Breakfast of Champions
The First Meal of the Day and Academic Success

Michael Discenza, Timothy Haley, and Jonah Smith

May 6, 2014

Abstract

We investigate whether eating breakfast influences academic performance using data from the National Longitudinal Study of Adolescent Health. Using a simple two-sample comparison, we estimate that breakfast eaters have, on average, a higher grade point average (GPA) by 0.17 on a four-point scale (95 percent confidence interval: 0.10 to 0.24; two-sided p-value: <0.001). Because these data are observational, we adjust for non-random treatment assignment using linear regression approaches and propensity score matching. Controlling for the predictors identified programmatically (stepwise regression using a large variable space), we find that students who ate breakfast have, on average, higher GPAs by 0.091 points (95 percent confidence interval: 0.0026 to 0.179; two-sided p-value: <0.05). Adjusting for established demographic predictors of educational outcomes, we estimate this effect to be 0.20 (95 percent confidence interval: 0.14 to 0.27; two-sided p-value: <0.001). Using a one-sample t-test on propensity score-matched samples, we estimate the breakfast effect to be 0.22 (95 percent confidence interval: 0.13 to 0.31; two-sided p-value: <0.001). We conclude that eating breakfast has a positive effect on academic performance. Given the survey's sampling scheme, we suggest that these results are generalizable to the larger American adolescent population.

1 Introduction

1.1 Problem of Interest

Growing up, we remember our parents, our teachers, and the media repeatedly telling us that we could not do our best if we did not eat breakfast. We, however, believe the success of our nation's youth is far too important to leave to speculation and folklore. We chose to explore the connection between eating breakfast and success using statistical tools and observational data.

Success is a multifaceted concept. For our purposes, we focused on academic success, which we operationalized using one important metric for a young-person: grade point average (GPA).

In particular, we ask two questions. First, do students who eat breakfast tend to get higher grades in school? This is a simple observational question, but it has important implications for popular perceptions about the relationship between eating breakfast and success. Second, we ask: is there an independent breakfast effect on high school grades? In other words, is the relationship spurious—that is, generated by confounding variables—or is there actually a probable causal link between eating breakfast and doing well in school?

1.2 Data Overview

To answer these questions, we used data from the National Longitudinal Study of Adolescent Health (Add Health), the "largest, most comprehensive longitudinal study of adolescents ever undertaken" [1] [10]. Our sample included a total of 3918 subjects who responded to questions in the first wave of the survey, administered in seventh grade, and the third wave, administered near the time of the subjects' graduation from high school. The first wave included most of the demographic information we use in our models, as well as information about breakfast-eating behavior. Although the survey collected more detailed information about breakfast habits, we used a simple binary variable, which indidcated whether students generally eat 'nothing' for breakfast. Cumulative high school GPA was self-reported in the third wave.

Among the subjects included in our dataset, 3181 reported that they generally eat breakfast, while 737 reported that they generally did not. Because we only had information about breakfast eating at one point in time, we had to make the assumption that students' breakfast-eating remained consistent throughout their secondary school years. We are also unable to adjust our analysis for frequency of breakfast eating, as subjects are asked to generalize about their breakfast habits.

Because subjects were selected using a probability sampling scheme, namely cluster sampling, we are able to generalize our findings to the population of American adolescents.

1.3 Analytical Approach

First, we conducted a simple two-sample test. We tested whether people who eat breakfast tend to do better in school, without controlling for other variables, in order to determine whether there is any observable association between breakfast-eating and GPA.

Though this type of test is a valid way to compare the two groups, it does not allow us to make conclusions about the *effect* of breakfast, as the data are observational. Subjects are not randomly assigned to eat breakfast, but rather eat breakfast for a complex variety of reasons. Thus, we try to identify and control for these complex reasons to uncover a 'pure' breakfast effect. We take two broad approaches: regression-based and propensity-score-based analyses.

