# Methods and Results

Ling Dai

## Data and Methods

The dataset used in this research was originally collected by Urminsky and Goswami through a 9-week field experiment in 2015 at four on-campus cafeterias at the University of Chicago, namely Gordon Cafe at the Gordon Center for Integrative Science, Harris Cafe at the Harris School of Public Policy, Law School Cafe at the University of Chicago Law School, and Stuart Cafe at the Stuart Hall. Because raw data were collected in the form of digital receipts, to make effective use of the data, I performed in this study subsequent data processing and analysis, including text parsing, data cleaning, coding of caloric information, and analysis of the causal effects of nutritional poster interventions.

In these cafes, many of the food items were already labeled with non-salient calorie information. During the experiment, four different signs, displayed at the entrance to the cafe, were tested. Two of the signs were informational and two were “mere reminders”. Specifically, Sign 1 showed “Do you know? Total Per-Meal (3 meals per day) Calories recommended is typically between 650-800 Calories.” Sign 2 displayed pictures of albacore tuna wraps, turkey and gouda wraps, and chicken Caesar salad, with their Calories per serving information listed. In contrast, Signs 3 and 4 had non-informational, mere-reminder content, with “Calorie information is available for many of the pre-packaged items we carry in this cafe” and “Do you know how many calories are there in your lunch today?” printed, respectively. (Figure 1)

