# Target journal = [Ecology of Food and Nutrition](https://www.tandfonline.com/action/authorSubmission?journalCode=gefn20&page=instructions#words)

# Abstract [required word count = 100, current word count = 109]

The complexity of dietary patterns necessitates use of innovative designs and statistical methods. We use a recently proposed statistical design method to estimate the causal relationship between healthy eating identity and diet and its robustness to unmeasured confounders. Among a sample of participants living in urban neighborhoods defined as food deserts, belonging to a high healthy eating identity group was causally associated with about five point [95% confidence interval: 1.18, 8.28] improvement in diet quality. This study provides stronger of evidence of the relationship between perceived healthy eating identity and diet to inform public health nutrition interventions.

# Background

Dietary patterns are complex and are not always driven by physiological or nutritional needs.[1] In particular, healthy eating identity has been found to have a meaningful impact on what individuals eat. [2–4] Healthy eating identity refers to the degree to which an individual believes he or she emphasizes healthy eating as an important aspect of self-identity [5] In their paper examining eating identity and two eating habits (consumption of fruits and vegetables vs intake of low-nutritional value foods), Strachan & Brawley found a significant association between healthy eating identity and healthy eating habits.[6] Kendzierski & Costello found that participants who classified themselves as healthy eaters were more likely to meet dietary guidelines for fiber, dietary cholesterol and total fat.[5] Blake and colleagues found a significant association between healthy eating identity type and dietary intake. [7]

These prior studies explored the association between healthy eating identity and diet quality, but estimates are primarily based on regression models. While such analytical methods are able to detect significant associations, one is unable to infer causal associations between factors such as healthy eating identity and dietary patterns without substantial assumptions. Estimates from regression models can also be misleading when covariate distributions in the groups being compared do not overlap, and the models require highly precise specification.[8–10] To our knowledge, no attempt has been made to carefully estimate a causal relationship between healthy eating identity and diet quality.

There is growing recognition of the methodological challenges to estimating causal associations in studies and a call for increased awareness of and improved methods to reduce bias and facilitate causal inference, particularly in obesity prevention research. [11] In the present study we used an analytical approach, which can be employed in other observational studies, to estimate a causal effect of healthy eating identity on diet quality among participants residing in “food desert” neighborhoods. Re-examination of the association between healthy eating identity and diet quality using modern methods is useful for several reasons. [11] First, there is value in using an alternative statistical framework that imposes fewer assumptions than prior methods used to estimate this association to see if these new methods replicate prior findings. Such a replication would provide supporting evidence to the existence of this association, which has import for interventions to reduce obesity trends and disparities. Secondly, since no randomized controlled trial has been done to assess this association, it would be useful to apply an analytical framework that accounts for unmeasured confounders of the association between healthy eating identity and diet. Lastly, it demonstrates new methods for examining causal inference that can be used to evaluate natural experiments more broadly.

The purpose of this study is to estimate a causal association between perceived healthy eating identity and diet, to inform public health nutrition interventions, using a method that accounts for unmeasured biases during design of an observational study.

# METHODS

Subjects and setting

This study used baseline data from the Future of Food in your Neighborhood Study (foodNEST), an ongoing natural experiment examining how changes to the food environment influence dietary behaviors. foodNEST participants are primary food shoppers living in two neighborhoods in Ohio, one with a plan to implement a food hub (in Cleveland) and the other one with no food hub (in Columbus). A primary food shopper does more than half of the food shopping for a household. The two neighborhoods were selected because they had similar socio-demographic profiles and were within a “food desert” (i.e., low-income census tracts with a substantial number or share of residents having low access to a supermarket or a large grocery store)[12] . A total of 1,395 primary food shoppers were screened of whom 655 met the study inclusion criteria (Figure 1) and 515 participated in all baseline surveys in 2015-16.

These 515 subjects form the basis of our study. Our primary analytic goal will be to compare 105 subjects who demonstrated high levels of healthy eating identity to a matched sample of subjects from the remaining 410 who displayed lower levels of healthy eating identity to examine how these groups differed in terms of their adherence to dietary recommendations.