For the regression-based strategies, we constructed base regression models for GPA, and then included breakfast to see if the effect was still significant, or whether the variation in GPA could be explained by confounding variables. These base models were constructed in two different ways: programmatically and theoretically. In the programmatic approach, we build a strong base model for GPA using stepwise selection. In the theoretical approach, we build our base model using commonly cited predictors of academic performance. In both cases, breakfast remains significant when added to the base model.

For the propensity-matched strategies, we construct a matched sample of individuals based on their a priori propensity for eating breakfast. This simulates randomized assignment to breakfast-eating by pairing individuals with opposite breakfast-eating habits who have otherwise equal probability of eating breakfast. If we then compare the GPAs of the two groups, we can estimate the pure breakfast effect. Alternatively, we can compute the estimated effect of the treatment on the treated, which is conceptually similar to a weighted average of the treatment effects, to adjust for the unequal distribution of demographic characteristics amongst the two different groups. Both of these propensity-matching strategies also suggest an independent breakfast contribution to GPA.

2 Two-Sample Group Comparison

We began with a very simple model to compare the GPAs of students who do and do not eat breakfast.

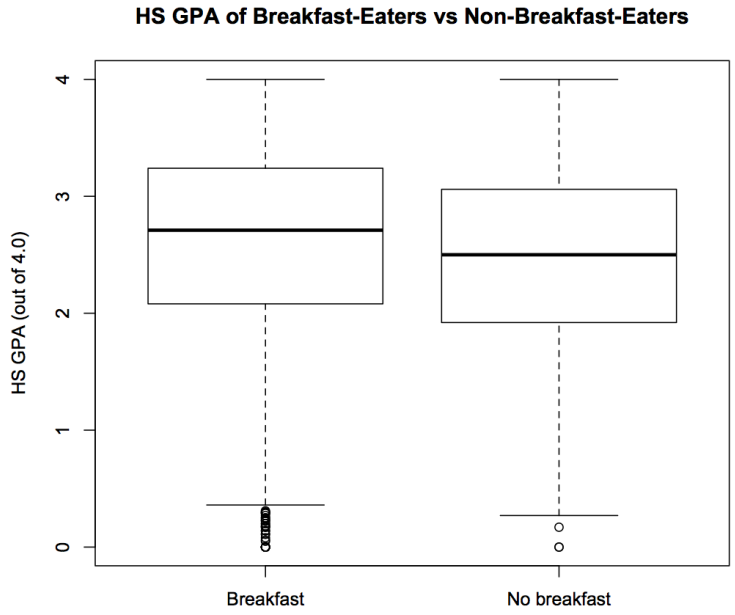

Figure 1: Box Plots of GPA by Breakfast Eating

As we can see, the two samples seem to have reasonably symmetric GPA distributions. Given this and the large n, we feel confident that deviations from normality are inconsequential and no transformations are needed. The more important assumption, equality of variance, is more difficult to justify (particularly because the F-test for equality of variance is very sensitive to deviation from normality), so we use the Welch correction, as implemented in R, to make conservative statistical judgments.

The two-sample t-test with Welch correction returns an estimated difference of means of 0.17 on a four point scale in favor of those who ate breakfast. This finding explains why people might *think* that eating breakfast is

Two-Sample T-Test with Welch Correction
t = 5.0509, df = 1116.32, **p-value** = 5.133e-07
mean of the differences = 0.169474
95 percent confidence interval: (0.1036393, 0.2353081)

Table 1: The results of a two-sample t-test

directly related to success. But to test the hypothesis that they are directly linked, we must apply techniques that do not have such strong requirements of random assignment to treatment, and that will allow us to control for variables that are not of direct interest but might confound the relationship between breakfast and high school GPA.

3 Regression models

3.1 Programmatic regression model

In order to control for the effects of variables other than breakfast behavior that might have associations with GPA, we first took a programmatic approach. Using stepwise variable selection, we fit a base model with GPA as the response and twenty-six candidate variables in the selection search space. Because finding the true 'best' model would have required fitting 2^{26} (more than 67 million) models, we determined that a practical solution was to use stepwise selection to approximate the best subset [5]. Incidentally this also helped us avoid the problem of multicollinearity that might have resulted from including highly correlated predictors.