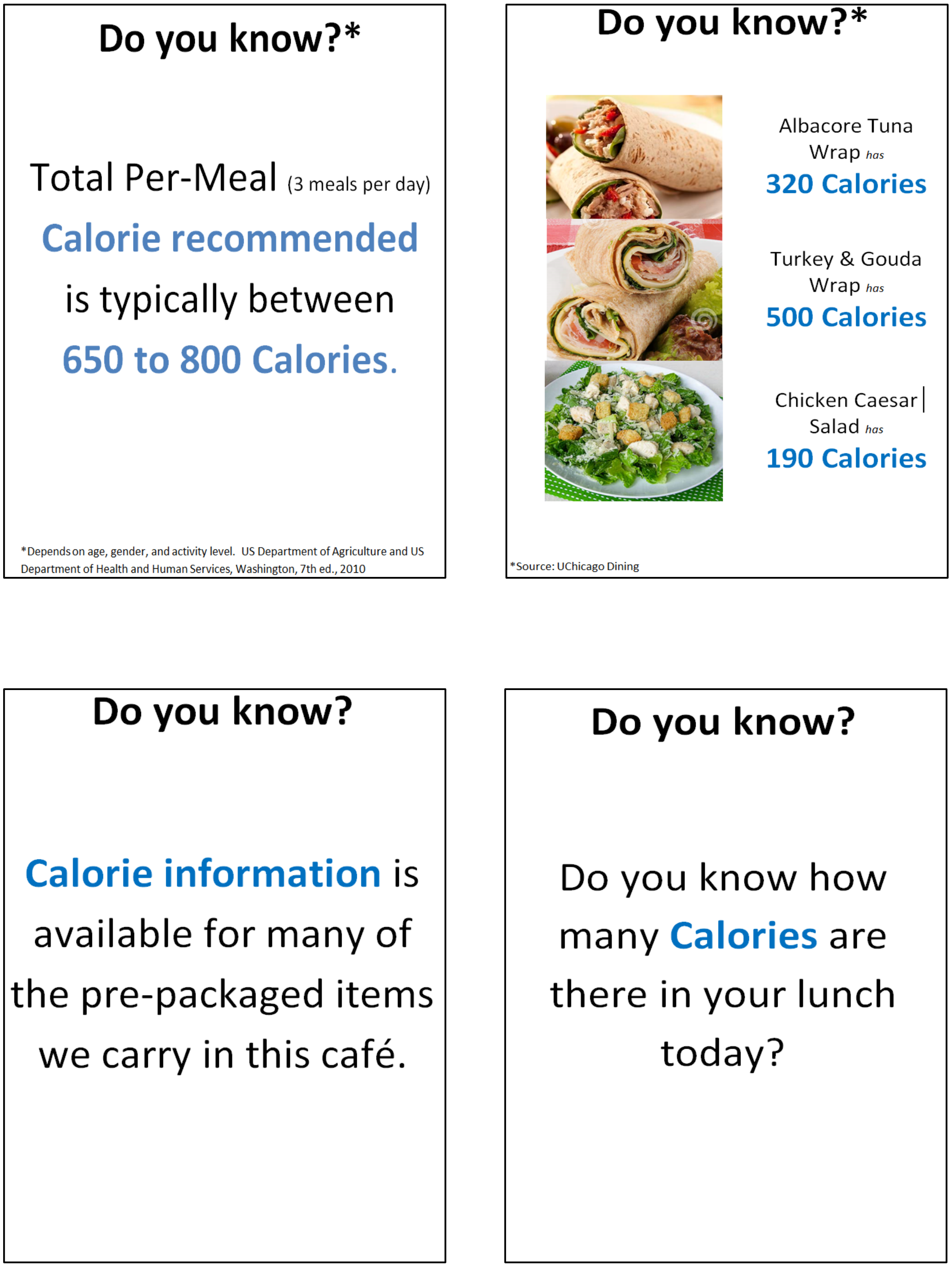


*Figure 1: Treatment Conditions: Sign 1 (Upper-left), Sign 2 (Upper-right), Sign 3 (Lower-left), and Sign 4 (Lower-right).*

| Week of: | 12th Jan | 19J | 26th Jan | 2F | 9th Feb | 16th Feb (no signage) | 19F | 2nd March |
|---|---|---|---|---|---|---|---|---|
| | W1 | W2 | W3 | W4 | W5 | W6 (Coupon Week) | W7 | W8 |
| **Law School Cafe** | Signage 1<br>Do you know?*<br>Total Per-Meal (3 meals per day) Calorie recommended is typically between 650 to 800 Calories. | | Signage 2<br>Do you know?*<br>Albacore Tuna Wrap 320 Calories<br>Turkey & Gouda Wrap 500 Calories<br>Chicken Caesar Salad 190 Calories | | Signage 3<br>Do you know?<br>Calorie information is available for many of the pre-packaged items we carry in this café. | $1 off Grab-n-Go Wraps Today<br>**(Mon, Tues, Wed)** | | Signage 4<br>Do you know?<br>Do you know how many Calories are there in your lunch today? |
| **Harris School Cafe** | Signage 4<br>Do you know?<br>Do you know how many Calories are there in your lunch today? | | Signage 3<br>Do you know?<br>Calorie information is available for many of the pre-packaged items we carry in this café. | | Signage 2<br>Do you know?*<br>Albacore Tuna Wrap 320 Calories<br>Turkey & Gouda Wrap 500 Calories<br>Chicken Caesar Salad 190 Calories | $1 off Lunch Today<br>**(Mon, Tues, Wed)** | | Signage 1<br>Do you know?*<br>Total Per-Meal (3 meals per day) Calorie recommended is typically between 650 to 800 Calories. |
| **Stuart Cafe** | Signage 3<br>Do you know?<br>Calorie information is available for many of the pre-packaged items we carry in this café. | | Signage 1<br>Do you know?*<br>Total Per-Meal (3 meals per day) Calorie recommended is typically between 650 to 800 Calories. | | Signage 4<br>Do you know?<br>Do you know how many Calories are there in your lunch today? | Eat Healthy! $1 off Lunch Today<br>**(Mon, Tues, Wed)** | | Signage 2<br>Do you know?*<br>Albacore Tuna Wrap 320 Calories<br>Turkey & Gouda Wrap 500 Calories<br>Chicken Caesar Salad 190 Calories |
| **Gordon Cafe** | Signage 2<br>Do you know?*<br>Albacore Tuna Wrap 320 Calories<br>Turkey & Gouda Wrap 500 Calories<br>Chicken Caesar Salad 190 Calories | | Signage 4<br>Do you know?<br>Do you know how many Calories are there in your lunch today? | | Signage 1<br>Do you know?*<br>Total Per-Meal (3 meals per day) Calorie recommended is typically between 650 to 800 Calories. | Eat Healthy! $1 off Salad Today<br>**(Mon, Tues, Wed)** | | Signage 3<br>Do you know?<br>Calorie information is available for many of the pre-packaged items we carry in this café. |

