Audit Report: Workflow `9c1cab79a94242f68c5aab8136037a28`

Overview

- Source JSON: `reports\workflow-report-9c1cab79a94242f68c5aab8136037a28.json`
- Run ID: `9c1cab79a94242f68c5aab8136037a28`
- Papers input: `papers\Calorie Posting in Chain Restaurants - Bollinger et al. (2011).pdf`
- Started at: `2026-02-16T05:20:22.366907+00:00`
- Finished at: `2026-02-16T05:28:51.886045+00:00`
- Duration: `0:08:29.519138`

Effective Configuration

- Chat model: `gpt-5-nano`
- Embedding model: `text-embedding-3-large`
- Top K: `10`
- Chunk words / overlap: `350 / 75`
- Batch size: `64`
- Database URL configured: `True`

Step Outcomes

- `prep: completed`
- `ingest: num_pdfs=1, num_papers=1`
- `enrich: openalex=0, citec=0`
- `econ_data: fetched`
- `agentic: completed`
- `index: indexed`
- `report_store: pending`

Agentic Summary

- Status: `completed`
- Main question: What are the paper's main contribution, identification strategy, key results, and limitations?
- Report question set: `both`
- Structured questions generated: `84`
- Confidence mean/median: `0.2796909731048519 / 0.28416116738125424`
- Confidence labels: low=74, medium=9, high=0

Final Answer

- Main contribution

    - Establishes large-scale, granular evidence on the impact of NYC point-of-purchase calorie posting using a broad Starbucks dataset

across three cities (plus Seattle as placebo), documenting heterogeneous and primarily extensive-margin effects on calories per transaction. (Scope and data: page 4, words 275-576; Dataset: page 9, words 0-349)

- Identification strategy

  - Primary specification exploits a post-treatment indicator POST = 1 for NYC stores after April 1, 2008, with store and week fixed effects, weather controls, and week-specific effects; Seattle serves as a placebo/contrast. This constitutes a difference-in-differences–like setup with robust controls. (Key empirical setup: page 13, words 275-528)

- Key results

  - NYC: calories per transaction decline after posting, but the absolute decline is modest; a drop in log(calories per transaction) around posting is observed relative to controls. (Main findings: page 9, words 0-349; page 10, words 0-186)
  - Overall effect at the store-day level: calories per store-day fall about 4.6 calories, indicating part of the per-transaction decrease is offset by higher transaction volume in NYC. (Main findings: page 13, words 275-528)
  - Seattle results (for comparison): beverage calories per transaction fall by 4.6 calories; food calories per transaction rise by 0.8 calories, signaling item-type heterogeneity when not all items are posted. (Main findings: page 10, words 0-186)
  - Item-level effects (Table 4):
    * log(beverage calories per beverage) -0.008 (significant)
    * log(food calories per food item) -0.039 (larger, significant)
    * Net effect driven largely by the extensive margin (fewer food items purchased). (Main findings: page 18, words 0-349)
  - Extensive-margin channel: nearly three-quarters of the total calorie reduction comes from opting not to buy food items. (Main findings: page 18, words 0-349)
  - Distributional results: larger reductions at higher quantiles; roughly 5–6% declines from the 75th to the 99th percentile, with absolute reductions rising in the upper tail (e.g., about -77 calories at the 99th percentile). (Main findings: Table 6, page 22, words 0-349)

- Heterogeneity and implications

  - Effects persist up to 11 months post-posting; the paper discusses short-run versus longer-run persistence and variation by consumer response, product substitution, and consumer knowledge. (Heterogeneity/time path: page 9, words 0-349; page 13, 275-528; page 15, 0-254; page 18, 0-349)

- The data permit analyzing product substitution patterns and implications for profitability, reinforcing heterogeneity across cities and item types. (Heterogeneity implications: page 9, 0-349; page 15, 0-254; page 18, 0-349)

- Limitations and caveats

  - Some prior findings in the literature report non-significant effects with large standard errors; the paper emphasizes its larger scale and granularity relative to those studies. (Methodological notes: page 9, 0-349)
  - Confidentiality: exact values for some internal counts (N) are not disclosed; noted in the Appendix. (Methodological notes: page 18, 0-349)
  - Policy interpretation is nuanced: the measured effects are modest in absolute terms and partly offset by changes in transaction volume. (Main findings: page 9-10; page 13, 275-528)

- Quick formula/objective summary

  - Conceptual model: $\log(\text{Calories per transaction}) = \alpha_{\text{store}} + \gamma_{\text{week}} + \beta \cdot \text{POST} + X^\top \delta + \varepsilon$, with POST = 1 if NYC store after 1 Apr 2008; captures the posting effect. The same framework is extended to calories per transaction by item type and to calories per item. (Quick formula: page 13, 275-528; page 18, 0-349)

If you'd like, I can extract a compact table of key estimates, standard errors, and interpretation with precise page references. (Cited from: Primary model and Tables 3–6 descriptions: page 13, 275-528; page 18, 0-349)

Sub-Answers

Sub-answer 1

- Question: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')

- Question tokens estimate: `19`

- Answer:

- Scope and data (context for the study)

  - NYC calorie posting law required at point of purchase; discussion of costs and counterarguments. See context describing the policy and related views (page 4, words 275-576).
  - Dataset: universe of over 100 million Starbucks transactions in Boston, NYC, and Philadelphia over 14 months; 1.5 million

cardholder transactions; Seattle used as a placebo/contrast. See description of data and design (page 9, words 0-349).

- Key empirical setup
  - Primary specification uses a binary calorie-posting dummy POST = 1 for NYC stores after Apr 1, 2008; regression controls include fixed effects and weather, with week-specific effects. See discussion of Table 3/regression setup (page 13, words 275-528).

- Main findings
  - Calories per transaction declined in NYC after posting, but the decline is not large in absolute terms; NYC vs control comparison shows a drop in log(calories per transaction) around posting time. See Figure 1 and accompanying discussion (page 9, words 0-349; page 10, 0-186).
  - Overall calories per store-day fell by about 4.6 calories, reflecting that higher transaction volume in NYC offset part of the per-transaction decrease. See a key sentence tying the store-day figure to the per-transaction effect (page 13, words 275-528).
  - Seattle results: beverage calories per transaction fell by 4.6 calories; food calories per transaction rose by 0.8 calories, illustrating heterogeneous effects by item type when calories aren't posted for all items in that city (page 10, 0-186).
  - Beverages vs. foods (Table 4):
    * Number of beverages and foods per transaction; but calorie effects are item-specific:
      · log(beverage calories per beverage) -0.008 (significant)
      · log(food calories per food item) -0.039 (larger, significant)
      · Overall, substantial portion of calorie reductions come from consuming fewer food items (extensive margin). See Table 4 and associated notes (page 18, 0-349).
  - Extensive margin effect (dominant mechanism): nearly three quarters of total calorie reduction due to opting not to buy food items. See the explicit statement in Table 4 discussion (page 18, 0-349).
  - Distributional/quantile results: larger reductions at higher quantiles; the percent change is roughly 5–6% across 75th–99th percentiles, with absolute reductions rising in the upper tail (e.g., about -77 calories at the 99th percentile). See Table 6 and related text (page 22, 0-349).

- Heterogeneity and implications
  - Time path and longevity: the study tracks effects up to 11 months post-posting to assess persistence; results discuss potential short-run vs. long-run effects (page 9, words 0-349; page 13, 275-528).
  - Heterogeneity in consumer response and potential profit implications are analyzed; the data allow testing for product substitution patterns

4

and consumer knowledge (page 9, 0-349; page 15, 0-254; page 18, 0-349).

- Methodological notes

  – Prior related findings sometimes show no statistically significant effect due to large standard errors; Bollinger et al. position their approach as larger-scale and with more granular data (page 9, 0-349).
  – Confidentiality: exact values of some internal counts (N) are not disclosed; the Appendix notes this confidentiality constraint (page 18, 0-349).

- Policy context and cost considerations

  – The NYC calorie posting requirement is argued to have in-store/online calorie access; counterarguments cite implementation costs (e.g., about \$2,000 per Wendy's store) and potential indirect costs from multi-city menus (page 4, words 275-576).

- Quick formula/objective summary

  – Primary model (conceptual): log(Calories_per_transaction) = _store + _week + · POST + X +
  – POST = 1 if NYC store after 1 Apr 2008; otherwise 0
  – captures the effect of calorie posting on calories per transaction; same framework extended to items per transaction and per-item calories (as reported in Tables 3–6) (page 13, 275-528; page 18, 0-349).

If you want, I can extract specific numbers for a compact table (e.g., key estimates, standard errors, and qualitative interpretation) with precise page references.

Structured Q&A Appendix

This section mirrors `agentic.report_questions` for audit traceability.

A01: What is the main research question of the paper?

- Category: `A) Research question / contribution`
- Confidence: `low` (0.31602360829932447)
- Retrieval method: `local`
- Evidence type: `direct_quote_or_statement`
- Data source: Starbucks transaction data and cardholder data (NYC, Boston, Philadelphia), plus Seattle survey data.
- Answer:

Whether calorie posting affects people's behavior.

- Quote snippet:

Ultimately, whether calorie posting affects people's behavior is an empirical question.

- Citation anchors:
  - page=3, words=300-320, section=Introduction, note=Contains the claim that 'Ultimately, whether calorie posting affects people's behavior is an empirical question.'

A02: What is the paper's primary contribution relative to the existing literature?

- Category: `A) Research question / contribution`
- Confidence: `low (0.29785717396926265)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Bollinger, Leslie, Sorensen. Calorie Posting in Chain Restaurants, American Economic Journal: Economic Policy, 2011.
- Answer:

The paper's primary contribution is delivering a substantially larger and more comprehensive analysis of calorie-posting effects than prior work—using a dataset over 100 million Starbucks transactions, plus 1.5 million individual cardholder observations, across multiple cities, with long-run time variation, analysis of substitution and profit implications, and exploration of learning versus salience mechanisms.

- Quote snippet:

  the dataset we study is much larger and broader—the universe of over 100 million transactions at Starbucks in Boston, NYC, and Philadelphia over a 14-month period.

- Citation anchors:
  - page=8, words=0-349, section=Introduction, note=Authors claim their dataset is much larger and broader than prior NYC studies.
  - page=9, words=0-349, section=Introduction / Background, note=Prior work had smaller samples; authors emphasize breadth and depth of their data.

A03: What is the central hypothesis being tested?

- Category: `A) Research question / contribution`
- Confidence: `low (0.29013829293561877)`
- Retrieval method: `local`
- Evidence type: `central_hypothesis`
- Data source: Starbucks transaction and cardholder datasets (in-study)
- Answer:

The central hypothesis is that mandatory calorie posting in chain restaurants affects consumer behavior, notably reducing calories per transaction (through learning and/or salience).

- Quote snippet:

  whether calorie posting affects people's behavior is an empirical question.

- Citation anchors:
  - page=3, words=275-564, section=Introduction / Background, note=States that it is an empirical question whether calorie posting affects behavior and discusses learning vs salience as possible mechanisms.
  - page=29, words=0-349, section=Results – Commuters' and non-commuters' effects, note=Presents results showing reductions in calories per transaction, informing the hypothesis that posting affects behavior.
  - page=30, words=0-349, section=Discussion – Mechanisms, note=Concludes evidence supports both learning and salience as mechanisms, tying back to the central hypothesis about posting affecting behavior.

A04: What are the main outcomes of interest (dependent variables)?

- Category: `A) Research question / contribution`
- Confidence: `low (0.30457129143686446)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Transaction data (primary); cardholder data used to separate intensive-margin effects.
- Answer:

The main dependent variables are calories per transaction measures: beverage calories per transaction, food calories per transaction, and total calories per transaction (beverage + food calories), all in log form. The study also analyzes calories per beverage and calories per food item conditional on purchasing, as well as counts of items per transaction (number of beverages and number of food items).

- Quote snippet:

  In the top row, the dependent variable is log(beverage calories per transaction).