We chose AIC over BIC as the selection/removal and stopping criterion because we wanted to strongly test the significance of the breakfast effect, and for this purpose less parsimony was in order. The algorithm was to include the variable that causes the largest reduction in AIC, and then remove the variable that causes the largest reduction in AIC, and repeat iteratively until the AIC cannot be decreased any further through the addition or removal of a single variable.

Once we obtained our base model, we then added the breakfast variable and found it to be significant. According to the programmatically selected model, students who ate breakfast have GPAs that are on average higher by 0.091 points (95 percent confidence interval: 0.0026 to 0.179; two-sided p-value: <0.05).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.6359	0.1817	9.00	0.0000
female	0.3088	0.0396	7.80	0.0000
hrsSleep	-0.0264	0.0141	-1.87	0.0612
religiosity	-0.0890	0.0180	-4.94	0.0000
R:Black	-0.4438	0.0567	-7.83	0.0000
R:NativeAm	0.0826	0.1736	0.48	0.6342
R:Asian	-0.1319	0.1090	-1.21	0.2265
R:Other	-0.1421	0.1016	-1.40	0.1621
enoughSleep	-0.0912	0.0435	-2.09	0.0364
attractiveness	0.0557	0.0267	2.09	0.0371
troublePayAttention	-0.0363	0.0187	-1.94	0.0528
intelligence	0.2575	0.0185	13.91	0.0000
personality	0.0327	0.0285	1.15	0.2501
hrsTV	-0.0046	0.0013	-3.48	0.0005
safeSchool	-0.0515	0.0179	-2.88	0.0041
hrsComputerGames	-0.0046	0.0029	-1.62	0.1059
feltDepressed	-0.0779	0.0254	-3.07	0.0022
daysSmoked	-0.0108	0.0017	-6.45	0.0000
timesSmokePot	-0.0013	0.0011	-1.23	0.2208
breakfast	0.0908	0.0449	2.02	0.0435

Table 2: Predictors from Programmatic Regression

$$\begin{aligned} GPA = \beta_0 &+ \beta_1(sex) + \beta_2(hrsSleep) + \beta_3(religiosity) + \beta_4(race) + \beta_5(enoughSleep) \\ &+ \beta_6(physicalAttractiveness) + \beta_7(troublePayAttention) + \beta_8(intelligence) + \beta_9(personality) \\ &+ \beta_{10}(hrsTV) + \beta_{11}(safeSchool) + \beta_{12}(hrsComputerGames) + \beta_{13}(feltDepressed) \\ &+ \beta_{14}(daysSmoked) + \beta_{15}(timesSmokePot) + \beta_{16}(breakfast) + \epsilon \end{aligned} \quad (1)$$

Our analysis of typical diagnostic plots showed that this model largely conformed to the assumptions for linear regression:

- *Residuals vs. fitted plot of the programmatic model:* showed no concerning, non-random patterns
- *normal Q-Q plot of the standardized residuals plotted against the theoretical quantile*: very close to the desired straight line, with only minor deviation at the tails
- *square root of the absolute standardized residuals:* consistent with what was previously observed in the residual plot and normal Q-Q plots
- *residuals versus leverage plot with Cook's Distance:* identified a potentially influential observation with very high leverage

We discovered a few potential outliers using the above plots as well as leverage computations, studentized residual computations, and Cook's Distance computations. We looked into each of the observations individually, but none exhibited particularly alarming responses in the context of the survey, with the exception of a single point with very high leverage. This respondent self-reported smoking pot 534 times in the past thirty days. This is by far the largest value for the "smoked pot" variable in our data set. (After reviewing the rest of the subject's responses, we speculate that the cereal he ate for breakfast every morning helped him achieve his impressive 1.51 GPA.)