Data collection

Data were collected using a psychosocial survey, a meal preparation and shopping screener, and a 24-hour dietary recall assessment tool. The psychosocial survey, described in Table 1, included four scales related to factors influencing food shopping and diet. The meal preparation and shopping screener collected data on frequency of meal preparation involvement in the previous month using a five-point Likert scale (1: Not involved at all, 5: Involved in all meal preparation), shopping frequency at a farmer’s market in fall and summer and frequency of buying food outside the home (1: Never, 2: A few times a year, 3: Monthly, 4: Every 2 weeks, 5: Weekly, 6: Two or more time a week, 7: At least once a day, 8: Don’t know). Demographic data collected included age, race, income and education. The variables we extracted from these surveys are either known confounders or potential confounders of the relationship between healthy eating identity and adherence to dietary recommendations.[13–16]

### Dietary assessment data come from a standardized 24-hour dietary recall questionnaire administered three times over 30 days. Responses from these dietary recalls were processed using the 2015 version of the Nutrition Data System for Research (NDSR) software [17], then used to compute the 2010 Health Eating Index (HEI-2010).[18] The HEI-2010 ranges from 0 to 100 with higher scores indicating greater adherence to the 2010 Dietary Guidelines for Americans [18, 19] The average of the three total HEI-2010 scores was used as the outcome measure in this study.

Data on healthy eating identity were collected using a three-item scale (Cronbach’s alpha =0.82) with responses on a four-point Likert scale (1: strongly disagree, 4: strongly agree).[7] The items were: (a) I am a healthy eater, (b) I am someone who eats in a nutritious manner, (c) I am someone who is careful about what I eat. Individuals who strongly agreed with all three statements were considered to have high healthy eating identity (n = 105) otherwise they were regarded to have low healthy eating identity (n=410).

In this study, we assumed “don’t know” responses were data that was missing at random. All missing values were imputed using a random forest based algorithm for imputation of mixed data (i.e., continuous and categorical data).[20]

Statistical analysis

Baseline differences between high and low healthy eating identity groups were compared with standardized differences (Table 2), chi-square and t tests. To reduce observable differences between comparison groups, we first estimated a quantity called *propensity score* (PS), which was defined as the probability of membership to an eating identity group given information on observed covariates in Table 2. A multivariable logistic regression model was used to estimate the PS. The second step was to pair high healthy eating identity participants (treatment group) to low healthy eating identity participants (controls) on the basis of their propensity score (PS). This procedure is called matching on the PS.

The three matching algorithms considered were (1) nearest neighbor matching (NNM), (2) optimal matching, and (3) genetic matching. In the current study we considered 1:1 matching with and without replacement, for each of the three methods, for a total of 6 matching options for review in the design phase. Technical notes on these matching algorithms and their implementation in the R statistical software, are provided elsewhere.[21]

For each algorithm, low healthy eating identity individuals who could not be matched were discarded. Assessment of balance of baseline variables before and after matching was based on standardized mean differences for each baseline variable, as well as Rubin’s balance criteria. [22]

Given that we ran 6 algorithms in the design phase of this study, we obtained 6 matched datasets. Identifying an optimal dataset to use for estimating causal association was done in two phases. In the first phase we identified which of the six algorithms produced datasets that met the specified balance criteria. In the second data phase we selected the eventual dataset to use for analysis by computing and comparing the design sensitivities of the remaining algorithms using the matched datasets they produced.

The concept of using design sensitivity to make design decisions in observational studies has been discussed extensively by Heller et al who has shown that a design whose matched sampling process has a higher design sensitivity statistic is less sensitive to unmeasured biases if the treatment (intervention) effect is significant. While in his application Heller used this statistic to select a primary outcome from several potential coherent outcomes, in the current study we propose using this statistic to identify an algorithm whose matching process produces a matched dataset whose estimates have a greater likelihood of having the highest robustness to unmeasured confounders. To accomplish this a three step procedure was used;

1. For each of the remaining matched samples, after the exclusion of those not meeting the specified balance criteria, we randomly selected a third of the pairs (planning sample) for use to estimate design sensitivity statistic and held out the rest of the pairs (analysis sample) to estimate treatment effect.
2. Using the planning sample from step 1, we estimated the design sensitivity for each matched sample. Details for estimating this parameter are outlined by Rosenbaum.[24] Briefly, let and denote treatment status and outcome values , respectively, for the member of the pair so that denotes the difference in outcomes for the pair. Through Monte Carlo simulation, with 1000 iterations, we then estimate the probability with and design sensitivity as .
3. Select the approach which produces the highest design sensitivity in step 3, and then complete the outcome analysis using the analysis sample associated with that approach.