- Citation anchors:
  - page=11, words=0-349, section=Table 2—Changes in Cardholders' Beverage Choices Following Mandatory Calorie Posting (Treatment

and control results shown separately), note=Context for cardholder-based outcomes; framing of changes in beverage choices.
- page=12, words=0-349, section=Regression specification and Table 3 results, note=States dependent variables: log(beverage calories per transaction), log(food calories per transaction), and log(beverage + food calories).
- page=16, words=275-558, section=Table 4 – margins: number of items per transaction and calories per item purchased, note=Introduces dependent variables: number of beverages per transaction, number of food items per transaction, calories per beverage purchased, calories per food item purchased.
- page=20, words=275-551, section=Table 4 results for changes in calories and item counts, note=Reports results for calories per transaction and conditional calories per item.
- page=29, words=0-349, section=Table 9—Effects of Calorie Posting on Commuters' Log (Calories per transaction), note=Shows a regression specification with dependent variable log(calories + 1) for cardholder data.

A05: What are the key treatment/exposure variables (independent variables)?

- Category: `A) Research question / contribution`
- Confidence: `low (0.2439830212137138)`
- Retrieval method: `local`
- Evidence type: `Definition of treatment/exposure variable`
- Data source: transaction and cardholder datasets
- Answer:

The key treatment/exposure variable is POST, a dummy indicator set to one for NYC stores after the calorie posting began (April 1, 2008).

- Quote snippet:

  The POST variable is simply a dummy equal to one at NYC stores on every day after April 1, 2008

- Citation anchors:
  - page=11, words=0-349, section=Methods / Model specification, note=POS/T dummy equal to one after posting date (NYC stores after April 1, 2008)
  - page=11, words=275-522, section=Methods / Model specification, note=Explicit description that POST is a dummy equal to one if calories were posted (NYC after April 1, 2008)

A06: What setting/context does the paper study (country, market, period)?

- Category: `A) Research question / contribution`
- Confidence: `medium (0.3533264176203933)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Starbucks transaction and cardholder data from NYC, Boston, and Philadelphia (with Seattle surveys as controls).
- Answer:

United States; chain/fast-food coffee context (Starbucks); New York City calorie-posting policy implemented in mid-2008, with comparisons to Boston and Philadelphia, using data from NYC stores and those in two other cities during Jan 2008–Feb 2009 (policy effective April 1, 2008).

- Quote snippet:

  The law was first implemented in New York City (NYC) in mid-2008.

- Citation anchors:
  - page=1, words=0-20, section=Introduction, note= NYC calorie-posting law implemented in mid-2008; sets US setting.
  - page=5, words=0-60, section=Data Summary, note=Data cover NYC stores and Boston/Philadelphia stores; period before/after posting (Jan 1, 2008–Feb 28, 2009).
  - page=14, words=0-60, section=Results/Timeline, note=Calorie posting drop occurred around April 1, 2008; effects observed through Feb 2009.

A07: What is the main mechanism proposed by the authors?

- Category: `A) Research question / contribution`
- Confidence: `low (0.28583086382781325)`
- Retrieval method: `local`
- Evidence type: `textual evidence from the article`
- Data source: Starbucks transaction data
- Answer:

A combination of learning and salience effects.

- Quote snippet:

  Survey evidence and analysis of commuters suggests the mechanism for the effect is a combination of learning and salience.

- Citation anchors:
  - page=1, words=200-230, section=Introduction, note=Authors state the mechanism is a combination of learning and salience.

A08: What alternative mechanisms are discussed?

- Category: `A) Research question / contribution`
- Confidence: `low (0.22575046682164843)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Bollinger et al., American Economic Journal: Economic Policy, Calorie Posting in Chain Restaurants
- Answer:

The alternative mechanisms discussed are a learning effect and a salience effect.

- Quote snippet:

  One reason why calorie posting may affect consumer choice is a learning effect. Another possible explanation for the observed reduction in calories per transaction is a salience effect.

- Citation anchors:
  - page=25, words=0-349, section=A. Why is There an Effect?, note=Discusses learning and salience as alternative explanations for calorie posting effects.

A09: What are the main policy implications claimed by the paper?

- Category: `A) Research question / contribution`
- Confidence: `medium (0.3766236629941385)`
- Retrieval method: `local`
- Evidence type: `Policy implications (Discussion/Conclusion)`
- Data source: Starbucks NYC/Boston/Philadelphia transaction data; anonymous cardholder data; in-store surveys.
- Table/Figure: Figure 3 and Table 3 referenced in the results (sales and calorie effects by item).
- Answer:

Mandatory calorie posting yields modest direct health benefits: calories per Starbucks transaction fall about 6%, mainly due to changes in food purchases rather than beverages, while average profits are not significantly affected. The policy's direct impact on obesity is likely small, but posting is low-cost and may yield larger long-run benefits by spurring menu innovation toward lower-calorie items and increasing public nutrition awareness. Effects vary by consumer type and market context (e.g., near competitors), and results may differ for voluntary vs. mandatory posting.

- Quote snippet:

  "We find that mandatory calorie posting causes average calories per transaction to fall by 6 percent at Starbucks."

- Citation anchors:

- page=34, words=0-349, section=IV. Discussion, note=6% decrease in calories per transaction; supports the main quantified result and policy relevance.
- page=24, words=275-564, section=IV. Discussion, note=No significant effect on Starbucks profit on average.
- page=30, words=0-349, section=IV. Discussion, note=Evidence on learning and salience as mechanisms.
- page=34, words=0-349, section=IV. Discussion, note=Long-run benefits possible; costs are very low; potential for public education impact.
- page=34, words=0-349, section=IV. Discussion, note=Context: policy change to require calorie posting on menus across chain restaurants.

A10: What is the welfare interpretation (if any) of the results?

- Category: `A) Research question / contribution`
- Confidence: `low (0.31129440462685165)`
- Retrieval method: `local`
- Evidence type: `policy/welfare interpretation based on results`
- Data source: Starbucks transaction and cardholder data
- Answer:

The welfare interpretation of the results is that calorie posting may generate modest welfare gains through reductions in calories consumed, especially among high-calorie purchasers and certain demographics, but the overall welfare impact is uncertain because the study does not show obesity outcomes, finds limited effects on overall visit frequency, and raises the possibility of offsetting behavior outside posting locations. The policy relevance depends on whether calories reductions translate into meaningful obesity reductions; a crude calculation suggests potential non-trivial effects if baseline obesity relates to restaurant calories, but evidence is inconclusive. In short, some welfare gains are plausible, but robust conclusions require obesity data and cross-chain effects.

- Quote snippet:

  Could a 6 percent decrease in average calories per transaction at Starbucks conceivably translate into a non-trivial reduction in obesity?

- Citation anchors:
  - page=24, words=275-320, section=IV. Discussion, note=Summary of calorie reductions and lack of beverage impact
  - page=29, words=275-320, section=IV. Discussion, note=Obesity relevance and crude calculation about welfare impact of reductions

A11: What are the main limitations acknowledged by the authors?

- Category: `A) Research question / contribution`
- Confidence: `low (0.27605277059638345)`
- Retrieval method: `local`
- Evidence type: `limitations`
- Data source: Starbucks transaction data only
- Answer:

Two important limitations: (1) they do not directly measure the effect of calorie posting on obesity itself, as BMI data are not yet available; (2) they only have data for one chain (Starbucks), so they cannot know if effects would be similar at other chains or if consumers offset changes elsewhere.

- Quote snippet:

  There are two important limitations to this research. First, we do not directly measure the effect of calorie posting on obesity itself. A second limitation is that we have data for only one chain (Starbucks).

- Citation anchors:
  - page=3, words=275-564, section=Limitations, note=Two important limitations discussed in the limitations paragraph.

A12: What does the paper claim is novel about its data or identification?

- Category: `A) Research question / contribution`
- Confidence: `low (0.3098530512485742)`
- Retrieval method: `local`
- Evidence type: `data/identification novelty claim`
- Data source: Starbucks transaction data and anonymous cardholder data
- Answer:

The paper claims its data and identification are novel because it uses uniquely detailed transaction data across all NYC Starbucks stores and a large anonymous cardholder dataset, which together allow isolating calorie-posting effects on the intensive versus extensive margins and enable within-individual analysis of behavior changes.

- Quote snippet:

  "The detailed transaction data we use in this study are uniquely well-suited to answering this question."

- Citation anchors:
  - page=3, words=275-564, section=Data and identification novelty, note=The text states that 'The detailed transaction data we use in this study are uniquely well-suited to answering this question.'

- page=5, words=0-349, section=Data and identification novelty, note=The text describes the cardholder data as providing the ability to isolate effects on intensive and extensive margins from changes in consumer composition.

B01: What is the identification strategy (in one sentence)?

- Category: B) Identification strategy / causal design
- Confidence: low (0.27854413624147256)
- Retrieval method: local
- Evidence type: identification_strategy
- Data source: Starbucks transaction data and anonymous cardholder data
- Answer:

We identify the causal effect using a difference-in-differences design that exploits within-city variation over time and compares NYC to control cities (Boston and Philadelphia) around the calorie-posting date, with fixed effects to control for unobservables.

- Quote snippet:

  the policy variation we rely on for identification is at the city-week level.

- Citation anchors:
  - page=12, words=0-349, section=Identification strategy / Methodology, note=Policy variation for identification is at the city-week level; within-city variation used.

B02: Is the design experimental, quasi-experimental, or observational?

- Category: B) Identification strategy / causal design
- Confidence: low (0.30519829538385773)
- Retrieval method: local
- Evidence type: Describes policy implementation (calorie posting in NYC) and uses pre/post data with control cities and regression specification to compare outcomes, indicative of a quasi-experimental approach.
- Data source: Transaction data and cardholder data from Starbucks across NYC, Boston, and Philadelphia (with Seattle controls in related analysis).
- Table/Figure: Table 2
- Answer:

Quasi-experimental (natural experiment) design using difference-in-differences with NYC as the treatment group and control cities (Boston/Philadelphia, with Seattle as an additional control in related analysis).

- Quote snippet:

y_sct is a measure of calories per transaction at store s in city c on day t, POS_Tc t is a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008), and x_sct includes week fixed effects...

- Citation anchors:
  - page=page 7, words=275-566, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=In anticipation of the law change, the authors discuss NYC calorie posting and the use of Seattle and control cities as comparisons, indicating a quasi-experimental design with treatment and controls.
  - page=page 11, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Presents the regression specification including a POST_Tc t dummy for posting, enabling a pre/post comparison consistent with a difference-in-differences approach.

B03: What is the source of exogenous variation used for identification?

- Category: B) Identification strategy / causal design
- Confidence: low (0.2984012985569916)
- Retrieval method: local
- Evidence type: policy-change-based difference-in-differences with city-week level variation
- Data source: Transaction data and cardholder data from Starbucks stores in NYC, Boston, Philadelphia (and Seattle for robustness).
- Answer:

The exogenous variation comes from the NYC calorie-posting policy implemented on April 1, 2008, exploited through within-city (city-week) variation and a difference-in-differences design comparing NYC to control cities (Boston/Philadelphia) (and Seattle robustness).

- Quote snippet:

  the policy variation we rely on for identification is at the city-week level.

- Citation anchors:
  - page=11, words=275-522, section=Identification, note=within-city variation over time used to identify effects
  - page=12, words=0-349, section=Identification, note=policy variation at the city-week level used for identification

B04: What is the treatment definition and timing?

- Category: B) Identification strategy / causal design
- Confidence: low (0.13182204304609996)

- Retrieval method: `local`
- Evidence type: `Definition of treatment variable and timing from methods section`
- Answer:

The treatment is the NYC stores after calorie posting began, defined by a POST indicator that equals 1 for NYC stores on/after April 1, 2008 (pre-post comparison uses Boston/Philadelphia as controls).

- Quote snippet:

  POST Tc t is a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008)

- Citation anchors:
  – page=11, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption > A. Calories Per Transaction, note=Defines POST as a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008)

B05: What is the control/comparison group definition?

- Category: B) `Identification strategy / causal design`
- Confidence: `low (0.1799154587599786)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Transaction data and Cardholder data (Starbucks dataset)
- Answer:

The control/comparison group consists of Boston and Philadelphia (the control cities for NYC) used to compare against NYC; in robustness checks, Seattle's controls are Portland and San Francisco.

- Quote snippet:

  Qualitatively, however, it appears that Boston and Philadelphia are reasonable controls for NYC.

- Citation anchors:
  – page=6, words=0-349, section=Table 1: Summary Statistics, note=Boston and Philadelphia listed as controls for NYC.
  – page=14, words=275-590, section=Robustness checks / Seattle controls, note=Seattle controls include Portland, Oregon and San Francisco.