We re-ran the programmatic model selection with the five potentially outlying observations omitted. Breakfast was still significant at the 95 percent confidence level, with an estimate of 0.08725 (95 percent confidence interval: 0.0005 to 0.174; two-sided p-value: 0.0486). After omitting these observations, the diagnostic plots look reasonable. Therefore, we fit the model will all of the observations as omission did not change our conclusion.

These results were important to us because they showed that the model chosen to programmatically approximate the best subset of variables using AIC [5] didn't break the significance of the relationship between breakfast and GPA, even after outlying observations were removed. This suggested to us the strong possibility of a direct relationship between breakfast and GPA.

3.2 Theoretical regression Model

In addition to the programmatic approach to controlling for confounders, we construct a model of high school GPA by identifying extraneous variables, and then we specifically target characteristics that would 'explain away' the variation in GPAs that is explained by breakfast eating (confounding variables) evident in the simple two-sample results [9]. These variables include personal characteristics (family income, race, sex, and a measure of cognitive aptitude) and a measure of an important confounding variable: parental involvement. If one's parents are more involved, they are more likely to encourage breakfast eating and schoolwork.

$$GPA = \beta_0 + \beta_1(familyIncome) + \beta_2(race) + \beta_3(sex) + \beta_4(cognitiveAbility) + \beta_5(parentalInvolvement) + \beta_6(breakfast) + \epsilon \quad (2)$$

Most of these variables were trivial to operationalize, either because they were directly asked/observed or because certain measures were constructed by the study's designers (such as cognitive aptitude, which was measured using a standardized exam). Parental involvement was less straightforward. Recognizing that this is a potentially important confounder, we read through the codebook and selected all of the variables that seemed to measure this characteristic. We then inserted each, one-by-one, into the model, picking the one that was most significant. It was a measure of the student's assessment of how much his/her parents 'care about' him/her, with higher values representing that parents care 'a lot,' and lower values 'very little.' Thus, higher values suggest more involved parents, and lower values represent less involved parents.

After constructing a matrix of scatterplots, we found that the distribution of incomes was heavily skewed and in need of transformation. A direct log transformation is not possible because there are zero values for family income (they are measured in thousands of dollars). However, because we are not interested in interpretability of the income coefficient, we can preserve the general relationship by transforming a shifted version of the income variable. Thus, the transformation log(X+1) is applied, where X is the income variable. After this, the scatterplot matrix looks better.

We then ran the full model on this dataset with the transformed income variable. The diagnostic plots all looked reasonable:

- *The Q-Q plot* suggests approximate normality in the error terms, with only slight deviations at the tails.
- *The residuals vs. fitted plot* looks slightly suspect at the low-GPA range, but given the limited scope of the behavior, it seems safe to proceed with interpretation.

- *The leverage plot* did not show signs of any particularly influential points.

Finally, we checked the variance inflation factors for signs of multicollinearity. They are all acceptably small (just above 1).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.0424	0.1705	-6.11	0.0000
log(income)	0.1780	0.0166	10.73	0.0000
R:Black	-0.1890	0.0342	-5.52	0.0000
R:NativeAm	-0.2841	0.1260	-2.25	0.0243
R:Asian	0.2521	0.0737	3.42	0.0006
R:Other	0.0077	0.0710	0.11	0.9138
Female	0.3984	0.0262	15.19	0.0000
Cognitive	0.0191	0.0010	18.71	0.0000
Parental Investment	0.1036	0.0272	3.81	0.0001
Breakfast	0.2018	0.0338	5.96	0.0000

Fitting the full model yielded a significant and positive parameter estimate for breakfast eating of 0.20 (95 percent confidence interval: 0.14 to 0.27; two-sided p-value <0.0001). This is larger than that of the two-sample test above. We conclude that, given equal levels of the above demographic variables, those who eat breakfast have a higher GPA on average than those who do not by 0.20 points.

4 Propensity score methods

We can use propensity scores as a non-regression-based approach to controlling for the influence of confounding variables. Identifying a vector of probabilities that each subject ate or did not eat breakfast, the propensity scores, allows us to understand the true effect of breakfast in two ways:

1. Testing for the difference in sample means of propensity-matched samples of treated and untreated subjects, as described and justified in [8].

