension the HPI used here, assesses 12 dietary components (Guenther et al. 2013). These include nine adequacy components and three moderation components. The adequacy components reflect (1) total fruit, (2) whole fruit, (3) total vegetables, (4) greens and beans, (5) whole grains, (6) dairy, (7) total protein foods, (8) seafood and plant proteins, and (9) fatty acids. The moderation components include (1) refined grains, (2) sodium, and (3) empty calories. A higher value for each component of the index indicates a healthier nutrient intake. Moderation components are inversely coded so that lower consumption of these items leads to higher scores. The overall HPI measure computed over these 12 categories ranges from 0 to 100. Again, higher values of the overall index indicate healthier food purchases.

The FoodAPS survey collected detailed information about all foods purchased by the household over the course of seven days. The primary respondent in each household participated in two in-person interviews and up to three telephone interviews (Ver Ploeg et al. 2015). The dataset contains information on food items purchased or otherwise acquired, including brand, and package size, which allowed items to be matched to the USDA Food and Nutrient Database for Dietary Studies or the USDA National Nutrient Database for Standard Reference (USDA ERS 2017a). Economists at the USDA?s Economic Research Service developed computer code and intermediate datasets that aggregated these nutritional values into the HEI components and computed an overall index value at the household level. These datasets and computer code were then made available for our use in this research and were used to compute the HPI index we use as the outcome measure in this study.

Characteristics of the WIC and Eligible Non-WIC Samples

As shown in Table 1, the mean HPI score for WIC and eligible non-WIC households are virtually identical at 50.259 and 50.388, respectively. The last two columns of the table break the WIC households down into those who redeemed WIC benefits during the sample period and those who did not. Here the difference in HPI is meaningful. The average HPI computed over those who redeemed WIC foods is much higher at 55.958 in comparison to the average of 47.062 computed over those who did not. This is not surprising because foods provided by WIC help participants meet dietary guidelines. Nevertheless, the apparent importance of WIC foods to the magnitude of the index is striking and will receive further attention below. One thing that is noteworthy from Table 1 is that there are important differences between the WIC and eligible non-WIC samples. The average number of WIC-eligible individuals is similar between the two groups but on average, WIC households contain higher numbers of infants and eligible women than did non-WIC households. WIC households have lower levels of educational attainment, a lower percentage of married couples, and much lower household incomes. The proportion of Hispanic households is much higher among the WIC sample and the proportion of African American households is slightly higher. Finally, access to WIC clinics is higher among WIC households (40.4 percent compared to 35.8 percent) but a larger proportion of WIC households did not have access to a supermarket (22 percent compared to 14.5 percent).

To assess households? subjective evaluation of dietary quality, FoodAPS included the primary respondent?s self-assessment of whether the household is following a healthy diet. As shown in Table 1, a slightly higher percentage, 40.9 percent, of WIC households reported a healthy diet in comparison to 38.2 percent of the non-WIC households. Nevertheless, given the differences between the WIC and non-WIC samples, the interesting question remains whether WIC truly improves the healthiness of food purchase for participants or whether participants self-select into WIC because of preferences for healthier foods such as those provided by WIC. To answer this question, we use a matching algorithm to estimate the ATT of WIC participation.

Matching WIC Households to Eligible Non-WIC Households

As noted earlier, estimating the impact of WIC on nutritional quality is difficult with observational data such as FoodAPS because the treatment selection (in this case, WIC participation) is often influenced by subject characteristics. Consequently, baseline characteristics of treated subjects could differ systematically from those of untreated subjects. Therefore, to understand the effect of participation in WIC on HPI, we must first account for the systematic differences in baseline characteristics between WIC participants and eligible non-participants.

Matching methods provide a way to reduce selection bias among observational data (Rosenbaum 2002; DiPrete and Gangl 2004). The goal is to find a group of non-treated individuals who are similar to the treated individuals in all baseline characteristics ? then focus attention on estimating the effect of interest and consider all variables other than the treatment variable as potentially confounding. Balancing the vector of characteristics across treatment reduces the influence of confounding variables. Therefore, matching mimics a randomized experiment (conditional on a set of observables) so that the effect of the treatment is established (Drichoutis, Nayga, and Lazaridis 2009).

Matching methods are discussed at length in Rosenbaum and Rubin (1983). The matching algorithm we use constructs an artificial control group among eligible non-WIC households that have similar characteristics as those of WIC participants. Let T_i indicate the treatment, which equals one if household i participated in WIC (treated case) and zero if household i is WIC-eligible but does not participate in WIC (control case). Define HPI outcomes as Y_0i and Y_1i for the associated treatment status 0 and 1. The treatment effect for an household i can be written as: $t_i = Y_11 - Y_0i$. However, we do not know t_i because we can only observe the outcome of either t_1 or t_0 (we only observe $Y_i = t_iY_1i + (1 - t_i)Y_0i$), but not both. Therefore, we can estimate the average treatment effect on the treated (ATT) as $t_ATT = E(Y_1 \| T = 1) - E(Y_0 \| T = 1)$. Understanding the effect of WIC participation on those who ultimately participated is the relevant policy question in our study, not the effect of WIC participation averaged across all households regardless of whether or not they participated in WIC. Hence, we focus on the estimation of ATT in our analyses.

Notice that the term $E(Y_0 \| T = 1)$ is not observed because we do not observe the WIC effect on households who are not on WIC. Moreover, if one tries to substitute this with $E(Y_0||T=0)$, it would lead to self-selection bias (Drichoutis, Nayga, and Lazaridis 2009). We can assume that selection into treatment depends on observable covariates as long as the following two strong ignorability assumptions in treatment assignment are satisfied: (1) $\{(Y\}_1 i, Y_0 i) \perp T_i | X_i$; and (2) $0 < p(X) < 1$. The first condition implies that selection is solely based on observable characteristics and that all variables influencing treatment assignment and potential outcomes simultaneously are observed by the researcher (Caliendo and Kopeinig 2008). The second ensures a common support (to rule out perfect predictability of treatment given X) between the treatment and control groups. Rosenbaum and Rubin (1983) further demonstrated that under the assumptions of strong ignorability, treatment and control groups are exchangeable. The average treatment effect for the treated is estimated as $t_A TT = E\{E(Y_1||X, T=1) - E(Y_0||X, T=0)||T=1\}$, where the outer expectation is taken over the distribution of baseline covariates in the treated group (Rosenbaum and Rubin 1983). Therefore, outcome analysis on the matched data tends to produce unbiased estimates of treatment effects due to reduced selection bias through the balancing of the distributions of observed covariates.