B06: What is the estimating equation / baseline regression specification?

- Category: B) `Identification strategy / causal design`
- Confidence: `low (0.30867120445315527)`

15

- Retrieval method: `local`
- Evidence type: `text`
- Data source: Transaction data and cardholder data from Starbucks stores; NYC stores after April 1, 2008, with control cities.
- Answer:

Baseline estimating equation: y_sct = x_sct + POS_Tc t + _sct, where y_sct is calories per transaction for store s in city c on day t; POS_Tc t equals 1 if calories were posted (NYC stores after 1 April 2008); x_sct includes week fixed effects, day-of-week dummies, holiday dummies, temperature and temperature-squared, and precipitation and precipitation-squared. Store fixed effects are included for transaction data; individual fixed effects for cardholder data. Identification comes from within-city variation over time.

- Quote snippet:

  y_sct is a measure of calories per transaction at store s in city c on day t, POST_Tc t is a dummy equal to one if calories were posted, and x_sct includes week fixed effects (to control for seasonality), day-of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared.

- Citation anchors:
  - page=12, words=0-349, section=Estimating equation / baseline regression specification, note=Contains the regression specification (Equation 1) used in the analysis.

B07: What fixed effects are included (unit, time, two-way, higher dimensional)?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.25446196351693673)`
- Retrieval method: `local`
- Evidence type: `textual evidence from regression specification describing fixed effects used in the analyses`
- Data source: American Economic Journal: Economic Policy (Bollinger et al., 2011)
- Answer:

The regressions include unit (store or consumer) fixed effects and time fixed effects: store fixed effects for transaction data; individual fixed effects for cardholder data; plus week fixed effects, day-of-week dummies, holiday dummies, and weather controls (temperature, precipitation, and their squares). In robustness checks, date fixed effects are used; when date fixed effects are included, week fixed effects and day-of-week dummies may be dropped.

- Quote snippet:

  x_sct includes week fixed effects (to control for seasonality), day-

16

of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared.

- Citation anchors:
  - page=11, words=275-522, section=Regression specification / fixed effects, note=Specifies that x_sct includes week fixed effects, day-of-week dummies, holiday dummies, and weather controls.
  - page=11, words=275-522, section=Fixed effects: unit vs. cardholder, note=Store fixed effects used with transaction data; individual consumer fixed effects used with cardholder data (store fixed effects dropped).
  - page=29, words=0-349, section=Cardholder regression fixed effects, note=The regressions based on the cardholder data include individual, week, and day-of-week fixed effects, and weather controls.

B08: What standard errors are used (robust, clustered; at what level)?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2808421540637513)`
- Retrieval method: `local`
- Evidence type: `Table 9 notes explicitly state 'Robust standard errors in parentheses' and 'An observation is a transaction.'`
- Data source: cardholder data (Table 9 notes)
- Answer:

Robust standard errors (reported in parentheses); the unit of observation is a transaction.

- Quote snippet:

Robust standard errors in parentheses.

- Citation anchors:
  - page=29, words=Notes:-Robust standard errors in parentheses., section=Table 9 notes, note=Robust standard errors in parentheses; unit of observation is a transaction.

B09: What is the key identifying assumption (parallel trends, exclusion restriction, ignorability)?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2727193222361167)`
- Retrieval method: `local`
- Evidence type: `parallel trends assumption used for difference-in-differences identification`
- Data source: Starbucks transaction and cardholder datasets (NYC vs control cities)
- Table/Figure: Figure 2

17

- Assumption flag: `True`
- Assumption notes: The key identifying assumption is that, absent calorie posting, NYC trends would have followed the same path as the control cities (parallel trends). The text explicitly notes no evidence of pre-trend differences.
- Answer:

Parallel trends (no evidence of pre-trend differences between NYC and the control cities).

- Quote snippet:

  First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia.

- Citation anchors:
  - page=page 14, words=0-349, section=Pre-trend check / Figure 2 discussion, note=States there is no evidence of pre-trend differences between NYC and Boston/Philadelphia, supporting parallel trends.

B10: What evidence is provided to support the identifying assumption?

- Category: B) Identification strategy / causal design
- Confidence: `low (0.24037968296379172)`
- Retrieval method: `local`
- Evidence type: `parallel trends / pre-trend differences check`
- Data source: Transaction data and cardholder data from Starbucks study
- Table/Figure: Figure 2
- Assumption flag: `True`
- Assumption notes: Supports parallel-trends identification; pre-treatment trends are not different across NYC and control cities, bolstering causal interpretation.
- Answer:

The paper provides evidence of no pre-trend differences between NYC and control cities (Boston/Philadelphia) prior to calorie posting, as shown for both the transaction data and the cardholder data (Figure 2), supporting the identifying assumption of parallel trends.

- Quote snippet:

  First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia.

- Citation anchors:
  - page=14, words=First,-differences., section=Identification/Pre-trend check, note=No pre-treatment trend differences between NYC and controls reported for both datasets.

B11: Are there event-study or pre-trend tests? What do they show?

- Category: B) Identification strategy / causal design
- Confidence: low (0.3177639965137011)
- Retrieval method: local
- Evidence type: Event-study / pre-trend test results described in the paper.
- Data source: Starbucks transaction and cardholder data used in Bollinger et al. (2011) Calorie Posting study.
- Table/Figure: Figure 2
- Assumption flag: True
- Assumption notes: Parallel trends prior to April 1, 2008 are assumed to hold, as evidenced by the lack of pre-trend differences between NYC and control cities in the event-study analysis.
- Answer:

Yes. The study conducts an event-study/pre-trend test by comparing NYC to control cities (Boston and Philadelphia) using weekly fixed effects and a post-posting indicator. It reports no evidence of pre-trend differences between NYC and the controls, and finds the calorie-posting effect emerges around the posting date and persists over time.

- Quote snippet:

  First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia.

- Citation anchors:
  - page=13, words=275-528, section=Event-study and pre-trend results, note=No evidence of pre-trend differences between NYC and Boston/Philadelphia; Figure 2 shows the results.

B12: What instruments are used (if IV)? Define instrument and first stage.

- Category: B) Identification strategy / causal design
- Confidence: low (0.09978406030178358)
- Retrieval method: local
- Evidence type: no_instrument_used
- Data source: Transaction data and cardholder data from the Starbucks calorie-posting study
- Table/Figure: Table 3
- Answer:

No instrumental variables are used. The study uses a difference-in-differences-like approach with a post-treatment indicator for calorie posting (POST) and store/city fixed effects, not an instrument. Therefore there is no first-stage as in IV estimation.

- Quote snippet:

  The POST variable is simply a dummy equal to one at NYC stores on every day after April 1, 2008.

- Citation anchors:
  – page=page 13, words=275-528, section=Methods: Calorie Posting, note=Describes calorie posting as binary with POST = one at NYC stores after Apr 1, 2008.

B13: What is the first-stage strength (F-stat, partial R^2, relevance evidence)?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2451960650751884)`
- Retrieval method: `local`
- Evidence type: `No instrumental-variables first-stage statistics reported; methodology described as regression with a post dummy and fixed effects.`
- Answer:

Not reported; the article uses regression with a post dummy and fixed effects, not an instrumental-variable first stage.

- Quote snippet:

  To examine the effects of calorie posting while controlling for seasonality, we estimate regressions: ysct = xsct + POST t + sct, where ysct is calories per transaction.

- Citation anchors:
  – page=11, words=275-522, section=Regression specification, note=Shows regression form with POST_t and fixed effects; no IV first stage mentioned.

B14: If RDD: what is the running variable and cutoff? bandwidth choice?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.22357778569302011)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Starbucks transaction and cardholder data for NYC, Boston, and Philadelphia
- Answer:

The study does not implement a regression discontinuity design. It uses a difference-in-differences framework around the calorie posting date, with a post-treatment indicator (post-01April08) for NYC stores after April 1, 2008. There is no running variable or RD cutoff or bandwidth explicitly specified in the text.

- Quote snippet:

  POS Tc t is a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008)

- Citation anchors:
  - page=11, words=0-349, section=Regression specification, note=Definition of POS Tc t as a dummy equal to one if calories were posted (NYC stores after April 1, 2008)
  - page=29, words=275-442, section=RD-like specification, note=Discussion of post-01April08 dummy with interactions; use of post-treatment indicator
  - page=5, words=0-100, section=Data context, note=Calorie posting commenced date and NYC/Boston/Philadelphia data

B15: If DiD: what is the timing variation (staggered adoption)? estimator used?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2889714459786146)`
- Retrieval method: `local`
- Evidence type: `methodology description from article`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia)
- Table/Figure: Figure 2
- Assumption flag: `True`
- Assumption notes: Parallel trends between NYC and control cities prior to April 1, 2008; no pre-trend differences observed according to the study.
- Answer:

A standard difference-in-differences (DiD) estimator is used with a post-treatment indicator for NYC after April 1, 2008, and city- and calendar-week fixed effects; an alternative specification uses separate weekly dummies for NYC and the control cities (Boston/Philadelphia) and excludes the POST variable to trace timing (i.e., a staggered timing approach).

- Quote snippet:

  The POST variable is simply a dummy equal to one at NYC stores on every day after April 1, 2008.

- Citation anchors:
  - page=11, words=275-522, section=Methods; Identification, note=Baseline DiD specification with POST dummy for NYC after April 1, 2008 and calendar controls.
  - page=13, words=275-528, section=Robustness/Alternative specification, note=Alternative DiD formulation with separate week dummies for NYC and controls and exclusion of the POST variable to trace timing.

21

C01: What dataset(s) are used? (name sources explicitly)

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.28017797555208224)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction data; anonymous Starbucks cardholder data (NYC, Boston, Philadelphia).
- Answer:

Transaction data and cardholder data (anonymous Starbucks cardholder purchases).

- Quote snippet:

  We refer to the first dataset as the transaction data and the second dataset as the cardholder data.

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=Describes the two datasets: the transaction data and the cardholder data.
  - page=6, words=0-349, section=Table 1—Summary Statistics, note=Continuation of dataset description and context for comparison between datasets.

C02: What is the unit of observation (individual, household, firm, county, transaction, product)?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.32252825729598444)`
- Retrieval method: `local`
- Evidence type: `Data description from the paper showing the unit of observation is transaction data (with parallel cardholder data for individual-level analysis).`
- Data source: Starbucks transaction data and cardholder data
- Assumption flag: `False`
- Answer:

transaction

- Quote snippet:

  For each transaction we observe the time and date, store location, items purchased, and price of each item.

- Citation anchors:
  - page=5, words=0-349, section=Data Summary, note=Unit of observation is the transaction data; for each transaction we observe time/date, store location, items purchased, and price.

C03: What is the sample period and geographic coverage?

- Category: C) Data, sample, and measurement
- Confidence: low (0.23064608598535105)
- Retrieval method: local
- Evidence type: data_description
- Data source: Starbucks transaction data for NYC, Boston, and Philadelphia
- Answer:

Sample period: January 1, 2008–February 28, 2009; geographic coverage: all Starbucks locations in New York City (NYC) and all locations in Boston and Philadelphia (i.e., NYC, Boston, Philadelphia).

- Quote snippet:

  Our transaction data cover all 222 Starbucks locations in NYC, and all 94 Starbucks locations in Boston and Philadelphia.

- Citation anchors:
  - page=5, words=0-349, section=Data Summary, note=Describes geographic coverage and the pre/post period for calorie posting data.

C04: What are the sample restrictions / inclusion criteria?

- Category: C) Data, sample, and measurement
- Confidence: low (0.24504002347179643)
- Retrieval method: local
- Evidence type: Data description / sample restrictions
- Data source: Transaction data for all Starbucks locations in NYC, Boston, and Philadelphia; anonymous cardholder data
- Answer:

Sample restrictions: The cardholder analysis uses a subsample of anonymous cardholders who averaged at least one transaction per week in NYC, Boston, or Philadelphia in the pre-posting period, and the transactions are restricted to stores open for the entire data period; transactions with more than four units of a single item are excluded (balanced panel).

- Quote snippet:

  We define a subsample containing any individual that averaged at least one transaction per week in one of NYC, Boston, or Philadelphia, in the period before calorie posting in NYC.