2. Estimating the latent average effect of treatment on the treated subjects (ETT) by using the propensity scores to re-weight observations using principles from unequal probability sampling. [7] [6]

Each method allows us to interpret the results of our observational study as a "quasi-randomized experiment." We describe the application and report the results of both approaches.

Both of the uses of propensity scores require us to construct a model that can estimate the probability that each subject eats breakfast or not. Though this is an observational study, we refer to this variable using the term "treatment." The treatment variable is binary so we use a logistic regression model. We make the important assumption that we have observed all relevant pre-treatment factors affecting the assignment of subjects to the treatment or control groups.

$$logit(breakfast) = \beta_0 + \beta_1(familyIncome) + \beta_2(race) + \beta_3(sex) + \beta_4(cognitiveAbility) + \beta_5(parentalInvolvement) + \epsilon \quad (3)$$

4.1 Propensity score matching

For the first approach, with this vector of propensity scores, either in their logit form or back-transformed into odds ratios, which we call p, we can sort the subjects based on score. We next construct matched pairs from the treatment and control groups that have the most similar values of p. For regions of the distribution of p where there are only subjects that we observed as being in either one group or the other (no overlap), we drop these subjects from the analysis. The result is that we then have a sample of matched pairs where both the treatment and control subjects are balanced with respect to any observed variables that would affect treatment assignment.

The propensity matched sample was created using the R package `MatchIt` [4].

We can see from graph 3 that in creating the propensity-matched dataset, we need to ignore a number of observations in the high-density region of the breakfast distribution where there is a low probability of not eating breakfast (which makes intuitive sense). This means that our new matched sample includes breakfast-eating subjects selectively in regions of the propensity score distribution where there are not enough non-breakfast eaters. Again, the fact that the non-breakfast-eater density remains unchanged by the propensity matching procedure is evident in graph 4. This graph shows that the blue density curve and sample mean statistics

Breakfast_GPA	Breakfast_PropScore	Breakfast_GPA	Breakfast_PropScore
3.04	0.06	2.77	0.07
2.61	0.07	3.85	0.07
1.80	0.09	3.02	0.09
2.08	0.12	3.02	0.11
2.63	0.12	2.41	0.12
...	...	...	...
3.46	0.28	1.05	0.28
2.12	0.28	2.51	0.28
2.24	0.28	2.41	0.28
2.90	0.28	2.71	0.28
3.82	0.28	2.94	0.28

Table 3: The head and tail of the propensity matched data set, where each row contains a single matched pair in which the first subject ate breakfast and the second did not, but the two have approximately the same propensity toward breakfast-eating according to the model.