*Figure 2: Original Design of Interventions (Note: The interventions in red font color were not actually implemented).*

The four different signs were displayed at the on-campus cafes on a rotational basis during the first, third, fifth, and eighth week, each followed by a week of washout period with no signs. However, at the Law School Cafe, the sign was only displayed in the first week, as the experiment was halted by the Law School administration in the second week. As shown in Figure 2, during the first week, Sign 1 was displayed at the Law School Cafe, Sign 4 was displayed at the Harris School Cafe, Sign 3 was displayed at the Stuart Cafe, and Sign 2 was displayed at the Gordon Cafe. Similarly, during the third week, Signs 3, 1, 4 were displayed at the Harris School Cafe, the Stuart Cafe, and the Gordon Cafe, respectively. Signs 2, 4, 1 were displayed at the Harris School Cafe, the Stuart Cafe, and the Gordon Cafe, respectively, during the 5$^{th}$ week. Finally, during the 8$^{th}$ week, we displayed Sign 1 at the Harris School Cafe, Sign 2 at the Stuart Cafe, and Sign 3 at the Gordon cafe. No signage was displayed at any of the cafeterias during 2$^{nd}$, 4$^{th}$, 6$^{th}$, 7$^{th}$, and 9$^{th}$ week. During the 6$^{th}$ week, coupons were handed out to promote sales of a selected collection of low-calorie products, displaying different messages at the different Cafes.

The original data in the form of electronic receipts were stored in 561 separate text files downloaded from the Café vendors' proprietary POS system and came in two different formats. The data were scraped and parsed into an aggregated receipt-level dataset, which contains 132,294 observations after dropping duplicate records. To effectively store and analyze the information of each electronic check, I used a data structure that is commonly found in natural language processing (NLP). Because there are 348 unique items in our item database, I created a vector of length 348 for each electronic receipt, with an integer on each position of the vector indicating the number of the corresponding item purchased in that transaction. This data structure enables advantages in vector operations, thereby facilitating some important tasks such as calculating the number of items, summing up the total prices, as well as calorie lookup. The resulting aggregated dataset contains 49 days of sales data from the Spring of 2015, collected during the experiment, as well as 115 days of sales data from Spring and Winter of 2014 to serve as a baseline. For the remaining observations, a unique ID was created using the combination of date, cafe name, and the check number.

After data processing, the effectiveness of calorie labeling was evaluated by testing both average total Calories per transaction and average Calories per USD spent in the weeks when a sign was present to the same measures in the weeks when no sign was displayed. The calorie content for each sold item was matched using the databases at the cafeterias. Following that, data cleaning was performed to identify and exclude outliers and receipts that did not represent consumer purchases. Next, I compared average total calories per transaction during weeks with signage display to the average level during the baseline weeks, as well as the differences in food choices. Furthermore, for the Signs that are found to be effective, I also want to test if there is a spillover effect of calorie labeling. Last but not least, if the results suggest that the "mere-reminder" effect is a primary mechanism of effective calorie labeling, I would also like to test if such "mere-reminder" effect is context-intendent, in contrast to the context-dependent effect of the calorie labels with benchmarking information.

**Analysis and Results**

Some key independent variables that were controlled for in the regression analyses include total value of the transaction in USD, without tax (*Subtotal*) and the cafeteria where the transaction took place (*cafe*). The largest recorded single transaction has a *Subtotal* of 252.8 USD, compared to a mean *Subtotal* is 4.049 USD. The median, the 99th percentile, and the 99.9th percentile of *Subtotal* is 3.08, 12.78, and 21.41 USD, respectively. The small number of large transactions (e.g., above $20) are a concern in the subsequent data analysis for multiple reasons. First, these transactions are statistical outliers that can have a disproportionate effect on the results. Second, I am skeptical that these transactions reflect actual purchases by consumers for their own consumption. While we are unable to get definitive confirmation, based on our understanding of the Cafes, we believe that the very large transactions are either individuals purchasing for multiple people (e.g., a student who is picking up lunch for their study group) or may represent catering orders.

To identify outliers for exclusion in a way that is relatively independent of the hypotheses being tested, we focused on the number of items per receipt. In the data, the number of items per receipt (*#Items)* has a median value of 1 and a 99th percentile of 4. Based on these statistics, we decided that the few (< 1%) of transactions with more than 5 items are very unlikely to be a single meal for an individual and thus should be excluded from our subsequent analysis. We will also discuss the robustness of the results to this assumption.