After selecting the analytic sample, a mixed effects model was used to estimate the healthy eating identity effect on diet quality by regressing the HEI-2010 score on the healthy eating identity group indicator (as a fixed effect) and a matched pair ID of each observation (as a random effect). To assess the impact of using some of the data during design, we also ran the outcome model in the full matched sample and compared effect estimates from the two analyses (i.e., using only the planning sample and using the full matched sample). Using Rosenbaum’s sensitivity analysis framework [24, 25], we quantified the magnitude of hidden bias due to unmeasured confounders () that would need to be present to explain away the effect of healthy eating identity on diet quality so that this effect would no longer be due healthy eating identity.

# RESULTS

Prior to matching, we had 105 individuals in the high healthy eating identity group and 410 in the low healthy eating identity group. The high group had a better perception of healthy food access (2.97 vs 2.69; *P* < 0.005), better perception of control of food shopping (3.43 vs 3.01, *P*<0.01), and better family and neighborhood social support and norms for healthy eating (3.35 vs 3.03, *P* < 0.01 and 2.08 vs 1.77, *P* < 0.01, respectively). A higher proportion of participants with high versus low healthy eating identity were involved in the majority of meal preparation activities (89.50% vs 76.30%; *P* =0.005) and shopped at a farmer’s markets more frequently per year (74.30% vs 57.80%; *P* =0.003). Low healthy eating identity participants were more likely to shop at fast food restaurants (78.10 % vs 62.40 %; *P* <0.01).

In general, matched samples from all algorithms other than NNM with replacement reduced the difference between comparison groups on baseline covariates based on standardized differences (Figure 2) with difference constrained to within [-0.1,0.1] for most covariates. In Figure 3 we present results on Rubin’s balance criteria recommended to ensure reliability of estimates from linear regression-based outcome models (as is the case in this study). Prior to matching, the values of the first two Rubin’s balance criterion values (97.34% and 1.59) were outside acceptable range. The first criterion values were 5.94% for optimal matching with and without replacement; 23.40% and 6.00% for nearest neighbor matching with and without replacement, respectively; and 20.80% and 9.30% for genetic matching with and without replacement, respectively. More details of Rubin’s balance criterion are shown in Figure 2. Since NNM without replacement had two variables with standardized differences considerably outside the ideal range [-0.10,0.10] and acceptable range [-0.20,0.20], it was not considered part of competing algorithms for estimation of design sensitivity.

Genetic matching with replacement had the highest design sensitivity of 1.70 followed by NNM without replacement (1.65), genetic with replacement (1.38), optimal with and without replacement (1.22). Genetic matching with replacement was selected to be the design algorithm for this study because it had the highest design sensitivity. A description of full and analysis matched samples from genetic matching with replacement, stratified by healthy eating identity, are shown in Table 4.

For the analysis sample, the mean HEI-2010 score was 49.73 (95% confidence interval: 47.07, 52.40) with high healthy eating identity associated with a 4.73 (95% confidence interval: 1.18, 8.28) increase in HEI-2010 scores. Using the full sample, the mean HEI-2010 score was 49.40 (95% confidence interval: 46.26, 51.54) with high healthy eating identity associated with a 4.40 (95% confidence interval: 1.54, 7.09) increase in HEI-2010 scores. Results from hidden bias sensitivity analysis indicated that rejection of the null hypothesis would be sensitive to unmeasured bias that would alter the odds of being selected into a particular eating identity group by a magnitude of at least 3.50.

# DISCUSSION

We found the use of modern statistical methods was an effective tool for uncovering causal associations between two variables: healthy eating identity and diet quality. These two variables have been examined cross-sectionally finding strong associations. This research adds further evidence supporting the relationship between healthy eating identity and diet quality among a sample of primary food shoppers living in neighborhoods considered to be food deserts.

Existing literature indicates the need to try as many matching solutions as possible and ﻿choose the one with the best balance to ensure that a matching method that best reduces effect of measured bias is selected for design of an observational study.[28] During the design phase of an observational study, where matching is involved, it is typical to dedicate a lot of effort on specification of the PS model that will balance baseline variables. However, given the diversity in how matching is implemented by different algorithms, it is also important to assess the quality of matches produced by the selected PS model across these algorithms. Thus, assessing quality of matches produced by a specified PS model using only one matching algorithm might lead to an inferior study design since no consideration is given to other possible matching algorithms that might have superior balance.