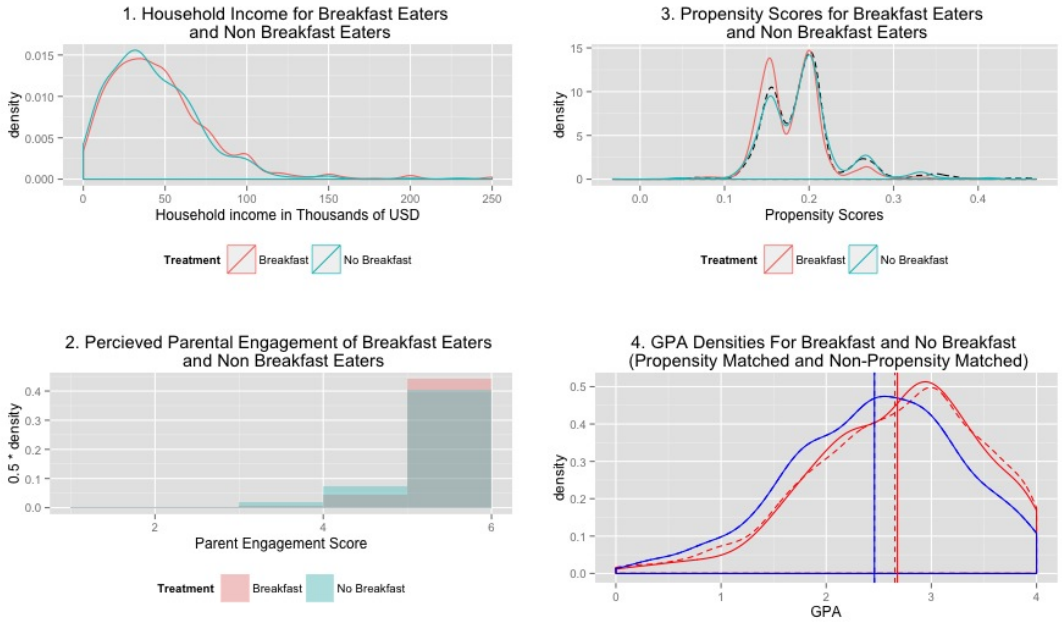

Figure 2: Diagnostic plots for propensity score matching. Note that in plot 3 the probability of no breakfast is graphed as opposed to breakfast and the dotted line represents the distribution of the propensity-matched dataset. In plot 4, red again represents breakfast and blue no breakfast. The solid lines represent the densities of the propensity-matched sample and the dashed, the raw samples. For breakfast eaters, these two densities are the same.

do not change under the application of propensity matching. The result is a set of observations in which the samples are balanced with respect to propensity. We have created a situation where the assignment of subjects to treatment and control groups is strongly ignorable [8].

We were then able to run a paired, one-sample t-test. We found that in compensating for non-random treatment assignment, we achieve an effect size of approximately 0.22, which is highly significant.

Matched Pairs T-Test for Mean Difference
t = 4.7521, df = 535, **p-value** = 2.591e-06
mean of the differences = 0.2186381
95 percent confidence interval: (0.1282579, 0.3090182)

Table 4: The results of a matched pairs t-test

4.2 ETT estimation using propensity scores

For our second use of propensity scores, we leverage the same p scores in a reweighting procedure that grew out of the field of survey sampling. We are able to use the p scores as weights to calculate an estimate of the "effect of treatment on the treated" (ATT or ETT) also known as the "average treatment effect on the treated" (ATET) [3] [2]. Just as weights are used to adjust for unequal probabilities of inclusion in samples, here the weights can allow us to compensate for non-independent assignment to treatment or control groups. We are thus constructing an unbiased estimate for the unobserved (or latent) counterfactual. [7] We achieve this by weighting the outcome variable by the p scores

$$f(T = 1|y, x) = f(T = 1|x) = p(x),$$

where T is the treatment value which can take values one or zero, and the same relationship holds true for the control group with probability $1 - p(x)$. The treatment assignment is independent of the outcome given our vector of pre-treatment xs.

This helps us approximate a situation in which there is random assignment and allows us to directly calculate the ETT. Our adjusted estimator for the effect size in a synthetic counterfactual setting is

$$ETT = E(Y_1|T = 1) - E(Y_0|T = 1),$$

which we calculate using the following formula:

$$ETT = E[E[Y_1 i|T = 1, p(x_i)] - E[Y_0 i|T = 0, p(x_i)]|T = 1].$$

The second term of the equality above is the unobserved counterfactual estimate. It can be expressed as a weighted average in terms of w which is a function of the weights, $w = p(x)/p(1 - x)$:

$$E[Y_0|T = 1] = \frac{\sum_{i=1}^{N} y_i w_i (1 - T_i)}{\sum_{i=1}^{N} w_i (1 - T_i)}$$

We then subtract that quantity from the observed sample mean of the treated subjects and obtain our unbiased estimate of the treatment effect on the treated.

The R code used to calculate the ETT was written by Columbia Professor Sid Dalal. Applying the weights as described, we found that the ETT is 0.113, indicating that eating breakfast increased a student's GPA by 0.113, on average, as compared to the unadjusted difference of 0.170 from our first analytical approach. Characterizing the distribution of this statistic is non-trivial, so we did not build a p-value or confidence interval.