The primary intended dependent variable in this study is the total calories of the items on each receipt. However, many receipts reported "miscellaneous" as a product category. Each item was supposed to be rung up in the café by pressing one of the buttons on a large programmable keypad linked to that item (the Cafes did not use UPC scanning). Our understanding is that items were rung up as "miscellaneous" primarily when the cashier could not find the item on the keypad and found it easier to just type in the price.  When individual items could be uniquely matched to a price, we used the price to infer the item and assign the calories. Among all 130,891 transactions with a positive *Subtotal* value, 45,856 transactions contained at least one miscellaneous item that does have a corresponding calorie content in our database and thus require imputation.

To impute the calorie content of the miscellaneous items, I first calculated *Subtotal* and total Calories for each receipt after excluding all the miscellaneous items, and then fitted a linear model with no intercept to predict the calorie content of miscellaneous items based on their prices (Figure 3). The training set of the linear model consists of all the transactions that have positive *Subtotal* values after excluding miscellaneous items, and the resulting linear model had a slope of 69.142 (e.g., 69 calories per US dollar). The new variable that denotes the calorie content of each receipt after the imputation procedure is named as *Predicted_Total_Calories*. The median of *Predicted_Total_Calories* is 318 kcal, and the mean is 294 kcal. Other dependent variables to analyze include calories per dollar spent (*Calories_per_Dollar*), which is calculated as *Predicted_Total_Calories / Subtotal* for all transactions with positive transaction value (e.g., excluding reimbursement receipts).

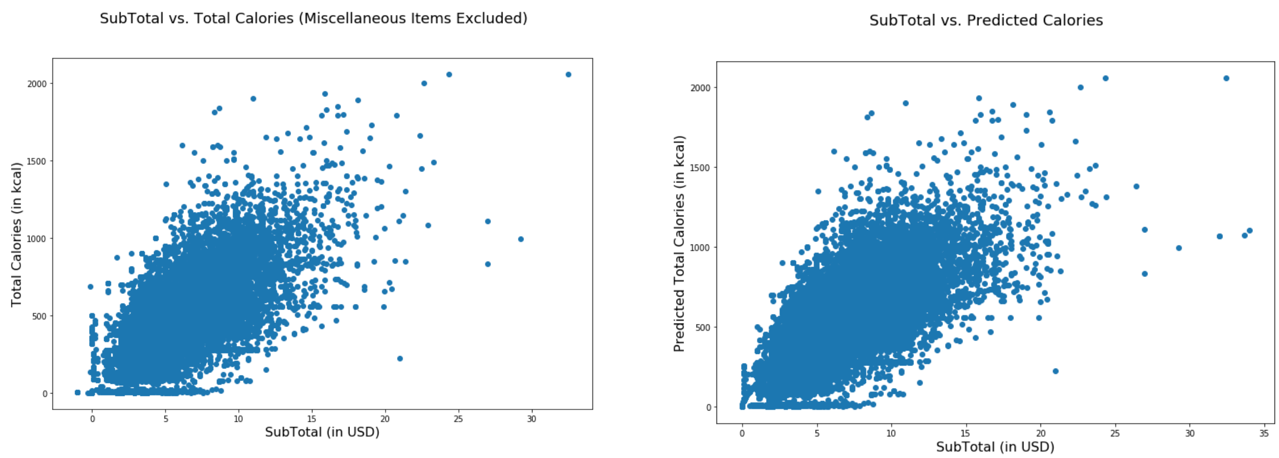


*Figure 3: Subtotal vs. Total Calories with Miscellaneous Items Excluded (Left); Subtotal vs. Predicted Total Calories (Right).*