In this study, we use nearest neighbor propensity score matching (PSM). Propensity score matching has been popularly applied in economics, statistics and medical research (Hong and Yu 2008; Ye and Kaskutas 2009; Wyse, Keesler, and Schneider 2008; Staff et al. 2008). PSM forms matched sets of treated and untreated subjects who share a similar probability to be treated (Rosenbaum and Rubin 1983; 1985). We use a logistic regression model to estimate the propensity score. Specifically, WIC participation status is regressed on observed household characteristics. Afterwards, we use the ?Matching? package in R to obtain the matched samples (Sekhon 2011). The algorithm we use matches each treatment household to a control household with replacement. We restricted matches to nearest neighbor within a caliper as small as 0.01, but our central findings and the overall quality of matches are robust to caliper restrictions. Consequently, the matched samples we report below include all WIC participating households in the sample or subsample that had common support. Once the matched sample is formed, the treatment effect can be estimated by directly comparing outcomes between treated and untreated subjects in the matched sample.

Results

In this section, we first present results from the logistic regressions that we use to model WIC participation. Second, we assess the quality of our matches by checking the balance between the WIC and eligible non-WIC samples. Third, we then present the ATT of WIC participation and assess Rosenbaum sensitivity of the ATT to hidden bias or unobserved heterogeneity. Fourth, we assess robustness of our main findings by repeating the analysis for a sample containing only households who shopped at a WIC-approved retailer and again using a modified HPI that excludes the majority of WIC purchases. We then present information on differences in ATT by self-reported healthiness of purchases and by access to supermarkets. Finally, we present the ATT estimates derived through an alternative generalized random forest algorithm.

Determinants of WIC Participation

Marginal effects from the logit models used to form the matched samples are reported in Table 2. The model in Table 2 includes state fixed effects (not reported). These fixed effects are included because earlier studies show that states with stricter WIC eligibility rules have lower take-up (Bitler, Currie, and Scholz 2003; Swann 2010). To formally test whether the existence of state effects matters, we conducted a log-likelihood ratio test comparing models with and without the state effects and reject the null hypothesis of no state effects.

Table 2 shows that households with higher income and the highest levels of educational attainment are less likely to participate in WIC. Hispanic households are more likely to participate in WIC. Household composition is also important. Specifically, households with eligible infants and women are more likely to participate in WIC. These findings are largely consistent with earlier work that has examined WIC participation. For example, there is evidence that WIC take-up is lower for children age 1 to 4 (Bitler, Currie, and Scholz 2003; Tiehen and Jacknowitz 2008) and higher among eligible postpartum women (Tiehen and Jacknowitz 2010). Earlier work also shows higher rates of participation among Hispanic households (Bitler and Currie 2004; Bitler, Currie, and Scholz 2003). WIC take-up is higher among socially disadvantaged women (Tiehen and Jacknowitz 2010; Swann 2010; Bitler, Currie, and Scholz 2003).

We find no evidence that access to WIC clinics affects participation among households in the FoodAPS sample. This is in contrast to Rossin-Slater (2013) who finds that access to clinics increases WIC food benefit take-up. However, Rossin-Slater (2013) has a much larger sample from a single state and exploited information about opening and closing of clinics, information which we were unable to obtain for the more geographically diverse FoodAPS sample. That we do not find a significant effect of clinic access could be due to differences in program delivery and clinic access across the many states represented in FoodAPS. The model in Table 2 does not include a covariate measuring supermarket access, but we look further at the issue of supermarket access in a follow-up analysis below[27]. In sum, we do not find evidence that geographic barriers (i.e., in terms of clinic access) meaningfully affect WIC participation in this sample.

Assessing the Quality of Matched Samples

As explained above, the goal of propensity score matching is to obtain a dataset that is similar to one that would result from a randomized experiment. For this reason, we want the distribution of covariates to be the same between the matched treated and control groups. One way to check this is assess the balance post-match. We use the standardized difference measure proposed by Rubin (1991). For each explanatory variable in the logit model, the standard difference of the sample means in the treated and matched controls are presented in Figure 1. The overall mean difference

[27]5. We did not include supermarket access to avoid the causal loop. The majority of supermarkets take WIC. It is unclear whether households take up WIC because they have access to supermarkets or whether they shop at a supermarket that takes WIC because of WIC participation.

before matching lies between 0.4 percent and 80.4 percent for WIC households and eligible non-WIC households. The bias is reduced to between 0.4 percent and 13.0 percent in the post-match sample.

Since the covariates include not only continuous variables, but also binary variables, a Kolmogorov-Smirnov test based on 2,000 bootstrap iterations is employed to provide correct coverage as recommend by Sekhon (2011). Based on the Kolmogorov-Smirnov test, before matching, the unbalanced variables are household income, households of Hispanic ethnicity, the existence of a WIC eligible woman in the household, and the number of WIC-eligible children in the household. After matching, these differences are reduced and there is no longer significant imbalance between WIC households and eligible non-WIC households. We also use a two-sample t-test, as proposed by Caliendo and Kopeinig (2008), to check if there are significant differences in covariate means between treated and matched. These two-sample t-tests provide additional evidence that covariate balance is achieved at the 5 percent level (see Appendix 1).

Finally, we test the assumption of common support by checking the distribution of the propensity scores for the treatment and control groups, as exhibited in Figure 2. As shown in Figure 2, almost all the treated observations could be matched with non-zero propensity score control observations and we restricted the matched samples to the region with common support.

Effect of WIC on Nutritional Quality of Household Purchases

Table 3 presents the ATT estimates for the effect of WIC participation regardless of whether or not households redeemed WIC vouchers during the interview week, the effect of WIC participation for households who had a WIC food redemption, and the effect of WIC participation for households who did not have a WIC food redemption. The control group in all these analyses comprises those who are eligible but non-WIC participants. We estimate a positive and statistically significant ATT of 2.742 for WIC participation. As noted above, the HPI ranges from 0 to 100. Given that the average HPI value for WIC participants is about 50.259, this estimate suggests that WIC participation improves the nutritional value of purchases by about 5.5 percent. This finding is consistent with earlier work showing that WIC participation is associated with improvements in diet quality, nutrient intakes, and biochemical indicators of nutritional adequacy (e.g., see review by Fox et al. 2004).