- Citation anchors:
  - page=5, words=0-349, section=Data Summary / Cardholder sample, note=Definition of cardholder subsample: at least one weekly transaction in NYC/Boston/Philadelphia before posting.

C05: What is the sample size (N) in the main analysis?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.29772083628688245)`
- Retrieval method: `local`
- Evidence type: `Figure 4 caption note`
- Data source: American Economic Journal: Economic Policy, Bollinger et al., calorie posting in chain restaurants
- Table/Figure: Figure 4
- Answer:

N 50

- Quote snippet:

  Note: To preserve confidentiality of the data, we are unable to specify the exact value of N, only that N 50.

- Citation anchors:

C06: How is the key outcome measured? Any transformations (logs, z-scores, indices)?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.32671005493514993)`
- Retrieval method: `local`
- Evidence type: `outcome definition and transformations description`
- Data source: Transaction data and cardholder data
- Table/Figure: Table 3; Table 4; Table 6
- Answer:

The key outcome is calories per transaction (y_sct). It is measured as calories per transaction at store s in city c on day t. In some specifications this outcome is used in raw form, while in others it is transformed to a log: log(beverage + food calories). Additionally, the analysis includes a quantile approach where the log of the nth quantile of calories per transaction is regressed.

- Quote snippet:

  y_sct is a measure of calories per transaction at store s in city c on day t

- Citation anchors:

– page=page 11, words=275-522, section=Regression specification, note=Definition of y_sct as calories per transaction
– page=page 12, words=275-624, section=Transformations, note=Use of log(beverage + food calories) as dependent variable in some specs
– page=page 21, words=275-421, section=Quantile analysis, note=Discussion of estimating effects at different calories-per-transaction quantiles
– page=page 22, words=0-349, section=Table 6, note=Quantile-based results table with calorie-posting effects

C07: How is treatment/exposure measured? Any constructed variables?

- Category: C) Data, sample, and measurement
- Confidence: low (0.2923503571693718)
- Retrieval method: local
- Evidence type: descriptive methodology explanation from the study
- Data source: Transaction dataset and cardholder dataset from the study (NYC stores during calorie posting; aggregation to store-day level and individual fixed effects where applicable).
- Assumption flag: True
- Assumption notes: Identification relies on within-city variation over time with store/week fixed effects, assuming no time-varying confounders that shift only NYC relative to control cities outside the posting period.
- Answer:

Exposure (treatment) is measured with a binary post-treatment indicator called POST, defined as 1 for NYC stores on every day after April 1, 2008 when calories were posted. Identification relies on within-city variation over time, using store fixed effects and week fixed effects (policy variation at the city-week level). In the transaction data, exposure is observed at the store-day level; in the cardholder data, fixed effects at the individual level are used. Constructed variables include POST (the exposure), various outcome transformations (e.g., log calories per transaction, log beverage calories per transaction, log food calories per transaction, log beverage + food calories), numbers of beverages and food items per transaction, calories per item purchased, and, in a robustness check, the nth quantile of calories per transaction at each store-day; the commuting analysis also uses a log-calories-per-transaction with interaction terms.

- Quote snippet:

  POST variable is simply a dummy equal to one at NYC stores on every day after April 1, 2008.

- Citation anchors:
    – page=11, words=0-80, section=Exposure definition and model, note=Defines POST as a dummy equal to one if calories were posted

25

(NYC stores after April 1, 2008).

- page=12, words=0-120, section=Identification strategy, note=Notes store fixed effects and that policy variation for identification is at the city-week level.
- page=5, words=275-400, section=Data context, note=Describes datasets (transaction vs cardholder) and overall exposure to calorie posting.
- page=16, words=275-420, section=Constructed outcomes, note=Lists constructed variables: number of items per transaction and log(calories per item purchased) for beverages and food.
- page=22, words=0-120, section=Quantile analysis, note=Describes constructing the nth quantile of calories per transaction as the dependent variable.
- page=29, words=0-120, section=Commuter analysis, note=Mentions dependent variable as log(calories + 1) in the commuter/ NYC vs non-NYC analysis.

C08: Are there key covariates/controls? Which ones are always included?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2871307076656878)`
- Retrieval method: `local`
- Evidence type: `textual evidence from regression specification`
- Data source: Starbucks transaction data and cardholder data used in the calorie posting analysis
- Answer:

Yes. The key covariates/controls include week fixed effects (seasonality), day-of-week dummies, holiday dummies, and weather controls (temperature and precipitation with their squared terms). In the cardholder analyses, individual fixed effects are also included.

- Quote snippet:

  x_sct includes week fixed effects (to control for seasonality), day-of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared.

- Citation anchors:
    - page=11, words=x_sct-squared., section=Regression specification, note=Explicit list of covariates/controls included in x_sct
    - page=29, words=The regressions are-weather controls., section=Cardholder data regression specification, note=Indicates inclusion of individual fixed effects and weather controls in cardholder analysis

C09: How are missing data handled (dropping, imputation, weighting)?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2779073847940623)`
- Retrieval method: `local`
- Evidence type: `data_cleaning`
- Data source: Starbucks transaction data and cardholder data
- Answer:

Missing data are handled by dropping observations rather than imputing. Specifically, transactions at stores not open during the entire data period are excluded to form a balanced panel, and transactions that included more than four units of any one item are also excluded. There is no discussion of imputation or weighting.

- Quote snippet:

  we exclude transactions at stores that were not open during the entire data period (i.e., we analyze the balanced panel)

- Citation anchors:
  - page=6, words=0-349, section=Data cleaning / Panel construction, note=Exclude stores not open during the entire data period (balanced panel) and exclude transactions with >4 units.
  - page=6, words=0-349, section=Data cleaning / Panel construction, note=The excluded transactions represent only 2.2 percent of all transactions.
  - page=5, words=0-349, section=Data summary / Confidentiality, note=To preserve confidentiality, values are normalized (NYC = 1).

C10: Are weights used (survey weights, propensity weights)? How?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2953880045202148)`
- Retrieval method: `local`
- Evidence type: `textual evidence from regression methodology describing fixed effects and weather controls; no weights mentioned.`
- Data source: Transaction data and anonymous cardholder data
- Answer:

No weights are described or used. The analysis uses regression specifications with fixed effects (week, day-of-week, holidays) and weather controls to account for seasonality, rather than survey or propensity weights.

- Quote snippet:

  To examine the effects of calorie posting while controlling for seasonality and other influences, we estimate regressions of the following

form: (1) y_sct = x_sct + POST Tc t + _sct, where y_sct is
a measure of calories per transaction

- Citation anchors:
    - page=5, words=0-349, section=Data Summary, note=Data description; no mention of survey/propensity weights.
    - page=5, words=275-591, section=Data Summary, note=Continuation of data description; no weights described.
    - page=11, words=0-349, section=Methods/Regression specification, note=Regression specification with y_sct, x_sct, POST, and fixed effects; weather controls mentioned.
    - page=11, words=275-522, section=Methods/Regression specification, note=Details of fixed effects and weather controls; no weighting discussed.
    - page=12, words=275-624, section=Methods/Results, note=Continuation of model controls; no weights described.

C11: Are data linked/merged across sources? How is linkage performed?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.22242793049253753)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction data; anonymous Starbucks cardholder data.
- Answer:

Yes, the study uses two datasets—the transaction data and the anonymous cardholder data—and analyzes them to study effects, implying linkage at the cardholder level, though the exact linkage method is not described in the provided text.

- Quote snippet:

  We refer to the first dataset as the transaction data and the second dataset as the cardholder data.

- Citation anchors:
    - page=5, words=275-320, section=Data description — datasets, note=Definition of transaction data vs cardholder data; implies linkage across datasets.

C12: What summary statistics are reported for main variables?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3270012043096938)`
- Retrieval method: `local`
- Evidence type: `Table`

- Data source: Table 1 in the paper; Summary statistics for Transaction Data and Cardholder Data (Prior to policy change)
- Table/Figure: Table 1
- Answer:

The paper reports Table 1: Summary Statistics for Transaction Data and Cardholder Data (Prior to policy change). It lists the main variables and their statistics, including: avg weekly transactions per store; avg weekly revenue per store; percent transactions with brewed coffee, beverage, and food; avg. items per transaction; avg. drink items per transaction; avg. food items per transaction; food attach rate; avg. dollars per transaction; avg. calories per transaction; avg. drink calories per transaction; avg. food calories per transaction. The data are normalized (1.00) for confidentiality and reflect data prior to calorie posting in NYC.

- Quote snippet:

  Table 1—Summary Statistics for Transaction Data and Cardholder Data (Prior to policy change) Transaction data Cardholder data Boston & Boston & New York City Philadelphia New York City Philadelphia Avg. weekly transactions per store 1.00 0.77 1.00 1.90

- Citation anchors:
  - page=6, words=0-349, section=Table 1, note=Summary statistics for main variables (prior to calorie posting)

C13: Are there descriptive figures/maps that establish baseline patterns?

- Category: C) Data, sample, and measurement
- Confidence: low (0.28341958679236584)
- Retrieval method: local
- Evidence type: Descriptive figures (Figure 1 and Figure 2) illustrating baseline patterns and pre-trend comparability
- Data source: Starbucks transaction data and cardholder data (NYC, Boston, Philadelphia).
- Answer:

Yes. Descriptive figures establish baseline patterns: Figure 1 shows average calories per transaction per week, distinguishing NYC from control cities; Figure 2 shows results for each dataset and indicates no pre-trend differences between NYC and controls.

- Quote snippet:

  Figure 1 shows average calories per transaction each week, distinguishing transactions in NYC from transactions in the control cities.

- Citation anchors:

– page=9, words=0-349, section=II.A Calories Per Transaction, note=Figure 1 description: average calories per transaction by week, NYC vs controls.

– page=14, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Figure 2 and discussion of no pre-trend differences between NYC and controls.

D01: What is the headline main effect estimate (sign and magnitude)?

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.4279760780782623)`
- Retrieval method: `local`
- Evidence type: `Text from results section (Table 3) indicating the effect on log(beverage+food calories).`
- Answer:

-5.8% decrease in average calories per transaction (negative).

- Quote snippet:

  we estimate a 5.8 percent decrease in average calories per transaction, equivalent to 14.4 calories.

- Citation anchors:
  - page=12, words=0-349, section=Main results / Table 3, note=Headline main effect: 5.8% decrease in average calories per transaction.

D02: What is the preferred specification and why is it preferred?

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.18144481849324717)`
- Retrieval method: `local`
- Evidence type: `regression specification with fixed effects and weather controls`
- Data source: Transaction data and cardholder data from Starbucks stores (NYC, Boston, Philadelphia) prior to calorie posting
- Answer:

The preferred specification is the regression with controls: $y\_sct = x\_sct + \_POS$ Tc t $+ \_sct$, where y_sct is calories per transaction; x_sct includes week fixed effects, day-of-week, holiday dummies, temperature and its square, precipitation and its square; and POS_t indicates calorie posting. They estimate versions separately for transaction data and cardholder data to control for seasonality and other influences while isolating the effect of posting.

- Quote snippet:

  To examine the effects of calorie posting while controlling for seasonality and other influences, we estimate regressions of the following form: (1) $y\_sct = x\_sct + \_POS$ Tc t $+ \_sct$

- Citation anchors:
  - page=11, words=0-349, section=Regression specification, note=Introduces the regression form and the basic variables.
  - page=11, words=275-522, section=Regression specification, note=Details the controls included (week fixed effects, day-of-week, holidays, weather) and that the specification is estimated for both data types.

D03: How economically meaningful is the effect (percent change, elasticity, dollars)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: `medium (0.45884789727602737)`
- Retrieval method: `local`
- Evidence type: `Quantitative estimates from Tables 3 and 6; discussion in IV. (Calorie posting effects, margins, and profit implications).`
- Data source: Starbucks NYC transaction data and anonymous cardholder data
- Table/Figure: Table 3; Table 6; Figure 3; Figure 4
- Assumption flag: `True`
- Assumption notes: Crude extrapolation uses a chain-restaurant calorie share of about 25% and applies a 6% reduction across all chain-restaurant calories to assess population-level impact; this is explicitly described as a crude calculation in the Discussion.
- Answer:

The effect is quantitatively modest but nontrivial: on average, total calories per transaction fell about 5–6% (roughly 5.8% in transaction data and 5.0% in cardholder data). Breakouts show beverage calories largely unchanged (0% to -0.3% not significant) while food calories fell more steeply (roughly -13.7% in transaction data and -11.2% in cardholder data). The reductions are larger for higher-calorie purchases (top quantiles show reductions around 5–6% in calories per transaction). The analysis also finds that most of the calorie drop comes from the extensive margin (not buying food items) rather than substitution to lower-calorie items, and profits for Starbucks did not show a statistically significant decline (revenues rose modestly with higher item-level prices). A crude extrapolation is discussed: because roughly 25% of Americans' calories come from chain restaurants, a 6% reduction at all chain restaurants could have population-level implications, though this is framed as a crude calculation and its real obesity impact remains uncertain.