Next, I aggregated the receipt-level data into a Café-date-level dataset, with each observation uniquely identified by the combination of date and cafeteria name. The resulting Café-date-level dataset contained 569 observations. For each observation in the Café-date-level dataset, *Predicted_Total_Calories* and *Subtotal* were calculated as the mean of the *Predicted_Total_Calories* and *Subtotal*, respectively, of all the recorded transactions of the corresponding observations. After that, *Calories_per_Dollar* was calculated as *Predicted_Total_Calories/Subtotal*. For each observation, we also created a variable *Count* to indicate the total number of recorded transactions that corresponds to the specific date and café. Within this dataset, the number of transactions per day varied substantially.  For example, the observation with the least transactions (Law School cafe on March 3, 2014) has only 21 recorded transactions, and the observation with the most transactions (Law School cafe on Feb 13, 2014) had 431 recorded transactions. Because of this large discrepancy within the total number of transactions for each observation in the Café-date-level dataset, instead of fitting OLS models, we choose to fit weighted least squares (WLS) linear models on the Café-date-level data, using *Count* as the weight.

Linear regressions were performed on the Café-date-level dataset using predicted total Calories per transaction as the outcome. Because the experiment was halted at the Law School cafe very early, we decide to first exclude all the observations at the Law School cafe completely from our regression analysis, and to later bring Law School cafe back into the analysis for the purpose of baselining and sensitivity analysis. The regression result suggests that there was a significant reduction in average total Calories per transaction during the display of Sign 1 and Sign 4 (Table 1, Model 5). Specifically, Sign 1 has an estimated effect of reducing total predicted Calories per transaction by 9.209 kcal ($p = 0.013$), and Sign 4 has an estimated effect of reducing total predicted Calories per transaction by 8.734 kcal ($p = 0.027$) (Appendix II, Table 1). According to the regression result, displaying Signs 2 and 3 was not associated with a significant reduction on calorie consumption during the experiment. However, it is noteworthy that the effects of Signs 1 and 4 are only statistically significant when we control for the fixed effect of week of quarter (as in Models 4 and 5). Because the estimated effects of the signs are confounded with week of quarter fixed effect, we believe that Models 4 and 5 yield more unbiased estimates of the poster effects compared to Models 1, 2, and 3. Furthermore, when we code the four signs as a single treatment called 'Signage', the regression results indicate the overall the Signs had a small but still significant impact on reducing the level of total Calories per transaction (-5.332 kcal, $p = 0.041$) (Table 2).

*Table 1: Summary of Linear Regression Results for Predicted Total Calories per Transaction (Café-date-level Data, with Law School Excluded)*

| | Confounders | | | | |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Variable | *Qtr* | *Qtr + café* | *café:Qtr* | *café:Qtr + WkofQtr* | *café:Qtr + WkofQtr + DayofWk* |
| *Intercept* | -12.126 | -35.450*** | -36.54*** | -38.8123*** | -39.071*** |
| *SubTotal* | 72.566*** | 78.6397*** | 74.647*** | 74.4841*** | 74.538*** |
| *2014Spring* | 1.315 | 2.0633 | 16.593*** | 15.1887*** | 15.193*** |
| *2014Winter* | 11.235*** | 12.5543*** | 18.912*** | 15.9747*** | 15.887*** |
| *Signage1* | -5.906 | -5.3032 | -5.668. | -9.1235* | -9.209* |
| *Signage2* | -0.272 | 0.0196 | -0.321 | -1.2564 | -1.442 |
| *Signage3* | -0.845 | -0.2728 | -0.391 | -1.7268 | -1.855 |
| *Signage4* | -7.081 | -7.3593. | -6.669. | -8.5440* | -8.734* |
| *GCIS:Coupon* | 3.274 | 0.5977 | 2.876 | -0.0634 | -0.224 |
| *Harris:Coupon* | 22.010* | 21.8715*** | 10.441 | 7.8146 | 7.690 |
| *Stuart:Coupon* | -1.886 | 4.2914 | 12.002 | 9.0801 | 9.905 |
| $R^2$ | 0.823 | 0.848 | 0.884 | 0.891 | 0.893 |