To shed light on the importance of WIC foods, Table 3 also reports ATT estimates for the sample of WIC participating households who used WIC benefits to redeem foods during the interview week and those households who did not. The ATT is much higher at 9.443 among the households who redeemed WIC foods. In contrast, the ATT for the sample of WIC households who did not redeem WIC foods is -0.843. It is not surprising that the ATT is higher when WIC foods are redeemed. After all, foods eligible for purchase through WIC are those that help recipients meet dietary guidelines. The important finding is that there is no evidence that WIC participation improves nutritional quality among those participating households who do not redeem benefits. If households with healthier food preferences self-selected into the program, we should see a higher ATT even when WIC benefits are not redeemed. The ATT from this group is small and not statistically different from zero.

As noted above, the HPI we use to assess quality of household purchases is an aggregate of the 12 HEI components. Table 4, reports ATT estimates for each of these components estimated from the matched sample of WIC households redeeming WIC benefits. Table 4 suggests that households redeeming WIC foods scored significantly higher on total fruit, whole fruit, whole grains, dairy, seafood and plant protein, refined grains, and empty calories. In interpreting the table, recall that empty calories and refined grains are inversely coded with lower consumption of these items resulting in higher component scores. This is not surprising because these categories are emphasized/deemphasized in the WIC food packages. Earlier evidence shows reduced intakes of fats and added sugars among WIC participants (Basiotis, Kramer-LeBlanc, and Kennedy 1998; Wilde, McNamara, and Ranney 1999; Kranz and Siega-Riz 2002; Siega-Riz et al. 2004). The

2009 changes to the WIC food packages resulted in increased purchases of whole grain products (Oh, Jensen, and Rahkovsky 2016).

Rosenbaum Bounds to Assess Hidden Bias

When referring to hidden bias, we assume that some characteristics are unobserved and are not in the vector of covariates used in the matching model. Propensity-score matching estimators are based on the assumption that selection is on observable characteristics. This means that conditional on the observed covariates, the process by which units are selected into treatment is independent of unmeasured variables that affect the outcome variable. In order to estimate the extent to which such selection on unobservable or hidden bias may affect the estimates, we conducted a Rosenbaum bounds sensitivity analysis (Rosenbaum 2002; DiPrete and Gangl 2004; Drichoutis, Nayga, and Lazaridis 2009). This method assesses the sensitivity of the significance levels of the ATT and estimates the magnitude of hidden bias it would take to change inference assessments from statistical significance to insignificance. Details about computing Rosenbaum bounds can be found in Rosenbaum (2002) and DiPrete and Gangl (2004). We used the ?rbounds? package in R to conduct the sensitivity analysis (Keele 2010). Tables 3 and 4 present Rosenbaum?s gamma, the measure of hidden bias that that could potentially switch an inference decision at the 5 and 10 percent critical values.

Gamma is interpreted as the magnitude by which an unobserved variable would need to affect the odds ratio of treatment in order to cause an inference decision to switch from being significant to insignificant. As the Rosenbaum test reveals, our ATT from the sample of all WIC participants switches from being statistically significant to insignificant at a gamma value of 1.13 and 1.17 at the 5 and 10 percent levels, respectively. This indicates that the estimate would remain significant at the 10 percent level in the presence of hidden bias up to 17 percent. Table 3 shows that the significance of the large ATT estimated for the sample of WIC households who redeemed WIC benefits is very robust to hidden bias with gamma values of 2.82 and 3.05 at the 5 and 10 percent levels, respectively.

The statistically significant ATTs on the component measures estimated from this subsample in Table 4 are also robust to hidden bias. The refined grain component is the most sensitive with a gamma of 1.22 corresponding to the 5 percent critical value. Gammas for the other component scores range from 1.28 to 2.25 at this critical value. Hence, the ATTs on most of the significant component measures would remain significant at the 5 percent level even in the presence of substantial hidden bias.

Additional Evidence on the Importance of WIC Foods

To summarize, the evidence presented so far supports the conclusion that the improvements in nutritional quality attributable to WIC participation are driven by WIC food packages and are not simply a reflection of selection bias. This conclusion is strengthened by the fact that the FoodAPS sample was not stratified by time of month or date of delivery of WIC benefits. For this reason, the samples of WIC households redeeming and not redeeming their WIC benefits should be nearly random. Indeed, these two groups of households have very similar characteristics as exhibited in the last two columns of Table 1[28].

However, there are three potential issues that deserve further attention. First, WIC foods must be redeemed at WIC approved retailers. If these retailers stock healthier foods in general, then differences in shopping venue could account for some of the improvements in nutritional quality attributable to the subsample that redeemed WIC foods. Second, our finding could reflect a secondary selection problem wherein some households who enroll in the program are systematically more likely to only partially redeem food benefits. This could occur if shopping venues available to

[28]The two groups of households differ statistically only on the household members classified as African American (at 0.05 level).

the household stock some but not all of the foods in the WIC package or if some households deem some WIC foods to be undesirable. Third, infant formula is included in the calculation of HPI. The inclusion of infant formula in HPI could affect the measures of HPI and its components.

To address the first issue, we restricted the sample to include only participating and non-participating households who shopped at a WIC-approved retailer during the interview period. This restriction resulted in the removal of three households from the eligible but non-participating sample and four households who were on WIC but did not redeem WIC benefits during the interview week. Thus, it is unlikely that venue explains the results because the overwhelming majority of our sample shopped at a WIC approved retailer. However, the exclusion of these few households did alter the matched samples as can be seen by comparing the numbers of observations in Table 5 to those in Table 3. Nevertheless, as shown in Table 5, the ATT estimates are very close to those reported above. For WIC participants without regard for redemptions, the ATT estimate was 2.671. For those who redeemed and did not redeem foods during the interview week, the estimates were 9.385 and -0.963, respectively. Thus, in comparison to Table 3, no materially different conclusions are reached from the analysis summarized in Table 5.

To address the second issue, we re-estimated the ATT from the original matched samples in Table 3 but using an HPI that excludes items from shopping events where WIC redemptions accounted for more than 50 percent of the total expenditures. In FoodAPS, each shopping event is flagged as to whether WIC benefits were redeemed during the purchase event and the dollar value of WIC redemptions is indicated[29]. Unfortunately, FoodAPS does not provide item-level information on which items were purchased with WIC benefits and which were purchased using other forms of payment. However, of the 273 WIC purchase events, the overwhelming majority were solely WIC events. These could be identified by the fact that the total value spent on the shopping

[29] In October 2009, the USDA revised the WIC food package and introduced cash-value vouchers (CVV) for fruits and vegetables. Since FoodAPS does not separately identify CVV, the relatively small value recorded for fruits and vegetables expenditures when WIC benefits were used may reflect misreporting of the vegetables and fruits expenditures when using the CVV (National Academies of Sciences, Engineering, and Medicine. 2017). Therefore, our estimates of the effect of WIC benefits may be underestimated (i.e., a lower bound).

occasion was equal to the dollar value of WIC redemptions[30]. In only 28 shopping event cases did WIC redemptions account for less than half of the total value of the food purchases. Thus, by excluding majority-WIC shopping events, we effectively remove most WIC foods from the HPI calculation.