When comparing several matching algorithms, it can be a challenge to select the best among them especially when they balance baseline variables comparably as in our study. Given this challenge, we have illustrated that our proposed design strategy is able to select a data generating process that achieves acceptable balance of baseline variables and accounts for the need to reduce sensitivity of our study design to hidden bias from unmeasured confounders (if H0 is rejected). Using a third of the data to make design decisions on which matched sampling algorithms to use for a study design did not lead to drastic differences in treatment effect estimates that would have been obtained if the entire sample had been used relative to analysis sample (4.40 vs 4.73). This finding is similar to that of Keller *et al* where design sensitivity was used to select an outcome for primary analysis among several outcomes (30). The objective of that study was similar to ours, in that a decision involving more than two design options (selecting outcome for primary analysis) had to be made with the aim of designing a study more robust to hidden bias relative to other potential options.

The three step we propose, incorporating design sensitivity statistic, can be used to improve the design and analysis of natural experiments for examining causal relationships. In the present study we illustrate its use in a study of healthy eating identity and diet quality as captured by the 2010-Healthy Eating Index. The average HEI-2010 score (48.4) of our study population of primary food shopper living in food deserts falls well below ﻿the target score of 74 required to meet ﻿Healthy People 2020 objectives.[26] The average score in our study sample is similar to that of a recent study comparing HEI-2010 scores among middle aged adults at high risk of poor diet quality to those at a lower risk.[27] There, the average HEI-2010 scores for those at nutritional risk was about 50. However, that study’s population was almost exclusively non-Hispanic white (97%) while in our study the population is predominantly African American (74%). To our knowledge, this study is the first to estimate causal association of healthy eating identity and diet quality using a matched sampling design.

The findings of this study, that healthy eating identity is significantly associated with diet quality, are in agreement with prior association studies on this topic.[5, 6] It has been suggested that healthy eating identity could have an impact on diet quality through the moderating effect of planning such that healthy eaters may also be good planners who are able to come up with better arrangements to obtain healthier foods. [6]

This study has some limitations. Reduction of sample size through discarding unmatched control and further using a portion of the matched data to estimate design sensitivity, might lead to reduction of the power of the study. Since our treatment group (high healthy eating identity) remained unchanged, and only the control group ((low healthy eating identity)) decreased in size, the overall power may not have reduced very much [28] as precision, in a two-sample mean comparison study, is driven by the smaller of the two groups.[29] Matching will also likely increase power since it ensures that groups being compared are more alike relative to before matching (39). In one study, Lim *et al* showed that for matched sampling designs based on PS, it is possible that the matched sample fails to represent the characteristics of the population it was drawn from, thus negatively impacting external validity of findings.[30] This study makes no attempt to address this issue, however during determination of which algorithms are competitive in our proposed strategy, one can address possible external validity issues by eliminating algorithms whose matched samples differ significantly from the study population. The sensitivity analysis to unmeasured biases in this study was based on Wilcoxon rank test statistic, however it has been noted that poor choice of test statistic can lead to a mistaken view of the magnitude of unmeasured bias to which an observational study is sensitive (or insensitive).[31] It is possible that doing sensitivity analysis using a different test statistic would have led to different sensitivity analysis results. Our definition of healthy eating identity was also conservative since individuals had to “strongly agree” with all the three questions on the healthy eating identity scale in order to qualify to be in the high healthy eating identity group. A less conservative definition of healthy eating identity might lead to different results. Future research should focus on how healthy eating identity impacts individual’s components of the HEI-2010 and 2015-HEI rather than the total HEI score to allow more targeted interventions.

# CONCLUSION

﻿ Evidence has shown that diet is a critical component in diseases of public health concern in midlife including obesity. Understanding the causal impact of eating identity on diet quality is important in designing interventions and advocating for policies to prevent obesity and other comorbidities. Comparing the performance of multiple matching algorithms can be an important step in designing observational studies to evaluate causal effects in natural experiments. Using a predetermined balance criterion together with an assessment of design sensitivity is an effective way to optimize the robustness of a study design to both measured and unmeasured biases on the treatment effect estimate.

# REFERENCES:

1. Drewnowski A, Kawachi I. Diets and Health: How Food Decisions Are Shaped by Biology, Economics, Geography, and Social Interactions. Big data. 2015;3:193–7. doi:10.1089/big.2015.0014.

2. : U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015-2020 Dietary Guidelines for Americans. 2015. http://health.gov/dietaryguidelines/2015/guidelines/. Accessed 2 May 2018.

3. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. Obesity. 2015;23:1331–44.