$*p < .05$ $**p < .01$ $***p < .001$

*Table 2: Summary of Linear Regression Results for Predicted Total Calories per Transaction (Café-date-level Data, all Signs as a Single Treatment, with Law School Excluded)*

| | Confounders | | | | |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Variable | *Qtr* | *Qtr + café* | *café:Qtr* | *café:Qtr + WkofQtr* | *café:Qtr + WkofQtr + DayofWk* |
| *Intercept* | -12.02 | -35.334** | -36.39*** | -37.9716*** | -38.208*** |
| *SubTotal* | 72.54*** | 78.610*** | 74.62*** | 74.4607*** | 74.508*** |
| *2014Spring* | 1.31 | 2.061 | 16.56*** | 15.1684*** | 15.170*** |
| *2014Winter* | 11.23*** | 12.552*** | 18.88*** | 15.9341*** | 15.843*** |
| *Signage* | -3.52 | -3.203 | -3.26 | -5.1863* | -5.332* |
| *GCIS:Coupon* | 3.27 | 0.610 | 2.93 | -0.0151 | -0.177 |
| *Harris:Coupon* | 22.00* | 21.846*** | 10.38 | 7.7448 | 7.620 |
| *Stuart:Coupon* | -1.87 | 4.292 | 11.98 | 9.0652 | 9.889 |
| $R^2$ | 0.822 | 0.847 | 0.883 | 0.890 | 0.891 |

$*p < .05$ $**p < .01$ $***p < .001$

Linear regression models were also fitted to predicted Calories per USD spent as the outcome, using *TotalSales* as the weight for each observation in the Café-date-level data. According to the regression results, both Sign 1 and Sign 4 were both associated with a significant decrease in the level of predicted Calories per USD spent during the experiment (Table 3, Model 5). The final model reported an estimated effect of -2.100 kcal/USD for Signage 1 ($p$ = 0.013) and an estimated coefficient of -1.986 kcal/USD for Sign 4 ($p$ = 0.027), suggesting that the customers chose lower-calorie items during the display of Signage 1 and Signage 4 compared to the baseline level. Moreover, regression results predicting *Subtotal* finds no significant effects of any of the signs on the average level of spend or on the average number of items purchased per transaction. These results suggest that Signs 1 and 4 reduced the calories purchased primarily by prompting customers to purchase lower-calorie items, rather than reducing the amount they spent.

*Table 3: Summary of Linear Regression Results for Avg. Calories per USD Spent (Café-date-level Data, with Law School Excluded)*

| | Confounders | | | | |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Variable | *Qtr* | *Qtr + café* | *café:Qtr* | *café:Qtr + WkofQtr* | *café:Qtr + WkofQtr + DayofWk* |
| *Intercept* | 67.8412*** | 62.4739*** | 62.1328*** | 61.428*** | 61.39198*** |
| *SubTotal* | 0.4454 | 1.8349** | 1.0286* | 1.018* | 1.02716* |
| *2014Spring* | 0.2887 | 0.4298 | 3.4127*** | 3.111*** | 3.10699*** |
| *2014Winter* | 2.5482*** | 2.8269*** | 3.9261*** | 3.286*** | 3.25614*** |
| *Signage1* | -1.3483 | -1.2062 | -1.2876. | -2.084* | -2.09962* |
| *Signage2* | -0.0515 | 0.0412 | -0.0564 | -0.233 | -0.27285 |
| *Signage3* | -0.2087 | -0.0759 | -0.0671 | -0.279 | -0.30826 |
| *Signage4* | -1.5621 | -1.6219. | -1.4540. | -1.943* | -1.98611* |
| *GCIS:Coupon* | 0.7466 | 0.1538 | 0.6690 | 0.030 | -0.00173 |
| *Harris:Coupon* | 5.4539** | 5.4294** | 2.7031 | 2.130 | 2.09100 |
| *Stuart:Coupon* | -0.3495 | 0.7880 | 2.3417 | 1.695 | 1.89614 |
| $R^2$ | 0.14 | 0.259 | 0.427 | 0.464 | 0.472 |

*$p$ < .05 **$p$ < .01 ***$p$ < .001

While the analyses above that I carried out in my MA thesis confirmed our initial hypothesis that effective visually salient labeling mainly affects the level of calorie consumption through changing the food choices of the customers, I also did follow-up analyses and additional hypothesis testing to further our understanding of the mechanism of effective labeling. For example, while I assumed in these analyses above that there are no effects of the posters when the posters are absent, it is in fact possible that consumers remain affected by the posters that had been previously displayed after they are removed. Therefore, to test these potential spillover effects of the posters, I also recoded the weeks following an intervention period as the respective washout periods. However, it is worth mentioning that doing this inevitably reduces the precision of the estimates of the poster effects, since only Week 1 of Spring 2015 can be used as the baseline, when controlling for Quarter fixed effects.