5 Conclusion

Method	Estimate	95% Confidence Interval	p-value
Two-Sample t-test	0.17	0.10 to 0.24	<0.001
Regression: Programmatic	0.091	0.0026 to 0.179	<0.05
Regression: Theoretical	0.20	0.14 to 0.27	<0.001
Propensity-Matched T-test	0.22	0.13 to 0.31	<0.001
Propensity-Matched: ATT Treatment	0.11	*point estimate only	

We found the effects to be sizable and positive in all cases and significant in those that were easily estimable with a known distribution. We were fortunate in that our effect size and sample size were sufficiently large that we did not need to be particularly mindful of the efficiency of the techniques that we used. The regression techniques, for instance, used data more efficiently (but also required the estimation of more parameters), and propensity score matching led us to estimate the effect size with only 30 percent of the data. In cases where less data are available, we would have to think about our choice of methodologies more strategically.

It is worth adding two caveats. First, it is clear that there is not a direct causal link between breakfast and GPA. You are not graded on eating breakfast, and the act of eating breakfast does not magically increase GPA. It must operate largely through its effects on other factors. We propose two such pathways, which are not mutually exclusive: ability to focus and general health. Eating breakfast helps students focus, which allows them to perform better in school. Eating breakfast can also contribute to overall health, which is important for attendance and, again, focus, both of which will boost grades. The first pathway—improved focus—is consistent with experimental findings on college-aged students, which found improvements in short term mental performance as a result of eating breakfast [9]. Though we are unable to present findings in this paper, a preliminary mediation analysis reveals that there is evidence for both of these pathways.

The second caveat is that this study should not be understood as definitive proof of a causal relationship between breakfast eating and high school academic performance. There are still many other potentially confounding variables that we were unable to control for because of the availability of data. Chief among these is personality. Those who eat breakfast are also probably more likely to be diligent with their school work, for example.

Thus, we come to a somewhat tentative conclusion. Insofar as GPA is a reasonable measure of success, we have found evidence that eating breakfast does, indeed, help people succeed. At the very least, we can conclude that the type of people who eat breakfast are also more likely to succeed. Future research could focus on uncovering some of the causal pathways through which breakfast operates to improve outcomes, and could also extend our operationalization of success to other definitions, like interpersonal success or emotional well-being.

References

[1] UNC Carolina Population Center. About add health. Accessed May 5, 2014.

[2] Richard K Crump, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199, 2009.

[3] Sara Geneletti and A Philip Dawid. *Defining and identifying the effect of treatment on the treated*. Oxford University Press, 2011.

[4] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8):1–28, 2011.

[5] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 103 of *Springer Texts in Statistics*. Springer New York, New York, NY, 2013.

[6] Robert J LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, pages 604–620, 1986.

[7] Daniel F McCaffrey, Greg Ridgeway, and Andrew R Morral. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9(4):403, 2004.

[8] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[9] Mickey T Trockel, Michael D Barnes, and Dennis L Egget. Health-related variables and academic performance among first-year college students: implications for sleep and other behaviors. *Journal of American college health*, 49(3):125–131, 2000.

[10] J Richard Udry, National Longitudinal Study of Adolescent Health, et al. Waves i & ii, 1994–1996; wave iii, 2001–2002 [machine-readable data file and documentation]. *Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill*, 2003.

A Appendix

A.1 Roles

Note: We collaborated extensively on this project, so none of these descriptions are exhaustive or exclusive. These are simply the parts of the project for which we each took primary responsibility. In addition, we worked together to typeset and edit the paper.

- **Michael** completed and documented the propensity score analysis, wrote the "Data Overview" section, and created graphics.
- **Timothy** extracted variables from codebook using regular expressions, conducted initial variable selection, and conducted and documented the programmatic regression analysis.
- **Jonah** downloaded and combined the data from the different waves, wrote the abstract, conducted and documented the two-sample analysis and the theoretical model.