- Quote snippet:

  calorie posting caused a trivial decrease in beverage calories per transaction of 0.3 percent. In the second row, the dependent variable is log(food calories per transaction). Based on the transaction

31

data we estimate that food calories decreased by 13.7 percent

- Citation anchors:
  - page=12, words=0-349, section=Results: Calorie posting effects (Table 3), note=Beverage calories per transaction: -0.3%; Food calories per transaction: -13.7%; Total calories per transaction: -5.8% (transaction data).
  - page=12, words=275-624, section=Results: Cardholder data, note=Beverage calories not significantly different from zero; Food calories per transaction: -11.2%; Total calories: -5.0% (cardholder data).
  - page=16, words=0-349, section=B. Substitution Effects, note=Discussion of intensive vs extensive margins; estimation of items per transaction and calories per item purchased.
  - page=18, words=275-541, section=Subsection on margins and distribution, note=Describes main finding that about three-quarters of calorie reductions come from the extensive margin (not buying food), and one-quarter from the intensive margin.
  - page=24, words=275-564, section=IV. Discussion, note=Summary: 14% reduction in food calories on average; 6% average reduction in total calories; profits largely unchanged; discussion of population-level implications via a crude calculation.

D04: What are the key robustness checks and do results survive them?

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2836299975851264)`
- Retrieval method: `local`
- Evidence type: `robustness_checks`
- Answer:

Robustness checks show the main results hold under several alternative specifications. Adding date fixed effects (dropping weekly/daily fixed effects) leaves the calorie-posting coefficients largely unchanged (beverage calories about -0.004, food calories about -0.152, total calories about -0.063). Using absolute calories instead of log calories yields nearly identical results. A serial-correlation robustness check that aggregates pre/post periods also yields essentially the same effect size. The cardholder (within-individual) data generally corroborates a calories-per-transaction reduction, though beverage calories are not significantly different from zero and food calories show a notable decline, suggesting some composition effects. A pre-trend check shows no evidence of pre-trend differences between NYC and control cities. Overall, the core finding—calorie posting reduces calories per transaction—persists across these robustness tests.

- Quote snippet:

As a robustness check, we include date fixed effects in the above

specification (and therefore drop the day-of-week and week fixed effects). The estimates based on the transaction data are barely changed.

- Citation anchors:
  - page=12, words=0-349, section=Robustness checks, note=Date fixed effects robustness: coefficients barely changed (beverage -0.004, food -0.152, total -0.063).
  - page=12, words=0-349, section=Robustness checks, note=Absolute calories robustness: results almost identical.
  - page=12, words=0-349, section=Robustness checks, note=Serial-correlation robustness: aggregation pre/post yields approximately the same effect.
  - page=14, words=0-349, section=Robustness checks, note=No evidence of pre-trend differences between NYC and control cities.
  - page=12, words=275-624, section=Robustness checks, note=Cardholder-data robustness: isolates intensive-margin; beverage calories not significant; total calories reduced.

D05: What placebo tests are run and what do they show?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.2121744163858382)
- Retrieval method: local
- Evidence type: robustness/placebo test (pre-trend check)
- Data source: Starbucks transaction data and cardholder data
- Table/Figure: Figure 2
- Assumption flag: True
- Assumption notes: Assumes parallel/trend equality between NYC and control cities prior to calorie posting; supported by lack of pre-trend differences in the placebo tests.
- Answer:

The authors run placebo tests by checking for pre-trend differences between NYC stores and the control cities (Boston/Philadelphia) before calorie posting, using two data sources (transaction data and cardholder data). They find no evidence of pre-trend differences, i.e., parallel trends prior to posting, and the observed decline in calories per transaction occurs only after posting (around April 1, 2008) and persists thereafter. This supports the interpretation that calorie posting caused the changes rather than pre-existing trends.

- Quote snippet:

  First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia.

- Citation anchors:
  - page=11, words=First,-differences, section=Robustness/Placebo

33

test, note=No pre-trend differences between NYC and control cities cited in the cardholder dataset.
- page=14, words=First,-differences, section=Robustness/Placebo test, note=No pre-trend differences; drop in calories per transaction occurs after posting date.

D06: What falsification outcomes are tested (unaffected outcomes)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: `low (0.2569710569233812)`
- Retrieval method: `local`
- Evidence type: `Falsification/robustness test description`
- Data source: Starbucks transaction and cardholder datasets
- Answer:

Weather variables (temperature, precipitation, and their squares) were tested as potential unaffected outcomes and found insignificant; robustness checks with date fixed effects also leave results unchanged.

- Quote snippet:

  We actually find that the weather variables have an insignificant impact on sales.

- Citation anchors:
  - page=page 11, words=0-349, section=Methods/Empirical model, note=Weather controls included; weather variables found insignificant, illustrating an unaffected outcome test
  - page=page 12, words=0-349, section=Robustness checks, note=Date fixed effects robustness; estimates barely changed

D07: What heterogeneity results are reported (by income, size, baseline exposure, region)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: `medium (0.37474580457139817)`
- Retrieval method: `local`
- Evidence type: `Heterogeneity analyses from Table 5 (income/education/baseline exposure interactions) and accompanying regional analyses; Table 2 (size-based beverage choices) and Table 9 (regional commuter effects).`
- Data source: American Economic Journal: Economic Policy (Bollinger et al., 2011)
- Answer:

Heterogeneity results show: (i) by income and education, higher-income ZIPs and areas with more college graduates experience larger decreases in calories per transaction; (ii) by size, changes in beverage size accompany calorie shifts (e.g., some customers switch to smaller sizes with lower calories and others to larger

sizes with higher calories); (iii) by baseline exposure, high-calorie and high-frequency customers exhibit larger reductions in calories per transaction than lower-calorie or lower-frequency customers; (iv) by region, the calorie-posting effects are concentrated in NYC relative to control cities, with no large change in visit frequency overall and commuters showing somewhat larger reductions than non-commuters in NYC.

- Quote snippet:

  We find that the decrease in calories per transaction was larger in zips with higher income and in zips with more education.

- Citation anchors:
  - page=21, words=Table 5—Heterogeneity-high calorie customer, section=Heterogeneity, note=Shows interaction terms for income, education, age, gender, and baseline exposure in the transaction data; highlights larger effects for higher income and higher education, and for high-calorie customers.
  - page=21, words=Posting × median income-−0.012**, section=Heterogeneity, note=Income interaction in the transaction data indicating larger reductions in calories per transaction in higher-income areas.
  - page=21, words=Posting × percent with college degree-−0.020**, section=Heterogeneity, note=Education interaction indicating larger effects where more residents have college degrees.
  - page=21, words=Posting × high calorie customer-−0.444***, section=Heterogeneity, note=Baseline exposure interaction showing greater reductions for high-calorie customers.
  - page=11, words=Table 2—Changes in Cardholders' Beverage Choices-Philadelphia, section=Beverage size heterogeneity, note=Demonstrates changes across smaller, same, and larger beverage sizes and associated calorie content in NYC vs control cities.
  - page=20, words=We found no statistically significant change-control cities, section=Regional results, note=Region-based comparison showing no large change in purchase frequency in NYC relative to controls.
  - page=20, words=We found no statistically significant change in the frequency of cardholders' purchases in NYC relative to the control cities; indeed- NYC, section=Regional results, note=Regional heterogeneity note supporting limited effect on visit frequency.
  - page=21, words=for those who averaged more than 250 calories per transaction-26 percent, section=Baseline exposure, note=High-calorie baseline exposure shows particularly large reductions (26% in post-period).
  - page=22, words=Table 6—Estimated Effects-quantile, section=Baseline exposure (quantile analysis), note=Higher-calorie

35

quantiles show larger absolute reductions; percent change roughly stable across higher quantiles.

D08: What mechanism tests are performed and what do they imply?

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.14271505518105457)`
- Retrieval method: `local`
- Evidence type: `Mechanism tests distinguishing salience vs learning via cross-city commuter analysis, regression with posting indicator and fixed effects, margin analyses (extensive vs intensive), quantile analysis, and survey data.`
- Data source: Starbucks transaction data; cardholder-level data; city controls (NYC, Boston, Philadelphia, Seattle, etc.).
- Answer:

They conduct mechanism tests to distinguish salience from learning: (1) they examine whether calorie postings in NYC influence purchases outside NYC (i.e., among commuters) to infer learning rather than salience, (2) they use regression analyses with post indicators and fixed effects across cities to isolate the posting effect, (3) they analyze substitution versus extensive vs intensive margins to see how calories per transaction change (e.g., fewer items vs lower-calorie choices), (4) they perform a quantile analysis to assess heterogeneity across the calories-per-transaction distribution, and (5) they use survey data on calorie knowledge to contextualize the information effect. Overall, the tests imply a learning mechanism (calorie information affects choices beyond posting sites), with reductions driven largely by the extensive margin (fewer high-calorie items) and some substitution to lower-calorie options, and no change in purchase frequency.

- Quote snippet:

  exposure to calorie information affects their choices even at nonposting (i.e., non-NYC) stores, which is consistent with a learning effect but inconsistent with the salience effect.

- Citation anchors:
  - page=3, words=commuters-effect, section=Mechanism tests (salience vs learning), note=Commuter evidence supports learning over salience.
  - page=7, words=In anticipation-data for Seattle, section=Research design / Data, note=Cross-city design including Seattle and control cities to identify effects.
  - page=11, words=To-influences, section=Methods: Regression specification, note=Post-date regression framework controlling for seasonality and weather.
  - page=16, words=B. Substitution Effects-items., section=Substitution / Margins, note=Assesses extensive vs intensive margins via item-

level regressions.

- page=21, words=That-))., section=Quantile analysis, note=Examines effects across quantiles of calories per transaction.
- page=24, words=IV.-percent., section=Discussion / Findings, note=Food calories per transaction fall; informs mechanism magnitude.
- page=25, words=The-trends., section=Appendix: Survey design, note=Survey data context for information effects and knowledge.
- page=29, words=The-percent., section=Commuter analysis (Table 9), note=Commuters' calorie reductions in NYC and outside NYC; robustness check.

D09: How sensitive are results to alternative samples/bandwidths/controls?

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.359343916221238)`
- Retrieval method: `local`
- Evidence type: `robustness checks and sensitivity analyses`
- Data source: Starbucks transaction data and anonymous cardholder data (NYC, Boston, Philadelphia)
- Assumption flag: `True`
- Assumption notes: Assumes Boston and Philadelphia provide valid counterfactuals for NYC and that no other city-specific shocks drive the results; robustness relies on week fixed effects to control for seasonality and on within-individual variation (cardholder data).
- Answer:

The results are robust to alternative samples, bandwidths, and controls. The drop in calories per transaction after calorie posting persists across transaction and cardholder samples, shows no pre-trend differences, and remains when using alternative specifications (including absolute calories) or a conservative serial-correlation approach; commuter and non-commuter subsamples show similar patterns.

- Quote snippet:

  There are a few points of interest to note from Figure 2. First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia.

- Citation anchors:
  - page=14, words=0-349, section=Pre-trend robustness, note=Figure 2 shows no evidence of pre-trend differences between NYC and control cities.
  - page=12, words=0-349, section=Cardholder data robustness, note=The cardholder data isolates intensive-margin effects by tracking same individuals over time.

- page=12, words=550-630, section=Alternative measures robustness, note=Using absolute calories yields similar results.
- page=16, words=0-349, section=Serial-correlation robustness, note=A conservative approach to serial correlation still yields roughly the same effect and significance.
- page=29, words=0-349, section=Subsamples robustness, note=Commuters and non-commuters show reductions similar to the NYC effect; some estimates are imprecise.
- page=5, words=0-349, section=Data overview robustness, note=Two datasets (transaction and cardholder) provide consistent results around the posting date.