4. Higgs S, Thomas J. Social influences on eating. Curr Opin Behav Sci. 2016;9:1–6. doi:10.1016/J.COBEHA.2015.10.005.

5. Kendzierski D, Costello MC. Healthy Eating Self-Schema and Nutrition Behavior. J Appl Soc Psychol. 2004;34:2437–51.

6. Strachan SM, Brawley LR. Healthy-eater Identity and Self-efficacy Predict Healthy Eating Behavior A Prospective View. J Health Psychol. 14:684–95. doi:10.1177/1359105309104915.

7. Blake CE, Bell BA, Freedman DA, Colabianchi N, Liese AD. The Eating Identity Type Inventory (EITI). Development and associations with diet. Appetite. 2013;69:15–22. doi:10.1016/j.appet.2013.05.008.

8. Cochran WG. Analysis of covariance: its nature and uses. Biometrics. 1957;13:261–81.

9. Cochran WG, Rubin DB. Controlling bias in observational studies: A review. In: Matched Sampling for Causal Effects. 2006. p. 30–58.

10. Rubin DB. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies Author ( s ): Donald B . Rubin Published by : International Biometric Society Stable URL : https://www.jstor.org/stable/2529685 International Biometric Society i. 2018;29:185–203.

11. Emmons KM, Doubeni CA, Fernandez ME, Miglioretti DL, Samet JM. National Institutes of Health Pathways to Prevention Workshop : Methods for Evaluating Natural Experiments in Obesity. 2017;:0–21.

12. Ver Ploeg M, Breneman V, Dutko P, Williams R, Snyder S, Dicken C, et al. Access to affordable and nutritious food: updated estimates of distance to supermarkets using 2010 data. 2012.

13. McKinnon L. Household size: An overlooked confounder when considering inequalities in food choice. Australian and New Zealand Journal of Public Health. 2014.

14. Oyserman D, Fryberg SA, Yoder N. Identity-Based Motivation and Health. J Pers Soc Psychol. 2007;93:1011–27.

15. Gittelsohn J, Rowan M, Gadhoke P. Interventions in small food stores to change the food environment, improve diet, and reduce risk of chronic disease. Prev Chronic Dis. 2012;9.

16. Ball K, Jeffery RW, Abbott G, McNaughton SA, Crawford D. Is healthy behavior contagious: associations of social norms with physical activity and healthy eating. Int J Behav Nutr Phys Act. 2010;7:86. doi:10.1186/1479-5868-7-86.

17. University of Minnesota, Minneapolis MN. Nutrient Data System for Research (NDSR). 2015. http://license.umn.edu/technologies/ndsr87072\_nutrition-data-system-for-research-nutritional-analysis-software. Accessed 10 Jun 2018.

18. Nutrition Coordinating Center (NCC), University of Minnesota. Guide to Creating Variables, Needed to Calculate Scores for Each Component of the Healthy Eating Index-2015 (HEI-2015). http://www.ncc.umn.edu/healthy-eating-index-hei/. Accessed 17 Jun 2018.

19. U.S. Department of Health and Human Services USD of A. 2015-2020 Dietary Guidelines. 2015. https://health.gov/dietaryguidelines/2015/guidelines/. Accessed 1 Apr 2018.

20. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–8. doi:10.1093/bioinformatics/btr597.

21. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivar Behav Res. 2011;46:399–424.

22. Rubin DB. Using propensity scores to help design observational studies: Application to the tobacco litigation. Health Serv Outcomes Res Methodol. 2002;2:169–88.

23. Heller R, Rosenbaum PR, Small DS. Split Samples and Design Sensitivity in Observational Studies. J Am Stat Assoc. 2009;104:1090–101. doi:10.1198/jasa.2009.tm08338.

24. Rosenbaum PR. Design sensitivity in observational studies. Biometrika. 2004;91:153–64. doi:10.1093/biomet/91.1.153.

25. Rosenbaum PR. Observational studies. New York: Springer; 2002.

26. Wilson MM, Reedy J, Krebs-Smith SM. American Diet Quality: Where It Is, Where It Is Heading, and What It Could Be. J Acad Nutr Diet. 2016;116:302-310.e1. doi:10.1016/J.JAND.2015.09.020.

27. Ventura Marra M, Thuppal S, Johnson E, Bailey R. Validation of a Dietary Screening Tool in a Middle-Aged Appalachian Population. Nutrients. 2018;10:345. doi:10.3390/nu10030345.

28. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit Anal. 2007;15:199–236.

29. Cohen J. Statistical Power Analysis for the Behavioral Sciences Second Edition. http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf. Accessed 28 Sep 2018.