In order to minimize the confounding effect of Week 1 on the estimates, it is thus necessary to introduce week of quarter (*WkofQtr*) as a covariate in the models. The regression results with washout periods is shown below (Table 4). While Signs 1 and 4 both had similar effects in reducing calorie consumption during the intervention period, Sign 4 seemed to have a stronger spillover effect compared to Sign 1. This may be explained by the fact that because Sign 4 only serves as a "mere reminder", its effect may persist even when Sign 4 is no longer exhibited, whereas Sign 1 may have a relatively small spillover effect because few people remember the informational content of Sign 1 when it is no longer exhibited.

*Table 4: Summary of Linear Regression Results for Predicted Total Calories per Transaction (Café-date-level Data, with Law School Excluded, with Washout Periods)*

| | Confounders | | | |
|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 |
| Variable | *Qtr* | *Qtr + café* | *café:Qtr* | *café:Qtr + WkofQtr* |
| *Intercept* | -6.08 | -34.475*** | -32.50*** | -32.223*** |
| *SubTotal* | 72.30*** | 79.191*** | 74.51*** | 74.478*** |
| *2014Spring* | -4.67 | -4.307 | 8.77* | 8.539* |
| *2014Winter* | 4.60 | 5.604 | 9.44* | 9.142* |
| *Signage1* | -13.89* | -14.130** | -15.56** | -15.848*** |
| *S1 Wash-out* | -3.36 | -3.388 | -6.79 | -7.022 |
| *Signage2* | -5.96 | -6.394 | -7.69 | -8.080. |
| *S2 Wash-out* | -5.18 | -6.228 | -8.57. | -8.745. |
| *Signage3* | -7.39 | -7.368 | -8.16. | -8.476. |
| *S3 Wash-out* | -6.21 | -6.209 | -8.60. | -8.813. |
| *Signage4* | -14.90* | -15.611** | -14.97** | -15.346** |
| *S4 Wash-out* | -8.83 | -9.960. | -10.21* | -10.489* |
| *GCIS:Coupon* | -1.80 | -5.739 | -5.42 | -5.852 |
| *Harris:Coupon* | 16.96. | 15.551. | 2.46 | 2.086 |
| *Stuart:Coupon* | -7.02 | -2.230 | 3.71 | 4.325 |
| $R^2$ | 0.83 | 0.856 | 0.892 | 0.894 |

$^*p < .05$ $^{**}p < .01$ $^{***}p < .001$

Moreover, because the previous analyses results suggest that even non-informational calorie labels can reduce the average level of calorie consumption through reminding people of their health and diet, I am also curious about whether such effect is dependent on the context of labels. For example, while Sign 4 only asks people to think about the calorie content in their lunch, if the sign is able to make people aware of their diet, then people may also consume significantly less calories even when it is not lunch time. On the contrary, informational labels such as Sign 1 and Sign 2 should effectively reduce people's calorie intake in their matching contexts. According to my hypothesis, Sign 1 should only be effective around lunch time but Sign 4 should have relatively uniform level of effectives throughout a day.

To test my hypothesis above, I introduced the interaction term between the treatments and hour into my linear regression models. While the result suggests that Sign 1 indeed had a significant effect on reducing calorie intake around the lunch time (from 12:00 to 13:00), it also shows that Sign 1 significantly raised the average level of calorie intake between 9:00 and 10:00 (Figure 4). I think it might be caused by the fact that many students who have late breakfast during that time instead purchased more food after seeing the benchmarking data of Sign 1 (because it is about calorie content of a meal). Another surprising result is that, while Sign 2 was found to be ineffective in our previous analyses, it in fact strongly reduced the average level of calorie intake during dinner time (from 17:00 to 18:00). I believe it is due to the fact that people mostly purchase the hot food on Sign 2 during this time, but it needs to be confirmed in a follow-up analysis. Last but not least, the results also suggest that Sign 4 was effective only from 11:00 to 14:00, as opposed to what I theorized before. The result suggests that contrary to my hypothesis, even non-informational calorie labels are subject to the limitations of their contexts and tend to be ineffective when the context becomes irrelevant. Moreover, Sign 3 was found to be ineffective throughout the day.