As shown in Table 6, when this revised HPI is used as the outcome, there is no longer a significant WIC effect. Among all WIC participants without regard for redemptions, the ATT estimate is -0.087. For those who redeemed and did not redeem foods, the ATT estimates are 1.420 and -0.843, respectively. The estimate for households who did not redeem WIC foods matches that in Table 3 because these households acquired no foods through WIC during the interview week and so the HPI score remains unchanged for these households. The magnitude of each estimate is close to one index point and not statistically different from zero, suggesting that there is no difference between WIC and eligible non-WIC households once WIC foods are effectively removed from the measure of nutritional quality. Again, this reinforces the conclusion above that WIC foods are the best explanation for observed improvements in dietary outcomes associated with the program, not systematic differences in the characteristics or behaviors of participating and eligible non-participating households.

Finally, to address the third concern about infant formula, we re-estimated the ATT from the original matched samples but excluding infant formula from the HPI calculation. The results are presented in Appendix 2. As shown, there was no meaningful difference in ATT when infant formula was excluded from HPI. Somewhat surprisingly, the ATT in Appendix 2 for total dairy is slightly higher. This could be explained by the fact that some of the formula items were non-dairy, specialty formulas. Also, exclusion of infant formula removes energy from the index calculation, which can increase component scores. Thus we conclude that the primary findings are robust to the exclusion of infant formula.

[30] The number of shopping events will not match the number of households because households can have multiple shopping events during the interview period.

Additional Insights on WIC from FoodAPS

FoodAPS includes the response to a self-assessed question about whether the household is following a healthy diet. As reported in Table 1, the proportion responding yes to this question is similar across WIC participants and eligible non-participants but is a bit lower among those who redeemed WIC benefits during the reporting period in relation to those who did not. As a follow-up, we matched those with yes and no responses to this question separately to eligible non-participants to obtain ATT estimates for each group. For those responding yes, the ATT is 3.720 (Std. Err=1.710) and is statistically significant at the 5 percent level. For those responding no, the ATT estimate is about 1.5 index points lower at 2.137 (Std. Err=1.474) and is not significantly different from zero. Nevertheless, this is not a substantial difference in the ATT estimates and the fact that these estimates are similar provides some additional context to the selection issue we explore above.

Another concern is whether households in lower income neighborhoods without access to nearby supermarkets may benefit less from nutritional programs like WIC. Given the geographic component of FoodAPS, we explore the heterogeneity that may exist because some households have limited access to supermarkets. We match 93 WIC households with limited supermarket access and 330 WIC households with supermarket access to eligible non-WIC households. We estimate the ATTs for each subgroup. The ATT estimate from the limited access subgroup is 2.340 (Std. Err=1.971) and is not statistically different from zero while the ATT from the subgroup without limited access is 2.984 (Std. Err=1.417) and is significant at the 5 percent level. Given the similarity of these ATT estimates, there is no compelling evidence from this sample that nutritional improvements from WIC are adversely affected by supermarket access.

Robustness Check Using Machine Learning

With propensity matching, we assumed a logistic function between the likelihood of WIC participation and the covariates. In reality, this relationship can be complex and unknown, and assuming a parametric relationship can lead to biases (Busso, DiNardo, and McCrary 2014). Thus, to further check the robustness of our findings presented above, we apply an alternative method, a non-parametric supervised machine learning approach called the ?honest? random forest that places no restrictions on model complexity to derive the ATTs (Athey, Tibshirani, and Wager 2017; Athey 2015; Wager and Athey 2018). An ?honest? tree splits a randomly selected subsample for the use of model structure estimation. Therefore, the asymptotic properties of treatment effect estimates within the splits are the same as if the partition had been exogenously given (Athey, Tibshirani, and Wager 2017; Athey and Imbens 2016; Wager and Athey 2018). We use an algorithm developed by Athey, Tibshirani, and Wager (2017), which is available in the R package ?grf?[31]. Since we are not concerned about the interpretability of the prediction model, we included additional covariates to check if they can improve estimation[32]. Essentially, these covariates help us detect similarities among WIC participants and among eligible non-participants and assign them to different groups (leaves). This method relies on a type of residual-on-residual regression in the leaves to eliminate the effect of confounding. Intuitively, if our main findings from the propensity score matching analysis are a result of the true data-generating process, we should not see different effects when we use the alternative approach provided through this machine-learning based method.

Generalized random forest estimates are reported in Table 7 and are consistent with the main results discussed above[33]. Specifically, we find similar ATTs on the overall WIC sample (ATT=2.267), the

[31]The ?grf? package version used in this study is 0.9.5.

[32]These additional covariates included alternative measures of supermarket accessibility, household use of nutrition information, number of meals consumed at home during the interview period, and additional indicators of household economic status in addition to those covariates listed in Table 2. The trees were also estimated using only the covariates in Table 2, but results (not reported) are similar to those reported below in Table 7.

[33]Appendix 3 presents the generalized random forest estimates of the ATTs without infant formula.

sample that redeemed WIC (ATT=7.720), and the sample that did not redeem WIC (ATT=-0.903). These results confirm our conclusion that only those who redeemed WIC observe a significant WIC effect, or that the WIC package is the driver of increased HPI. Our examination of the HPI that excluded majority WIC purchase events yields an ATT of -0.109, which is close to zero, statistically insignificant, and similar to that reported in Table 6. This supports our conclusion that WIC-provided foods are driving the improvement in diet and that this improvement is not simply a reflection of healthier households selecting into the program. When reviewing the effect of WIC on the HPI components, the list of components with positive and statistically significant improvements is similar to that reported above and includes total fruit, whole grain, dairy, seafood and plant protein, sodium, and empty calories. The total vegetables component was negative in Table 4 and is negative and significant in Table 7. The estimate for the total-vegetable component is notable because at the time the FoodAPS data were collected, white potatoes were not permitted for WIC purchases (Oliveira and Frazo 2015). White potatoes are also the most widely consumed fresh vegetable in the United States (USDA-ERS 2017c). The exclusion of white potatoes from WIC-eligible fresh vegetables is a plausible explanation for the negative ATT on the total vegetable component. There is also evidence that WIC recipients prefer to purchase fruits than vegetables with their WIC vouchers and CVV (Andreyeva and Luedicke 2015).