30. Lim S, Marcus SM, Singh TP, Harris TG, Levanon Seligson A. Bias due to sample selection in propensity score matching for a supportive housing program evaluation in New York City. PLoS One. 2014;9:e109112. doi:10.1371/journal.pone.0109112.

31. Rosenbaum PR. Design of Observational Studies. New York: Springer US; 2009.

Define a selection model to determine similarity (distance)between cases and controls to be matched

Match cases to controls using all available matching algorithms

Assess balance of baseline variables for all matching algorithms

If more than one algorithm produces a matched sample meeting prespecified balance criteria, compute design sensitivity statistic using planning samples from each algorithm and determine algorithm with highest value.

Use analysis sample from algorithm with highest design sensitivity for outcome analysis

Need to improve balance of baseline variables?

High healthy eating identity

(n = 105)

Low healthy eating identity

n = (410)

Screened (n=1,395)

Eligible (n = 655)

Participated in psychosocial and dietary recalls surveys baseline surveys (n = 515)

Study inclusion criteria: study of primary food shoppers residing in selected food desserts in Cleveland and Columbus , 2016. After screening, participants who lived outside study boundaries, did not plan to stay at current address for more than 12 moths or moved within 30 days, did not do more than 50% of food shopping for household, English not first language, whose household members are already in the study participants are members of household, were not interested in participating or withdrew within 30 days’ and did not provide consent were excluded. During survey administration, we collected 3 dietary recalls over 3o days for each participant. Participants who strongly agreed with all the questions on the healthy eating identity scale were defined as having high healthy eating identity otherwise they were defined as having low healthy eating identity.

|  |  |  |
| --- | --- | --- |
| Table 1. Description of Scales in the Psychosocial Survey of Factors Influencing Dietary Behavior and Eating Identity | | |
| **Scale** | **Questions** | **Response scale** |
| Perceptions of Healthy Food Access(42) | 1. A large selection of fresh fruits and vegetables is available in my neighborhood.  2. The fresh fruits and vegetables in my neighborhood are of high quality.  3. A large selection of low-fat products is available in my neighborhood. | 4-Likert Scale  (Strongly Disagree to Strongly agree) with don’t know option |
| Perceived Control of Food Shopping(43) | 1. I have enough time to shop for fresh and healthy foods.  2. It is convenient for me to purchase fresh and healthy foods.  3. I think that fresh and healthy foods are expensive. | 4-Likert Scale  (Strongly Disagree to Strongly agree) with don’t know option |
| Social Support and Norms for Healthy Eating and Food Shopping(44) | How often during the past 12 months have members of your family/neighbors?  1. Eaten fresh and healthy foods with you,  2. Encouraged you to eat fresh and healthy foods,  3. Discouraged you from eating unhealthy foods,  4. Told you about fresh and healthy foods and how to prepare them,  5. Prepared fresh and healthy food with you | 5-Likert Scale (Never to Most of the time) with a don’t know option |
| Sense of Community(45) | Thinking about your neighborhood overall, please rate how much you agree or disagree with the following statements about your neighborhood?  1. You can get what you need in your neighborhood,  2. Your neighborhood helps you fulfill your needs,  3. You feel like a member of your neighborhood,  4. You belong in your neighborhood,  5. You have a say about what goes on in your neighborhood,  6. People in your neighborhood are good at influencing each other, g. You feel connected to your neighborhood,  7. There is a strong sense of community spirit in your neighborhood,  8. You believe your neighborhood is changing for the better,  9. Your neighborhood seems like the kind of place where one person can make a difference,  10. Your neighborhood is the kind of place you’d like to live. | 4-Likert Scale  (Strongly Disagree to Strongly agree) with don’t know option |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2. Baseline Characteristics by Healthy Eating Identity Group before Matching of Study Participants from Cleveland and Columbus, OH, 2016 | | | | |
|  | High Healthy Eating Identity a (N = 105) | Low Healthy Eating Identity b (N= 410) | P c | Standardized differences |
| *Demographic variables:* |  |  |  |  |
| Age in years, mean (SD) | 51.42 (12.73) | 48.80 (13.76) | 0.078 | 0.210 |
| Female , n (%) | 79 (75.2) | 298 (72.7) | 0.686 | 0.026 |
| At least high school level of education, n (%) | 33 (31.4) | 169 (41.2) | 0.085 | -0.010 |
| Less than $20,000 annual income, n (%) | 76 (72.4) | 276 (67.3) | 0.38 | -0.051 |
| African Americans, n (%) | 75 (71.4) | 282 (68.8) | 0.684 | 0.026 |
| Location: Cleveland, n (%) | 40 (38.1) | 216 (52.7) | 0.011 | -0.145 |
| *Food Environment and community connection variables:* |  |  |  |  |
| Perception of Healthy Food Access, mean (SD) | 2.97 (0.94) | 2.69 (0.91) | 0.005 | 0.300 |
| Perceived Control of Food Shopping, mean (SD) | 3.43 (0.60) | 3.01 (0.67) | <0.001 | 0.696 |
| Family SSN for healthy eating, mean (SD) | 3.35 (0.83) | 3.03 (0.81) | <0.001 | 0.382 |
| Neighborhood SSN for healthy eating, mean (SD) | 2.08 (1.03) | 1.77 (0.79) | 0.001 | 0.301 |
| Sense of Community, mean (SD) | 3.02 (0.77) | 2.65 (0.73) | <0.001 | 0.474 |
| *Shopping Behavior and meal preparation variables:* |  |  |  |  |
| Involvement in more than 1/2 of meal preparations | 94 (89.5) | 313 (76.3) | 0.005 | 0.132 |
| Participation in all shopping, n (%) | 95 (85.7) | 328 (80.0) | 0.232 | 0.057 |
| Shopped at a FM at least once, n(%) | 78 (74.3) | 237 (57.8) | 0.003 | 0.165 |
| Buy fast food less than once per week, n(%) | 82 (78.1) | 256 (62.4) | 0.004 | 0.157 |