D10: What are the main takeaways in the conclusion (bullet summary)?

- Category: `D) Results, magnitudes, heterogeneity, robustness`

- Confidence: `low (0.2697093448641606)`

- Retrieval method: `local`

- Evidence type: `conclusion_summary`

- Data source: Starbucks transaction data; cardholder data; surveys

- Assumption flag: `True`

- Assumption notes: Back-of-the-envelope calculation uses USDA data and census-based expenditure; assumes 25% of average American calories come from chain restaurants and a 6% reduction across all chain restaurants to estimate potential obesity impact.

- Answer:

- Mandatory calorie posting reduces average calories per transaction by about 6% at Starbucks and the effect is long-lasting.

- The reduction is driven mainly by changes in food calories, with beverage calories changing little.

- The effect is larger for high-calorie consumers; learning and salience both help explain the response.

- Consumers exposed to calorie information reduce calories even at non-posting stores; commuters show learning effects, and surveys indicate increased sensitivity to calories.

- The impact on Starbucks profits is negligible on average; in some near-competition areas, posting may increase revenue.

- Policy takeaway: mandatory posting has low costs and potential public education benefits; long-run benefits may grow if restaurants innovate with

low-calorie options, though results may differ for voluntary posting.

- A back-of-the-envelope calculation suggests the direct effect on obesity may be small, but long-run effects could be more substantial if the industry responds and consumer awareness grows.

- Quote snippet:

  We find that mandatory calorie posting causes average calories per transaction to fall by 6 percent at Starbucks.

- Citation anchors:
  - page=page 34, words=0-349, section=Conclusion, note=Main take-aways from the conclusion section.

E01: What are the most important prior papers cited and why are they central here?

- Category: E) Citations and related literature
- Confidence: low (0.273515049200881)
- Retrieval method: local
- Evidence type: literature review / background
- Answer:

The most important prior papers are those that establish how nutrition/calorie information at the point of purchase influences consumer choices and the mechanisms (information provision, learning vs. salience) that this study tests. Key works include Jacoby et al. (1977) on consumer use and understanding of nutrition information; Ippolito & Mathios (1990, 1995) on information, advertising, and health choices; Downs, Loewenstein, and Wisdom (2009) on strategies for promoting healthier choices; Bassett et al. (2008) and Elbel et al. (2009) on calorie information at fast-food chains and its effects; Currie et al. (2010) on the impact of fast-food restaurants on obesity; Variyam & Cawley (2006) on nutrition labels and obesity; and Jin & Leslie (2003, 2009) and Kiesel & Villas-Boas (2008) for labeling/information experiments in restaurants and retail. These papers are central because they (i) establish that information disclosure can shape choices, (ii) provide empirical estimates of labeling effects in real-world settings, and (iii) frame competing mechanisms (salience vs. learning) that the current study uses to interpret calorie posting effects.

- Quote snippet:

  The notion that increasing the provision of nutrition information may stimulate people to adopt healthier eating habits is an old idea, and numerous prior studies have sought to evaluate its merit.

- Citation anchors:
  - page=page 37, words=0-349, section=References, note=Jacoby et al. 1977 — early work on consumer use and comprehension of nutrition information

39

- page=page 37, words=0-349, section=References, note=Ippolito & Mathios (1990, 1995) — information, advertising and health choices in nutrition labeling
- page=page 37, words=0-349, section=References, note=Downs, Loewenstein, & Wisdom (2009) — strategies for promoting healthier food choices
- page=page 36, words=275-422, section=References, note=Currie et al. (2010) — effect of fast-food restaurants on obesity and weight gain
- page=page 36, words=275-422, section=References, note=Bassett et al. (2008) — purchasing behavior and calorie information at fast-food chains
- page=page 38, words=0-349, section=References, note=Elbel et al. (2009) — calorie labeling and food choices among low-income NYC residents
- page=page 37, words=275-624, section=References, note=Variyam & Cawley (2006) — nutrition labels and obesity (NBER Working Paper 11956)
- page=page 37, words=550-649, section=References, note=Jin & Leslie (2003/2009); Kiesel & Villas-Boas (2008) — information labeling experiments in restaurants/stores

E02: Which papers does this work most directly build on or extend?

- Category: `E) Citations and related literature`
- Confidence: `low (0.23853607389917472)`
- Retrieval method: `local`
- Evidence type: `citation evidence from literature review`
- Data source: NYC calorie posting literature (Bassett et al. 2008; Downs et al. 2009; Elbel et al. 2009; Currie et al. 2010)
- Answer:

It most directly builds on the NYC calorie-posting literature, notably Downs, Loewenstein, and Wisdom (2009) and Elbel et al. (2009); it also cites Bassett et al. (2008) as an earlier NYC study.

- Quote snippet:

  Two subsequent papers compare purchase data before and after calorie posting in NYC.

- Citation anchors:
  - page=8, words=275-554, section=Related Research / NYC calorie posting, note=Mentions two subsequent NYC pre/post calorie posting studies: Downs et al. 2009 and Elbel et al. 2009.

E03: Which papers are used as benchmarks or comparisons in the results?

- Category: `E) Citations and related literature`
- Confidence: `low (0.26014396709337884)`
- Retrieval method: `local`
- Evidence type: `Literature review / benchmark references in results`
- Answer:

The benchmark/comparison papers are Anderson and Matsa (2011) and Currie et al. (2010).

- Quote snippet:

  Two recent papers by Michael L. Anderson and David A. Matsa (2011) and Janet Currie et al. (2010).

- Citation anchors:
  - page=8, words=0-349, section=Background/Literature review, note=Mentions benchmark papers: Anderson & Matsa (2011) and Currie et al. (2010).

E04: What data sources or datasets are cited and how are they used?

- Category: `E) Citations and related literature`
- Confidence: `low (0.28904240295650285)`
- Retrieval method: `local`
- Evidence type: `Empirical data sources and supplementary data used in the analysis.`
- Data source: Starbucks transaction data; Starbucks cardholder data; Starbucks milk order data; in-store surveys; nutrition information (Starbucks website) for calorie calculations
- Table/Figure: Table 1
- Answer:

The study cites several data sources: (1) transaction data from Starbucks stores in NYC, Boston, and Philadelphia (222 NYC stores and 94 in Boston/Philadelphia) covering 3 months before to 11 months after calorie posting, with items, prices and calories; (2) anonymous Starbucks cardholder data comprising about 2.7 million individuals used to analyze how calorie information affects individual behavior and to separate intensive vs extensive margins from changes in consumer composition; (3) Starbucks milk order data across all stores in the three cities to assess milk-related calories; (4) in-store consumer surveys conducted in Seattle and San Francisco (792 completed surveys) plus an Appendix questionnaire to test knowledge and control for time trends.

- Quote snippet:

"Our transaction data cover all 222 Starbucks locations in NYC, and all 94 Starbucks locations in Boston and Philadelphia. At each location we observe all transactions... Using Starbucks nutritional information we can also calculate the calories in each purchase."

- Citation anchors:
  - page=5, words=0-349, section=Data sources: Transaction and Cardholder Data, note=Describes scope of transaction data (222 NYC + 94 Boston/Philadelphia stores), time window, and calories calculation; introduces cardholder data.
  - page=5, words=275-591, section=Cardholder data details, note=Cardholder dataset contains 2.7 million anonymous individuals; used to analyze individual-level responses and separate margins from composition.
  - page=6, words=0-349, section=Table 1 and data sources, note=Table 1 presents summary statistics for transaction and cardholder data; mentions milk order data as an additional data source.
  - page=25, words=275-599, section=Survey data in Seattle and San Francisco, note=Describes in-store surveys (792 completed) and the Appendix questionnaire; used to control time trends and test consumer knowledge.

E05: What methodological or econometric references are cited (e.g., DiD, IV, RDD methods)?

- Category: E) Citations and related literature
- Confidence: medium (0.38555247933843567)
- Retrieval method: local
- Evidence type: Econometric/methodological references (randomized field experiments, information labeling studies, salience-related methods)
- Answer:

Methodological/econometric references cited include: (i) Cai, Hongbin; Chen, Yuyu; Fang, Hanming (2009). Observational Learning: Evidence from a Randomized Natural Field Experiment (randomized field experiment); (ii) Chetty, Raj; Looney, Adam; Kroft, Kory (2009). Salience and Taxation: Theory and Evidence; (iii) Jin, Ginger Zhe; Leslie, Phillip (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards; (iv) Jin, Ginger Zhe; Leslie, Phillip (2009). Reputational Incentives for Restaurant Hygiene; (v) Downs, Julie S.; Loewenstein, George; Wisdom, Jessica (2009). Strategies for Promoting Healthier Food Choices; (vi) Ippolito, Pauline M.; Mathios, Alan D. (1990, 1995). Information and Advertising: Health Choices; (vii) Kiesel, Kristin; Villas-Boas, Sofia B. (2008). Another Nutritional Label: Experimenting with Grocery Store Shelf Labels and Consumer Choice; (viii) DellaVigna, Stefano (2009). Psychology and Economics: Evidence from the Field; (ix) Cur-

rie, Janet; Moretti, Enrico; Pathania, Vikram (2010). The Effect of Fast Food Restaurants on Obesity and Weight Gain.

- Quote snippet:

  "Observational Learning: Evidence from a Randomized Natural Field Experiment."

- Citation anchors:
    - page=37, words=0-7, section=References, note=Cai, Hongbin; Chen, Yuyu; Fang, Hanming (2009) Observational Learning: Evidence from a Randomized Natural Field Experiment
    - page=37, words=8-16, section=References, note=Chetty, Raj; Looney, Adam; Kroft, Kory (2009) Salience and Taxation: Theory and Evidence
    - page=37, words=17-22, section=References, note=Currie, Janet; Moretti, Enrico; Pathania, Vikram (2010) The Effect of Fast Food Restaurants on Obesity and Weight Gain
    - page=37, words=23-32, section=References, note=Downs, Julie S.; Loewenstein, George; Wisdom, Jessica (2009) Strategies for Promoting Healthier Food Choices
    - page=37, words=33-42, section=References, note=Ippolito, Pauline M.; Mathios, Alan D. (1990, 1995) Information and Advertising: Health Choices
    - page=37, words=43-52, section=References, note=Kiesel, Kristin; Villas-Boas, Sofia B. (2008) Another Nutritional Label: Experimenting with Grocery Store Shelf Labels and Consumer Choice
    - page=37, words=53-62, section=References, note=Jin, Ginger Zhe; Leslie, Phillip (2003) The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards
    - page=37, words=63-72, section=References, note=Jin, Ginger Zhe; Leslie, Phillip (2009) Reputational Incentives for Restaurant Hygiene
    - page=37, words=73-82, section=References, note=DellaVigna, Stefano (2009) Psychology and Economics: Evidence from the Field
    - page=37, words=83-92, section=References, note=Currie, Janet; Moretti, Enrico; Pathania, Vikram (2010) The Effect of Fast Food Restaurants on Obesity and Weight Gain

E06: Are there any seminal or classic references the paper positions itself against?

- Category: `E) Citations and related literature`
- Confidence: `low (0.24679432612986457)`
- Retrieval method: `local`
- Evidence type: `textual`
- Answer:

Yes. The paper engages with seminal nutrition-information literature and cites

classic works such as Jacoby, Chestnut, and Silberman (1977) on consumers' use of nutrition information, as well as Jacoby (1974) on information displays.

- Quote snippet:

  An early study by Jacob Jacoby, Robert W. Chestnut, and William Silberman (1977) presents evidence that consumers tend not to seek out nutrition information.

- Citation anchors:
    - page=7, words=275-566, section=B. Related Research, note=Cites an early, classic study (Jacoby, Chestnut, Silberman 1977) on consumer use of nutrition information
    - page=37, words=0-349, section=References, note=Cites Jacoby (1974) on consumer reaction to information displays

E07: Are there citations to code, data repositories, or appendices that are essential to the claims?