Limitations

There are a few limitations in this study. One is that other members of the household could consume foods purchased with WIC benefits. Unfortunately, it is not possible to study the consumption of foods among household members using FoodAPS data. If there exists program spillover from WIC recipients to non-recipients, the improvements in purchases we measure may not accrue exclusively to the targeted individual. Even though our study finds a meaningful effect of WIC foods on the healthiness of purchases, we are unable to make the connection between healthier purchase and consumption or better health. This would be an important area for future research with data availability.

Another limitation is that FoodAPS was conducted between 2012 and 2013, during which time a few states were moving to the Electronic Benefit Transfer (EBT) system. We find evidence that most WIC purchase events were exclusively WIC purchase events in that all or most items were purchased with WIC benefits. The availability of EBT could facilitate the spread of the benefits over several shopping events. At present, most states have adopted EBT. To the extent EBT facilitates a more complete redemption of WIC benefits, our findings may underestimate the effect of WIC foods on purchase quality.

Finally, the FoodAPS datasets constitute a unique and valuable source of data but there are some constraints inherent in their use. First, FoodAPS is cross-sectional, which makes it difficult to access longer-term effects of WIC. For example, we find that WIC households had healthier purchases only when WIC benefits were redeemed. It would be incorrect to conclude that other aspects of WIC, such as nutrition education, are unimportant because such a conclusion would require follow-up with time. In addition, FoodAPS contains a relatively small sample of WIC households. Because of this, further heterogeneity tests were not possible given limited statistical power.

Conclusion

Children of low-income households tend to lag behind other children on a wide range of health outcomes. They also tend to be food insecure and have inadequate intake of important nutrients. As previously discussed, the role of WIC in improving birth, health and nutritional outcomes has been studied extensively. The findings of these studies generally suggest that WIC participation is associated with improved outcomes. However, most earlier work showing beneficial effects of

WIC on diet have been unable to convincingly determine whether the observed beneficial effects on diet are due to the program or to self-selection into the program.

We addressed this important topic using the FoodAPS data and found that households participating in WIC have higher HPI value in comparison to eligible non-participating households. While this finding is consistent with earlier findings linking WIC to improvements in diet quality, we found that this difference is driven entirely by households who redeemed WIC foods during the interview week. Importantly, we did not find any difference between WIC participating households that did not redeem WIC foods and WIC-eligible but non-participating households. Moreover, there was no difference in purchase quality when majority WIC shopping events were excluded from the HPI calculation. These findings suggest that WIC foods explain the improvement in relative quality of household food purchases, not self-selection of more nutrition-conscious households into the program. This is the key contribution of the present study. To be clear, it is not surprising to find that household purchases were healthier when WIC program benefits were used to purchase foods. After all, foods eligible for purchase under WIC are specifically designed to help beneficiaries meet dietary guidelines. Instead, the evidence against self-selection being an explanation for improvements in the quality of purchases lies in two main findings. First, WIC households had healthier purchases only when benefits were redeemed. During weeks when benefits were not redeemed, the nutritional quality of purchases among WIC households was no different from that of eligible non-participating households. Second, among those redeeming WIC benefits, the improvement in purchase quality disappeared once WIC purchase events were excluded from the purchase quality index. After exclusion, there was no longer a difference between those households who redeemed WIC foods and eligible nonparticipating households. Taken together, the evidence presented here suggests that foods provided by WIC can explain most if not all of the improvement in the quality of food purchases. There is little if anything left that can be explained by systematic differences in the characteristics of participating and non-participating households.

Access to the FoodAPS data permitted us to look at food purchases, the first link in a chain connecting WIC participation to better child and maternal health outcomes. Earlier studies have looked at links further along this chain such as diet quality, indicators of nutritional adequacy, and birth outcomes. The focus on purchases is important because we were able to examine the point in the chain where behaviors related to self-selection into the program would be easiest to detect. Had there been evidence of healthier purchases either among participating households who did not redeem benefits or among the non-WIC purchases of households who did redeem benefits, it would have supported an argument that WIC households would have been likely to make healthier purchases regardless of the program. We found evidence of neither. Thus, our findings suggest that WIC-provided foods are central to the beneficial program outcomes documented in earlier studies. Moreover, our findings on purchase quality align nicely with earlier findings on diet quality.

Overall, the findings of this study point to the importance of the WIC program in helping participants acquire the foods needed for a healthier diet. The ability of WIC to continue serving eligible households could be curtailed, however, if current proposals to significantly cut WIC funding push through. This issue becomes even more relevant for eligible households when considering that WIC is different from other welfare programs in that it is not an entitlement program. Not everyone who is eligible for WIC assistance is guaranteed to receive it because the reach of the program depends on adequate funding.

Appendix

Figure 25: Pre and post-match balance comparisons by subsample.

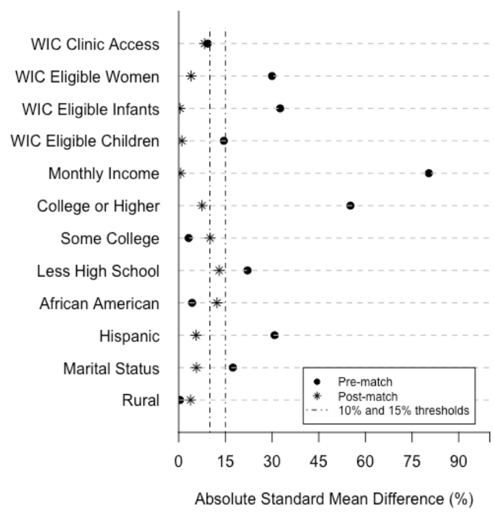


Figure 26: Distribution of propensity scores among WIC and eligible non-WIC households before the imposition of common support.