Abbreviations: SD, Standard Deviation; SSN = Social Support and Norms.

a Participants who strongly agreed with all questions on the healthy eating identity scale.

b Participants who did not strongly agree with all questions on the healthy eating identity scale.

Figure 1:

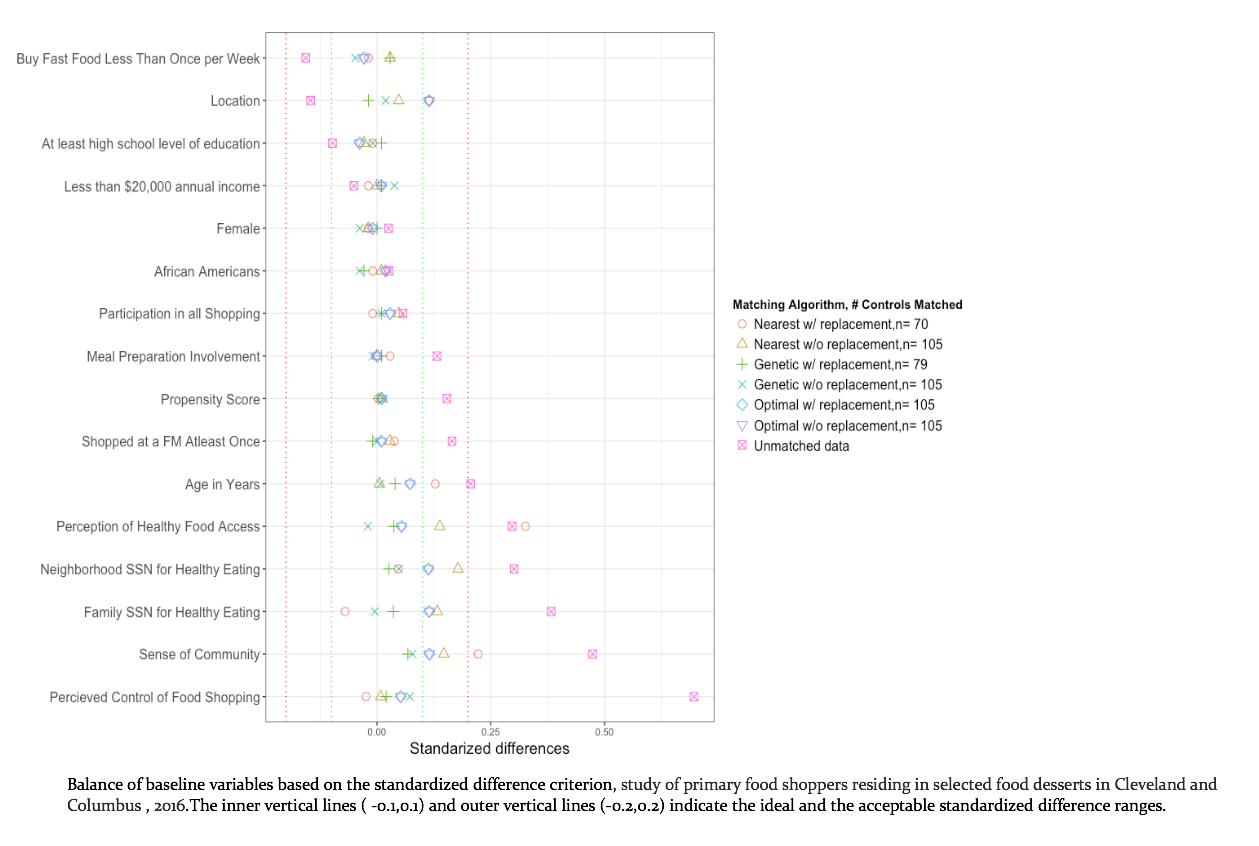
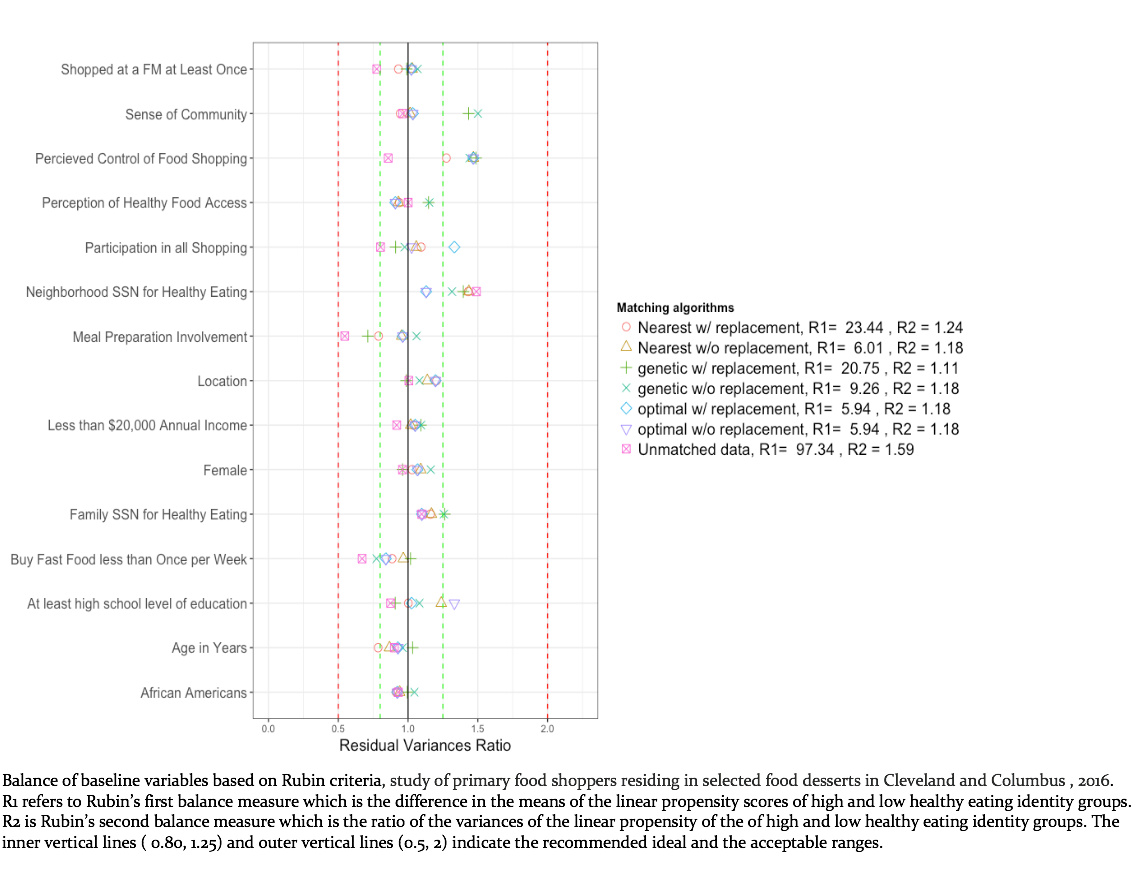


Figure 2:



|  |  |  |
| --- | --- | --- |
| Table 3. Design Sensitivity Values for Study Designs based on Competing Matching Algorithms | | |
| Matching Method | Design Sensitivity |
| Genetic without replacement | 1.70 |
| Nearest without replacement | 1.65 |
| Genetic with replacement | 1.38 |
| Optimal with replacement | 1.22 |
| Optimal without replacement | 1.22 |