- Category: `E) Citations and related literature`
- Confidence: `low (0.29701151678608206)`
- Retrieval method: `local`
- Evidence type: `Appendix reference (survey instrument)`
- Data source: Transaction data and cardholder data (Starbucks datasets).
- Assumption flag: `False`
- Answer:

Yes. The Appendix contains the survey instrument (the actual two-page questionnaire) used in the study, which is essential to substantiate the survey-based claims.

- Quote snippet:

  The actual two-page questionnaire is shown in the Appendix.

- Citation anchors:
    - page=25, words=275-599, section=Survey methodology / Appendix, note=The actual two-page questionnaire is shown in the Appendix.

E08: What gaps in the literature do the authors say these citations leave open?

- Category: `E) Citations and related literature`
- Confidence: `low (0.292567878010417)`
- Retrieval method: `local`
- Evidence type: `literature gaps`
- Data source: Starbucks transaction data (over 100 million transactions) across Boston, NYC, and Philadelphia over 14 months
- Answer:

Prior literature leaves open questions about the long-run effects and time-path of calorie posting, heterogeneity in consumer responses, substitution patterns (e.g., smaller sizes, lower-calorie items, fewer items, or less frequent purchases), and effects on restaurants' profits, as well as generalizability beyond NYC and direct obesity outcomes.

- Quote snippet:

  We examine the time-path and longevity of the effect, for up to 11 months after calorie posting. Furthermore, we analyze heterogeneity in consumers' responsiveness to calorie posting. Lastly, the data we study provides a unique opportunity to analyze the impact of calorie posting on restaurants' profits.

- Citation anchors:
  - page=9, words=0-349, section=Literature review / Prior research, note=Authors discuss how prior NYC studies used small samples and focused on a narrow set of outcomes, and highlight gaps: time-path, longevity, substitution, heterogeneity, profits.
  - page=7, words=275-566, section=Comparison to prior research, note=Authors emphasize that their approach differs by using a large, long-run dataset and by examining time-path, substitution, heterogeneity, and profits.

F01: Are replication files or code provided? If so, where?

- Category: F) `Replication and transparency`
- Confidence: `low (0.16762116484088999)`
- Retrieval method: `local`
- Evidence type: `not stated in context`
- Answer:

No replication files or code are provided in the provided materials.

- Quote snippet:

  There are over 100 million transactions in the dataset.

- Citation anchors:
  - page=5, words=0-349
  - page=5, words=275-591
  - page=7, words=0-349
  - page=9, words=0-349
  - page=15, words=0-254
  - page=25, words=275-599
  - page=29, words=0-349
  - page=35, words=0-161
  - page=36, words=0-349
  - page=38, words=275-448

F02: Is there a pre-analysis plan or registered trial? Provide details if mentioned.

- Category: `F) Replication and transparency`
- Confidence: `low (0.2584380995154346)`
- Retrieval method: `local`
- Evidence type: `absence_of_mention`
- Data source: Starbucks transaction data and anonymized cardholder data across NYC, Boston, and Philadelphia.
- Answer:

No pre-analysis plan or registered trial is mentioned in the provided context.

- Quote snippet:

  To examine the effects of calorie posting while controlling for seasonality and other influences, we estimate regressions of the following form: (1) y_sct = x_sct + _POS Tc t + _sct

- Citation anchors:
    - page=page 5, words=0-15, section=A. Data Summary, note=No pre-analysis plan or preregistration mentioned; describes data sources.
    - page=page 11, words=275-322, section=II. Effect of Mandatory Calorie Posting, note=Describes regression specification; no preregistration mentioned.

F03: Are data access constraints disclosed (restricted access, proprietary data, NDAs)?

- Category: `F) Replication and transparency`
- Confidence: `low (0.27490718558170246)`
- Retrieval method: `local`
- Evidence type: `confidentiality constraints (restricted access/proprietary data/NDAs)`
- Data source: Starbucks transaction data and cardholder data (confidential).
- Answer:

Yes. The context discloses confidentiality constraints on data access, noting confidentiality requirements prevent revealing more details about the data and analyses.

- Quote snippet:

  To preserve confidentiality of competitively sensitive information, for both datasets, we normalize the value for NYC to one.

- Citation anchors:
    - page=5, words=0-349, section=Data confidentiality, note=To preserve confidentiality of competitively sensitive information for both

datasets.

- – page=6, words=0-349, section=Data confidentiality, note=Due to confidentiality requirements, we are unable to reveal any more details about these differences.

F04: Are key steps in data cleaning and construction documented?

- Category: `F) Replication and transparency`
- Confidence: `low (0.21841212612701594)`
- Retrieval method: `local`
- Evidence type: `Description of data cleaning and data construction in the data/methods sections.`
- Data source: Bollinger et al., Calorie Posting in Chain Restaurants, American Economic Journal: Economic Policy, 2011.
- Answer:

Yes. The authors document explicit data cleaning and construction steps, including excluding incomplete/open stores and extreme transactions, and normalizing NYC values for confidentiality.

- Quote snippet:

  To preserve confidentiality of competitively sensitive information, for both datasets, we normalize the value for NYC to one.

- Citation anchors:
    - – page=5, words=0-349, section=Data Cleaning and Panel Construction, note=Stores not open for the full data period are excluded (balanced panel); transactions with more than four units of any one item are excluded.
    - – page=5, words=275-591, section=Data Cleaning and Panel Construction, note=Combined dataset description including transaction and cardholder data; sample sizes mentioned.
    - – page=6, words=0-349, section=Confidentiality and Data Construction, note=Normalization of NYC values to one for confidentiality across both datasets.

F05: Are robustness and sensitivity analyses fully reported or partially omitted?

- Category: `F) Replication and transparency`
- Confidence: `low (0.28416116738125424)`
- Retrieval method: `local`
- Evidence type: `Robustness/sensitivity analysis mentioned`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia, Seattle).
- Answer:

Partially reported; the authors note a robustness check showing findings unchanged when excluding weather controls, but there is no broader set of sensitivity analyses reported.

- Quote snippet:

  Our findings are unchanged if we exclude the weather controls.

- Citation anchors:
    - page=11, words=275-522, section=Methods/Robustness, note=Robustness check: findings unchanged if weather controls are excluded.

G01: What populations or settings are most likely to generalize from this study?

- Category: `G) External validity and generalization`
- Confidence: `low (0.28520470190987823)`
- Retrieval method: `local`
- Evidence type: `limitations and scope of generalizability`
- Data source: Starbucks transaction and cardholder data (NYC, Seattle, Boston, Philadelphia)
- Answer:

Generalizable to other urban settings with calorie labeling on large chain restaurants and to similar commuter populations who frequently visit such chains; however, generalization to all customers or other chains is limited because the study uses data from a single chain (Starbucks) and a loyalty-based cardholder sample.

- Quote snippet:

  we have data for only one chain (Starbucks). We can- not know if the effects of mandatory calorie posting at Starbucks are similar to the effects at other chains.

- Citation anchors:
    - page=3, words=275-564, section=Limitations and scope, note=Data are from one chain; not sure if effects generalize to other chains.
    - page=5, words=0-349, section=Data summary, note=Describes transaction and cardholder data used for analysis.
    - page=20, words=0-349, section=Heterogeneity, note=Demographic variation (income, education, gender) suggests differential generalizability.
    - page=29, words=0-349, section=Commuters analysis, note=Commuter results inform generalization to commuter populations.
    - page=30, words=0-349, section=Discussion, note=Addresses learning vs salience and broader applicability of findings.

G02: What populations or settings are least likely to generalize?

- Category: `G) External validity and generalization`
- Confidence: `low` (0.21431592640208633)
- Retrieval method: `local`
- Evidence type: `limitation`
- Data source: Starbucks transaction and cardholder data (NYC, Seattle, etc.)
- Answer:

Findings in this study may not generalize to other restaurant chains or to the broader Starbucks customer base, since the data come from a single chain (Starbucks) and from a subset of cardholders who may not be representative of all Starbucks customers.

- Quote snippet:

  we have data for only one chain (Starbucks). We can't know if the effects of mandatory calorie posting at Starbucks are similar to the effects at other chains.

- Citation anchors:
  - page=page 3 words 275-564, words=275-564, section=Limitations, note=Only one chain (Starbucks); generalizability to other chains is uncertain.
  - page=page 5 words 0-349, words=0-349, section=A. Data Summary, note=Cardholder data may not be representative of Starbucks customers more generally.

G03: Do the authors discuss boundary conditions or scope limits?

- Category: `G) External validity and generalization`
- Confidence: `low` (0.20199246581637512)
- Retrieval method: `local`
- Evidence type: `limitations/scope boundaries`
- Data source: cardholder dataset; Starbucks data (one chain)
- Answer:

Yes. The authors discuss boundary conditions and scope limits, noting limitations such as having data from only one chain (Starbucks) and not knowing whether effects generalize to other chains or obesity outcomes.

- Quote snippet:

  There are two important limitations to this research. First, we do not directly measure the effect of calorie posting on obesity itself. A second limitation is that we have data for only one chain (Starbucks).

- Citation anchors:

– page=3, words=275-564, section=Limitations / Boundary conditions, note=Authors discuss boundary conditions: data from only one chain (Starbucks); inability to generalize to other chains or obesity outcomes.

G04: How might the results change in different time periods or markets?

- Category: `G) External validity and generalization`
- Confidence: `medium (0.35761374108322)`
- Retrieval method: `local`
- Evidence type: `Generalizability and time-period sensitivity of calorie-posting effects`
- Data source: Starbucks transaction and cardholder data around NYC calorie posting (April 2008 onward)
- Answer:

Results may differ in other time periods or markets because the study relies on Starbucks data in NYC with two control cities (Boston and Philadelphia); it may not generalize to other chains or markets. The calorie-posting effect began around April 1, 2008 and persisted through February 2009 in NYC, with some holiday-related attenuation in the transaction data; commuters showed larger reductions than non-commuters, and profit effects depend on local competition. Therefore, in different markets or periods, the magnitude, timing, and the balance between extensive and intensive margins could vary, and spillovers to non-posting stores would depend on whether learning or salience drives behavior.

- Quote snippet:

  Second limitation is that we have data for only one chain (Starbucks). We can- not know if the effects of mandatory calorie posting at Starbucks are similar to the effects at other chains.

- Citation anchors:
    - page=3, words=275-564, section=Limitations and generalizability, note=Acknowledges data from only one chain (Starbucks); cannot know effects at other chains.
    - page=14, words=0-349, section=Time path after posting, note=Drop in calories per transaction occurred around April 1, 2008 and persisted through Feb 2009; holiday pattern differs between transaction and cardholder data.
    - page=16, words=0-349, section=Substitution effects, note=Calorie posting affects margins via extensive vs intensive margins; effects vary by item and are not uniformly negative across items.
    - page=29, words=0-349, section=Commuters analysis, note=Commuters show larger reductions in calories per transaction in NYC post-period.

- page=24, words=275-564, section=IV. Discussion, note=Finds a substantial drop in food calories but negligible impact on beverage calories; profitability effects depend on local conditions.
- page=11, words=275-522, section=Methods and controls, note=Uses fixed effects and within-individual variation; controls for seasonality; discusses pre-trends.

H01: Are key variables measured directly or via proxies?

- Category: H) Measurement validity
- Confidence: low (0.3077210377229498)
- Retrieval method: local
- Evidence type: Direct measurement from purchase data and nutrition information; some survey-based measures.
- Data source: Transaction data and cardholder data; plus in-store customer surveys
- Table/Figure: Table 3; Table 4; Figure 3
- Answer:

Key variables such as calories per transaction and calories per item are measured directly from transaction data (via item-level calories) rather than via proxies; the posting indicator and related analyses rely on observed data, and survey data provide direct perceptual measures.

- Quote snippet:

  Based on the transaction data, we compute that, prior to calorie posting, in NYC: average drink calories per transaction were 143; average food calories per transaction were 104; and average total calories per transaction were 247.

- Citation anchors:
    - page=5, words=275-591, section=Data and measurement description, note=Calorie-related measures are derived and discussed; cardholder vs transaction data context.
    - page=6, words=0-349, section=Data description and calorie per transaction, note=Prior to calorie posting, average calories per transaction are reported and are computed from transaction data.
    - page=11, words=275-522, section=Methodology/Regression specification, note=y_sct is a measure of calories per transaction; calories per transaction defined from data.
    - page=12, words=275-624, section=Methodology/Regression specification (continuation), note=Continued description of calories-per-transaction measurement from data.
    - page=25, words=275-599, section=Survey methodology, note=In-store customer surveys provide direct perceptual measures; supports direct measurement of beliefs about calories.