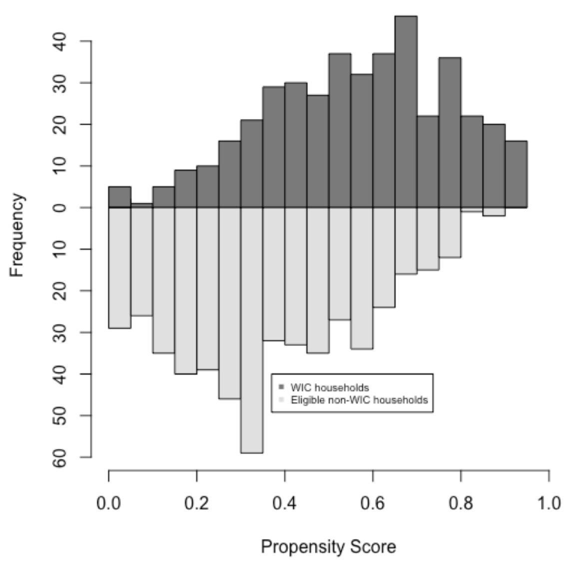


Table 35: **Descriptive Statistics for the Household (HH) Sample**

Variable	Non WIC	WIC	Redeemed WIC	Did Not
	N = 505	N = 423	N=152	N=271
Healthy Purch. Index	50.388	50.259	55.958	47.062
Rural	0.236	0.234	0.257	0.221
Marital Status	0.626	0.539	0.579	0.517
Hispanic	0.247	0.392	0.351	0.415
African American	0.156	0.172	0.112	0.205
Less High School	0.081	0.163	0.164	0.162
High School	0.196	0.324	0.309	0.332
Some College	0.384	0.369	0.355	0.376
College or Higher	0.339	0.144	0.171	0.129
Monthly Income	5.144	2.834	3.071	2.702
WIC Eligible Children	0.853	0.749	0.822	0.708
WIC Eligible Infants	0.129	0.286	0.342	0.255
WIC Eligible Woman	0.158	0.296	0.283	0.303
WIC Clinic Access	0.358	0.404	0.362	0.428
Supermarket Access	0.358	0.404	0.362	0.428
Self-reported Healthy Diet	0.382	0.409	0.375	0.428

Table 36: Effect of WIC Participation (Average Treatment Effect on the Treated (ATT)) on Healthy Purchasing Index Score from Matched WIC Subsamples: Healthy Purchasing Index Excludes Primary WIC Purchase Events

	All Participants	Redeemed WIC	Did Not Redeem WIC
ATT estimate	-0.087	1.420	-0.843
Standard error	1.277	1.520	1.421
p-value	0.946	0.351	0.553
N (post-match)	534	178	337

Table 37: Logit model used to Match WIC Participants to Eligible Non-Participants.

Variable	Marginal Effects	Standard Error
Rural	0.063	(0.044)
Marital Status	0.026	(0.033)
Hispanic	0.161***	(0.044)
African American	0.076	(0.048)
Less High School	0.024	(0.052)
Some College	-0.050	(0.037)
College or Higher	-0.193***	(0.046)
Monthly Income	-0.039***	(0.007)
WIC Eligible Children	-0.022	(0.024)
WIC Eligible Infants	0.157***	(0.038)
WIC Eligible Woman	0.172***	(0.037)
WIC Clinic Access	-0.025	(0.035)
Number of Observations	928	

Asterisks indicate significance: *, **, and *** at the 10, 5, and 1 percent levels, respectively

Table 38: Effect of WIC Participation (Average Treatment Effect on the Treated (ATT)) on Healthy Purchasing Index Score from Matched WIC Subsamples

	All Participants	Redeemed WIC	Did Not Redeem WIC
ATT estimate	2.742	9.443	-0.843
Standard error	1.351	1.573	1.421
p-value	0.042	< 0.001	0.553
N (post-match)	534	178	337
Critical value 0.05	$\gamma = 1.13$	2.82	-
Critical value 0.10	$\gamma = 1.17$	3.05	-

Magnitude of hidden bias (Gamma) from Rosenbaum sensitivity to change null hypotheses ATT=0

Table 39: Average Treatment Effect on the Treated (ATT) Estimates from Matched Subsample of Households who Redeemed WIC by Component of the Healthy Purchasing Index (HPI)

HPI Component	ATT Estimate	S.E.	p-value	(Gamma)	
Total vegetables	-0.101	0.189	0.594	-	-
Greens and beans	0.237	0.254	0.350	-	-
Total fruit	0.798	0.202	< 0.001	1.89	2.02
Whole fruit	0.534	0.247	0.031	1.30	1.39
Whole grains	1.829	0.338	< 0.001	2.25	2.43
Total dairy	1.001	0.415	0.016	1.28	1.36
Total protein	0.076	0.179	0.673	-	-
Seafood and plant protein	0.901	0.261	< 0.001	1.49	1.59
Fatty acids	0.398	0.473	0.400	-	-
Sodium	0.816	0.466	0.080	-	-
Refined grains	1.173	0.516	0.023	1.22	1.30
Empty calories	1.782	0.737	0.016	1.53	1.63

Higher values of each Adequacy component indicate an improvement in nutritional quality of purchases. Magnitude of (Gamma) from Rosenbaum sensitivity analysis to change null hypotheses that ATT=0

Table 40: Effect of WIC Participation (Average Treatment Effect on the Treated (ATT)) on Healthy Purchasing Index Score from Matched WIC Subsamples: Excludes Households not Shopping at a WIC-approved Store

	All Participants	Redeemed WIC	Did Not Redeem WIC
ATT estimate	2.671	9.385	-0.963
Standard error	1.353	1.579	1.425
p-value	0.048	< 0.001	0.499
N (post-match)	527	178	330
Critical value 0.05	$\gamma = 1.13$	2.80	-
Critical value 0.10	$\gamma = 1.17$	3.02	-

Magnitude of hidden bias (Gamma) from Rosenbaum sensitivity to change null hypotheses ATT=0

Table 41: Standard mean differences between WIC and eligible non-WIC households

Variable	Pre-match	Post-match
Rural	-0.378	3.802
Marital Status	-17.38***	5.732
Hispanic	30.783***	5.618
African American	4.325	-12.272*
Less High School	22.149***	-12.995*
Some College	-3.181	10.065
College or Higher	-55.273***	7.483
Monthly Income	-80.368***	0.642*
WIC Eligible Children	-14.496**	-1.043
WIC Eligible Infants	32.585***	0.392
WIC Eligible Woman	30.011***	4.037
WIC Clinic Access	9.33	8.292

Asterisks *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels

Table 42: Effect of WIC Participation (Average Treatment Effect on the Treated (ATT)) on Healthy Purchasing Index Score from Matched WIC Subsamples Excluding all Purchases of Infant Formula

	All Participants	Redeemed WIC	Did Not Redeem WIC
ATT estimate	2.868	9.412	-0.631
Standard error	1.368	1.633	1.428
p-value	0.036	< 0.001	0.659
N (post-match)	534	178	337
Critical value 0.05	$\gamma = 1.13$	2.64	-
Critical value 0.10	$\gamma = 1.17$	2.85	-