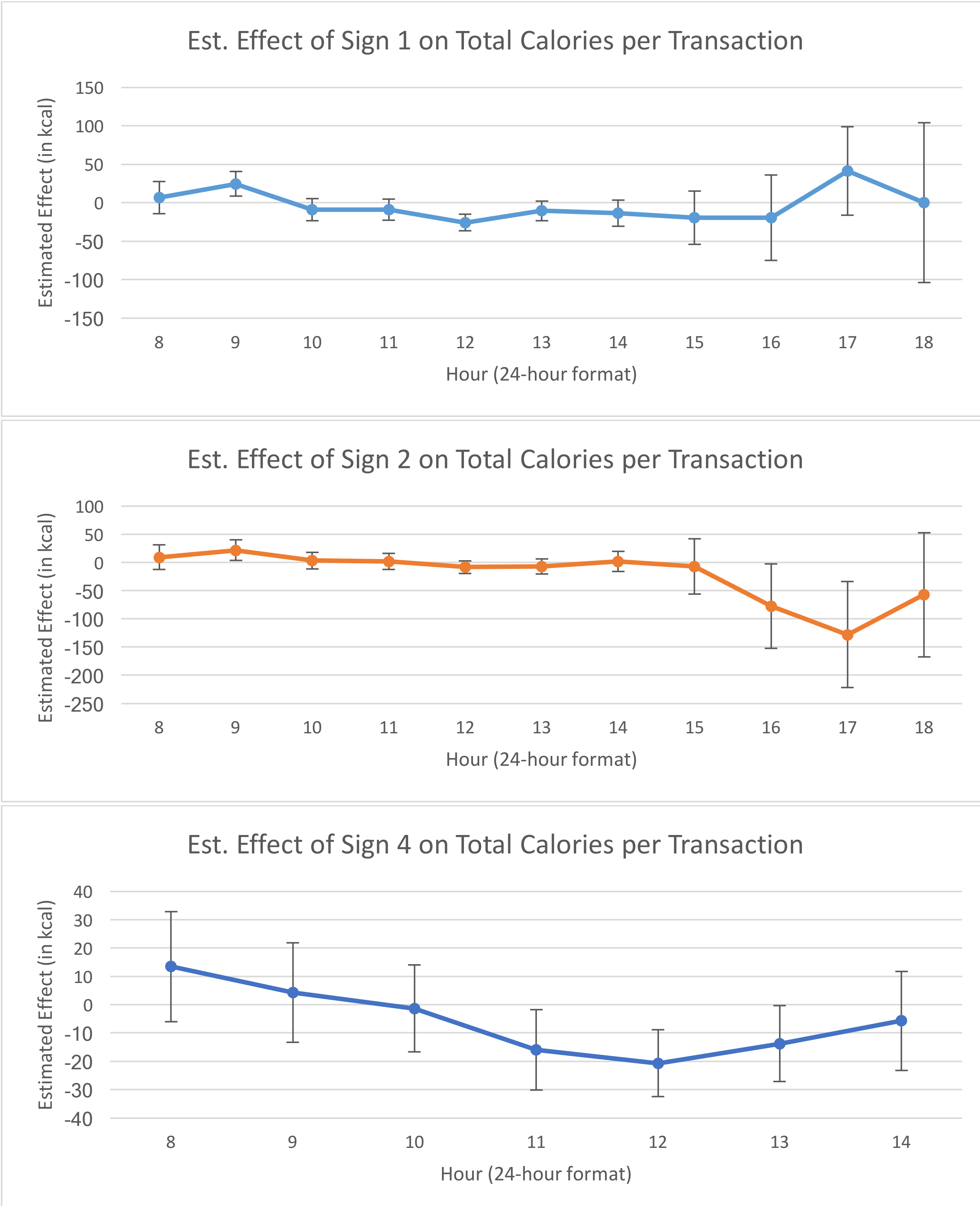


*Figure 4: Estimated Effect of Signs on Average Total Calories per Transaction During Different Time of a Day: Sign 1 (Upper); Sign 2 (Middle); Sign 4 (Lower).*