Magnitude of hidden bias (Gamma) from Rosenbaum sensitivity to change null hypotheses ATT=0

Table 43: Average Treatment Effect on the Treated (ATT) Estimates from Matched Subsample of Households who Redeemed WIC by Component of the Healthy Purchasing Index (HPI) Excluding all Purchases of Infant Formula

HPI Component	ATT Estimate	S.E.	p-value	(Gamma)	
Total vegetables	-0.042	0.192	0.826	-	-
Greens and beans	0.247	0.254	0.331	-	-
Total fruit	0.846	0.208	< 0.001	1.90	2.04
Whole fruit	0.553	0.252	0.028	1.29	3.55
Whole grains	1.910	0.346	< 0.001	2.31	2.49
Total dairy	1.076	0.421	0.011	1.31	1.40
Total protein	0.110	0.175	0.530	-	-
Seafood and plant protein	0.914	0.261	< 0.001	1.49	1.60
Fatty acids	0.505	0.479	0.292	-	-
Sodium	0.742	0.494	0.133	-	-
Refined grains	0.962	0.510	0.059	-	1.18
Empty calories	1.591	0.755	0.035	1.40	1.50

Higher values of each Adequacy component indicate an improvement in nutritional quality of purchases. Magnitude of (Gamma) from Rosenbaum sensitivity analysis to change null hypotheses that ATT=0

Table 44: Effect of WIC Participation (Average Treatment Effect on the Treated (ATT)) on Healthy Purchasing Index Score from Matched WIC Subsamples: Healthy Purchasing Index Excludes Primary WIC Purchase Events

	All Participants	Redeemed WIC	Did Not Redeem WIC
ATT estimate	-0.087	1.420	-0.843
Standard error	1.277	1.520	1.421
p-value	0.946	0.351	0.553
N (post-match)	534	178	337

References

Agranov, M., J. K. Goeree, J. Romero, and L. Yariv (2017). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association*.

Ahern, M., C. Brown, and S. Dukas (2011). A national study of the association between food environments and county-level health outcomes. *The Journal of Rural Health 27*(4), 367–379.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of economic literature 43*(3), 762–800.

Allcott, H., R. Diamond, J.-P. H. Dubé, et al. (2017). *The geography of poverty and nutrition: Food deserts and food choices across the United States*. National Bureau of Economic Research.

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Andris, C., D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden (2015). The rise of partisanship and super-cooperators in the us house of representatives. *PloS one 10*(4), e0123507.

Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, Volume 30. Springer Science & Business Media.

Baldwin, K. and J. D. Huber (2010). Economic versus cultural differences: Forms of ethnic diversity and public goods provision. *American Political Science Review 104*(4), 644–662.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Berinsky, A. J. (1999). The two faces of public opinion. *American Journal of Political Science*, 1209–1230.

Bishop, G. F. and B. S. Fisher (1995). Secret ballots and self-reports in an exit-poll experiment. *Public Opinion Quarterly 59*(4), 568–588.

Bitler, M. (2014). The health and nutrition effects of snap: Selection into the program and a review of the literature on its effects.

Bitler, M. P. and J. Currie (2005). Does wic work? the effects of wic on pregnancy and birth outcomes. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management 24*(1), 73–91.

Blair, G. and K. Imai (2012). Statistical analysis of list experiments. *Political Analysis 20*(1), 47–77.

Blisard, N., B.-H. Lin, J. Cromartie, and N. Ballenger (2002). America's changing appetite: Food consumption and spending to 2020. *FOOD REVIEW-WASHINGTON DC- 25*(1), 2–9.

Brown-Iannuzzi, J. L., M. B. Najle, and W. M. Gervais (2017). The illusion of political tolerance: Social desirability and self-reported voting preferences. *Journal of Experimental Social Psychology*.

Bursztyn, L., D. Cantoni, P. Funk, and N. Yuchtman (2017). Polls, the press, and political participation: The effects of anticipated election closeness on voter turnout.

Charness, G. and U. Gneezy (2008). What's in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization 68*(1), 29–35.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Claassen, R. L. and J. B. Ryan (2016). Social desirability, hidden biases, and support for hillary clinton. *PS: Political Science & Politics 49*(4), 730–735.

Coffman, K. B., L. C. Coffman, and K. M. M. Ericson (2016). The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*.

Connors, E., S. Klar, and Y. Krupnikov (2016, November). There may have been shy trump supporters after all. *The Washington Post*.

Coppock, A. (forthcoming). Did shy trump supporters bias the 2016 polls? evidence from a nationally-representative list experiment. *Statistics, Politics, and Policy*.

Corstange, D. (2008). Sensitive questions, truthful answers? modeling the list experiment with listit. *Political Analysis 17*(1), 45–63.

Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

Droitcour, J., R. A. Caspar, M. L. Hubbard, T. L. Parsley, W. Visscher, and T. M. Ezzati (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 185–210.

Dropp, K. (2015, December). Why does donald trump perform better in online versus live telephone polling? *Morning Consult*.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics 112*(4), 1203–1250.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research.

Enten, H. (2016a, May). Americans? distaste for both trump and clinton is record-breaking. *FiveThirtyEight*.

Enten, H. (2016b, November). ‘shy’ voters probably arent why the polls missed trump. *FiveThirtyEight*.

Esteban, J. and D. Ray (2008). Polarization, fractionalization and conflict. *Journal of peace Research 45*(2), 163–182.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *Review of Economic Studies 83*(2), 587–628.

Fang, D., M. R. Thomsen, R. M. Nayga Jr, and A. M. Novotny (2018). Wic participation and relative quality of household food purchases: Evidence from foodaps. Technical report, National Bureau of Economic Research.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research 20*(2), 303–315.

Fox News (2016, January). See which candidates qualified for the fox news-google gop debates.

Goffman, E. (1959). *The presentation of self in everyday life*. Harmondsworth.

Gundersen, C., B. Kreider, and J. Pepper (2011). The economics of food insecurity in the united states. *Applied Economic Perspectives and Policy 33*(3), 281–303.

Gundersen, C., B. Kreider, and J. V. Pepper (2017). Partial identification methods for evaluating food assistance programs: A case study of the causal impact of snap on food insecurity. *American Journal of Agricultural Economics 99*(4), 875–893.

Hales, C. M., M. D. Carroll, C. D. Fryar, and C. L. Ogden (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. US Department of Health and Human Services, Centers for Disease Control and ?

Halpin, J. and K. Agne (2009). State of american political ideology, 2009. *Center for American Progress*.

Heerwig, J. A. and B. J. McCabe (2009). Education and social desirability bias: The case of a black presidential candidate. *Social Science Quarterly 90*(3), 674–686.

Hoffman, E., K. McCabe, and V. L. Smith (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review 86*(3), 653–660.

Hopkins, D. J. (2009). No more wilder effect, never a whitman effect: When and why polls mislead about black and female candidates. *The Journal of Politics 71*(3), 769–781.

Huang, J. and C. Low (2017). Trumping norms: Lab evidence on aggressive communication before and after the 2016 us presidential election. *American Economic Review 107*(5), 120–24.

Janus, A. L. (2010). The influence of social desirability pressures on expressed immigration attitudes. *Social Science Quarterly 91*, 928–946.

Jekanowski, M. D., J. K. Binkley, et al. (2000). Food purchase diversity across us markets. *Agribusiness 16*(4), 417–433.

Kane, J. G., S. C. Craig, and K. D. Wald (2004). Religion and presidential politics in florida: A list experiment. *Social Science Quarterly 85*(2), 281–293.

Kantchev, G. and C. Whittall (2017, June). After another surprise election result, investors grow skeptical of pollsters.

Karlan, D. S. and J. Zinman (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics 98*(1), 71–75.

Keeter, S. and N. Samaranayake (2007). Can you trust what polls say about obama's electoral prospects. *Pew Research Center*.

Kinsey, J. D. (1994). Food and families' socioeconomic status. *The Journal of nutrition 124*(suppl_9), 1878S–1885S.

Kreider, B., J. V. Pepper, and M. Roy (2018). Does the women, infants, and children program (wic) improve infant health outcomes?

Krysan, M. (1998). Privacy and the expression of white racial attitudes: A comparison across three contexts. *Public Opinion Quarterly*, 506–544.

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. Vishny (1999). The quality of government. *The Journal of Law, Economics, and Organization 15*(1), 222–279.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social forces 13*(2), 230–237.

Lax, J. R., J. H. Phillips, and A. F. Stollwerk (2016). Are survey respondents lying about their support for same-sex marriage? lessons from a list experiment. *Public Opinion Quarterly*, 510–533.

Lo, Y.-T., Y.-H. Chang, M.-S. Lee, and M. L. Wahlqvist (2012). Dietary diversity and food expenditure as indicators of food security in older taiwanese. *Appetite 58*(1), 180–187.

Maccoby, E. E. and N. Maccoby (1954). The interview: A tool of social science. *Handbook of social psychology 1*, 449–487.

Mas, A. and E. Moretti (2009). Racial bias in the 2008 presidential election. *The American Economic Review 99*(2), 323–329.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of public Economics 89*(11-12), 2325–2368.

Miller, J. D. (1984). A new survey technique for studying deviant behavior. *Unpublished doctoral dissertation, George Washington University, Washington, DC*.

Miller, J. M. and J. A. Krosnick (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, 291–330.

Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review 95*(3), 796–816.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology 46*(3), 598.

Pew Research Center (2012, May). Assessing the representativeness of public opinion surveys.

Pew Research Center (2016, June). Partisanship and political animosity in 2016.

Powell, R. J. (2013). Social desirability bias in polling on same-sex marriage ballot measures. *American Politics Research 41*(6), 1052–1070.

Raghavarao, D. and W. T. Federer (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40–45.

Redlawsk, D. P., C. J. Tolbert, and W. Franko (2010). Voters, emotions, and race in 2008: Obama as the first black president. *Political Research Quarterly*.

Rothschild, D. (2009). Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly 73*(5), 895–916.

Rudman, L. A. and S. E. Kilianski (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin 26*(11), 1315–1328.

Santos, F. C., M. D. Santos, and J. M. Pacheco (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature 454*(7201), 213.

Schillinger, D., L. R. Barton, A. J. Karter, F. Wang, and N. Adler (2006). Does literacy mediate the relationship between education and health outcomes? a study of a low-income population with diabetes. *Public health reports 121*(3), 245–254.

Shepard, S. (2016, October). Gop insiders: Polls don't capture secret trump vote. *Politico*.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics 118*, 26–40.

Streb, M. J., B. Burrell, B. Frederick, and M. A. Genovese (2008). Social desirability effects and support for a female american president. *Public Opinion Quarterly 72*, 76–89.

Traugott, M. W. and V. Price (1992). A review: exit polls in the 1989 virginia gubernatorial race: Where did they go wrong? *Public Opinion Quarterly 56*(2), 245–253.

Tsuchiya, T., Y. Hirai, and S. Ono (2007). A study of the properties of the item count technique. *Public Opinion Quarterly 71*(2), 253–272.

Ver Ploeg, M. (2010). *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to Congress*. Diane Publishing.

Vespa, J., D. M. Armstrong, and L. Medina (2018). Demographic turning points for the united states: population projections for 2020 to 2060. *Washington, DC: US Census Bureau 42*.

Wagner, A., R. J. Zeckhauser, and A. Ziegler (2017). Company stock reactions to the 2016 election shock: Trump, taxes and trade. Technical report, National Bureau of Economic Research.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association 60*(309), 63–69.

Wolfers, J. and E. Zitzewitz (2004). Prediction markets. *The Journal of Economic Perspectives 18*(2), 107–126.

Wolfers, J. and E. Zitzewitz (2016). What do financial markets think of the 2016 election. Technical report, Working paper.

Yen, S. T. (2010). The effects of snap and wic programs on nutrient intakes of children. *Food Policy 35*(6), 576–583.

Conclusion

People respond to incentives and diversity is an incentive. While this may be a simple principle of economics, it drives much of the economic behavior we observe when you think of the contextual incentive of diversity. Individuals may not be truthful due to social desirability bias (contextual incentive) thus creating a difference in a person's stated versus revealed preference i.e. an incentive to not be truthful in a social situation. This is demonstrated in our polling and SDR paper where respondents would show more support or less support for a given candidate depending on how observable a response may be. Thus different expectations and stereotypes can begin to drive behavior which are based on the perception that other persons are different from themselves. Oppositional to this, we see that people modify revealed food choices (actual consumption) when given an environment that promotes different food choices; thus, showing that diversity of food offerings is associated with better health outcomes. Lastly, food assistance programs that promote more diverse foods show large gains in healthy purchase outcomes. While diversity seems to be a double edged sword incentivizing positive and negative outcomes, one thing is certain that diversity plays a large and significant role in many people's day to day lives and its effect will only get larger as the world grows more diverse and integrated.