H02: What measurement error risks are acknowledged or likely?

- Category: H) Measurement validity
- Confidence: low (0.33244058390053044)
- Retrieval method: local
- Evidence type: measurement_error and data limitations
- Data source: Starbucks transaction data and cardholder data; Seattle and San Francisco survey data
- Table/Figure: Figure 5
- Answer:

Measurement error risks acknowledged include: (a) large errors in self-reported calorie estimates of beverages (Figure 5) leading to potential mismeasurement of calories; (b) unobserved calorie sources such as milk additions not captured by transaction or cardholder data; and (c) potential sampling bias and changes in respondent composition across survey waves, which can bias estimates of knowledge and attitudes toward calories.

- Quote snippet:

  Figure 5 shows the distribution of errors—predicted minus actual calories—in respondents' best guess for their purchased beverage.

- Citation anchors:
  - page=26, words=0-178, section=Survey measurement errors, note=Discusses errors in respondents' calorie estimates (Figure 5) and potential sampling concerns across waves.
  - page=6, words=275-517, section=Measurement limitations in calorie data, note=Notes that milk additions are a source of additional calories not observed in data; potential measurement error.
  - page=27, words=0-349, section=Survey validity and sampling, note=Mentions possible sampling bias between waves and the possibility that observed effects reflect changes in who was surveyed.

H03: Are there validation checks for key measures?

- Category: H) Measurement validity
- Confidence: low (0.24862979131144963)
- Retrieval method: local
- Evidence type: robustness checks and validation described in methods and results
- Data source: Transaction data and Cardholder data (Starbucks NYC/Boston/Philadelphia) used to study calorie posting effects.
- Table/Figure: Table 3; Table 4; Table 6
- Answer:

Yes. The paper performs multiple validation checks for key measures, including data-cleaning exclusions, cross-dataset validation (transaction data vs. cardholder data), seasonality and weather robustness checks, dummy controls for

day of week and holidays, cross-city robustness (e.g., Seattle post-law comparison), and distributional (quantile) analyses that confirm the results across the calorie distribution.

- Quote snippet:

  we include day-of-week dummies and holiday dummies that differ
  for NYC and the controls, finding no difference in the estimates.

- Citation anchors:
  - page=6, words=0-349, section=Data cleaning/sample restrictions, note=Excluded transactions represent 2.2% of all transactions; data cleaning.
  - page=11, words=275-522, section=Validation across datasets, note=Cross-check: regressions estimated separately on transaction data and cardholder data.
  - page=12, words=550-630, section=Weather/seasonality controls, note=Weather controls included; findings unchanged if weather controls are excluded.
  - page=14, words=275-590, section=Robustness checks, note=Day-of-week and holiday dummies differ for NYC and controls; Seattle post-law robustness; estimates unchanged.
  - page=16, words=275-558, section=Distributional robustness, note=Quantile and within-individual analyses; controls for mean reversion with Boston/Philly.
  - page=21, words=275-421, section=Quantile analysis, note=Estimates by quantile; larger reductions at higher quantiles.
  - page=22, words=0-349, section=Quantile results details, note=Table 6 quantile results.

H04: Do the authors discuss construct validity for core outcomes?

- Category: `H) Measurement validity`
- Confidence: `low (0.20310003521525116)`
- Retrieval method: `local`
- Evidence type: `absence of explicit construct validity discussion; measurement definitions and limitations are described instead`
- Data source: Starbucks cardholder transaction data (limited to one chain), with surveys conducted at multiple locations
- Answer:

No. The provided excerpts do not show an explicit discussion of construct validity for core outcomes.

- Quote snippet:

  Ultimately, whether calorie posting affects people's behavior is an
  empirical question.

- Citation anchors:
  - page=3, words=275-564, section=Limitations and data scope, note=Discusses limitations and empirical focus; no explicit construct validity discussion.
  - page=7, words=275-566, section=Methods/Survey data, note=Describes knowledge surveys; no explicit construct validity discussion.
  - page=8, words=0-349, section=Literature review/Background, note=Reviews prior nutrition labeling literature; no construct validity discussion.
  - page=11, words=0-349, section=Methods/Empirical strategy, note=Defines outcome as calories per transaction; no explicit construct validity discussion.
  - page=11, words=275-522, section=Methods/Data/Identification, note=Presents regression specification; no construct validity commentary.
  - page=28, words=0-349, section=Results/Discussion, note=Table 8 on purchase-decision factors and calorie importance; measurement via ratings; no construct validity discussion.
  - page=32, words=0-349, section=Discussion/Limitations, note=Discusses challenges in assessing menu offerings; no construct validity discussion.
  - page=37, words=275-624, section=Discussion/References, note=References and synthesis; no explicit construct validity discussion.
  - page=38, words=0-349, section=References/Support, note=References; no explicit construct validity discussion.
  - page=38, words=275-448, section=References (continued), note=References; no explicit construct validity discussion.

I01: What policy counterfactuals are considered or implied?

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.34069116049166176)
- Retrieval method: local
- Evidence type: counterfactual analysis
- Data source: Starbucks calorie posting study dataset (transaction and cardholder data; NYC, Boston, Philadelphia) and related surveys
- Assumption flag: True
- Assumption notes: Assumes control cities provide a valid counterfactual and that within-city variation identifies the policy effect; assumes no differential pre-trends between NYC and control cities; uses fixed effects to control for confounders.
- Answer:

Counterfactuals considered include (1) using Boston and Philadelphia as control cities with no calorie posting to compare against NYC, and (2) within-city pre/post analysis in NYC to approximate the counterfactual absent posting;

the authors also discuss spillovers to nonposting stores among commuters as a learning-type counterfactual and evaluate learning versus salience through survey and commuter evidence.

- Quote snippet:

  To control for other factors affecting transactions, we also observe every transaction at Starbucks company stores in Boston and Philadelphia, where there was no calorie posting.

- Citation anchors:
  - page=2, words=0-349, section=Data/experimental design, note=Describes Boston and Philadelphia as control cities where there was no calorie posting.
  - page=12, words=0-349, section=Identification strategy, note=Policy variation identified at the city-week level; use of controls for identification.
  - page=11, words=275-522, section=Method: regression framework, note=Describes regression approach with NYC post indicator to assess within-city changes after posting.
  - page=30, words=0-349, section=Results/discussion on learning and salience, note=Survey and commuter evidence discuss learning versus salience as mechanisms.
  - page=30, words=0-349, section=Results; non-NYC store effects, note=Commuter evidence indicates learning if reductions occur only after prior NYC exposure; touches on nonposting store effects.

I02: What are the main welfare tradeoffs or distributional impacts discussed?

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.31026132151151437)
- Retrieval method: local
- Evidence type: empirical results from the study
- Data source: Transaction data and cardholder data (NYC and control cities)
- Table/Figure: Table 6 and Figure 4
- Answer:

Calorie posting lowers total calories mainly through the extensive margin (fewer food purchases), with beverage calories largely unchanged. Welfare implications include potential health benefits from reduced intake, but distributional impacts are notable: larger reductions among higher-income and more educated ZIPs and among female cardholders, and bigger absolute reductions for higher-calorie (top quantile) purchasers. Prices per item rise and profits are largely unaffected, and there is little or no change in visit frequency, indicating mixed welfare effects with notable inequality in who responds most.

- Quote snippet:

  Hence, nearly three quarters of the total calorie reduction can be attributed to people opting not to buy food items (i.e., the extensive margin of food demand).

- Citation anchors:
  - page=18, words=0-349, section=Discussion of margins, note=Food calories drop mainly due to the extensive margin (not buying food items); beverage calories largely unaffected.
  - page=20, words=0-349, section=Heterogeneity in the Effect of Mandatory Calorie Posting, note=Higher income and more education in ZIPs associated with larger calorie reductions; female cardholders more responsive.
  - page=22, words=0-349, section=Quantile effects, note=Calorie reductions larger in top quantiles; percent reduction roughly 5–6% from 75th to 99th percentile.
  - page=24, words=275-564, section=Profit discussion, note=Prices per item rise; profits largely unaffected; policy costs discussed.

I03: Are cost-benefit or incidence analyses provided?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.31295890298721923)`
- Retrieval method: `local`
- Evidence type: `Qualitative discussion; no formal cost-benefit or incidence analysis.`
- Table/Figure: Figure 1
- Answer:

No formal cost-benefit or incidence analysis is provided; the paper discusses potential costs (compliance costs and indirect costs) and potential benefits (learning and salience) in a qualitative way, and notes impacts on revenues/profits without presenting a quantified cost-benefit or incidence framework.

- Quote snippet:

  One news report indicated the cost of compliance for the Wendy's chain was about $2,000 per store.

- Citation anchors:
  - page=4, words=275-576, section=Costs of posting, note=Cites cost of compliance and indirect costs of menu differences across cities.
  - page=24, words=275-564, section=IV. Discussion, note=Discusses profits and qualitative implications of costs/benefits; no quantified CBA presented.

I04: What policy recommendations are stated or implied?

- Category: I) `Policy counterfactuals and welfare`
- Confidence: `low (0.33677666005489904)`
- Retrieval method: `local`
- Evidence type: `Conclusion / policy implications`
- Data source: Starbucks NYC transaction data; Boston and Philadelphia control stores; anonymous Starbucks cardholder data; in-store surveys
- Answer:

The authors advocate mandatory calorie posting on restaurant menus if the benefits outweigh the costs, arguing that the costs are very low, the policy could yield long-run benefits (including encouragement of lower-calorie options), and there may be public-education benefits; they also note that voluntary posting would likely yield different results.

- Quote snippet:

  the costs of calorie posting are very low—so even these small benefits could outweigh the costs.

- Citation anchors:
    - page=33, words=275-553, section=V. Conclusion, note=Policy implications and limits discussed in the conclusion
    - page=34, words=0-349, section=V. Conclusion, note=Direct policy recommendations: mandatory posting if benefits outweigh costs; low costs; potential long-run benefits

J01: What missingness or attrition patterns are reported?

- Category: J) `Data quality and integrity`
- Confidence: `low (0.315684807818624)`
- Retrieval method: `local`
- Evidence type: `exclusion_and_selection_patterns`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia); balanced panel construction
- Answer:

Two patterns of missingness/attrition are reported: (1) Exclusion of transactions to form a balanced panel—stores not open for the entire data period and transactions with more than four units—representing about 2.2% of all transactions; (2) Selection of a cardholder subsample—only anonymous cardholders with at least one transaction per week in NYC, Boston, or Philadelphia before NYC calorie posting are included, which may not be representative of the broader Starbucks customer base.

- Quote snippet:

  We exclude transactions at stores that were not open during the entire data period (i.e., we analyze the balanced panel). The excluded

transactions represent only 2.2 percent of all transactions.

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=Excludes stores not open during the entire data period to form a balanced panel and excludes bulky transactions; 2.2% of transactions are excluded.
  - page=5, words=275-591, section=A. Data Summary, note=Cardholder data sample restricted to cardholders with at least one weekly transaction in NYC, Boston or Philadelphia before NYC posting; potential non-representativeness.

J02: How are outliers handled (winsorization, trimming, exclusions)?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.31113196440612556)`
- Retrieval method: `local`
- Evidence type: `data exclusions / sample cleaning`
- Data source: Transaction data and cardholder data
- Answer:

Outliers are handled via data exclusions rather than winsorization or trimming: the analysis excludes transactions at stores not open for the entire data period (balanced panel) and transactions that included more than four units of any one item. These excluded transactions total about 2.2% of all transactions (and are believed to be well under 5%).

- Quote snippet:

  we exclude transactions at stores that were not open during the entire data period (i.e., we analyze the balanced panel), and we exclude transactions that included more than four units of any one item because we consider these purchases to be driven by fundamentally different processes (bulk purchases for an office, say). The excluded transactions represent only 2.2 percent of all transactions.

- Citation anchors:
  - page=5, words=275-591, section=Data exclusions and sample selection, note=Exclude stores not open during entire period and transactions with more than four units; 2.2% excluded; under 5% overall.

J03: Are there data audits or validation steps described?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.2431706842394831)`
- Retrieval method: `local`
- Evidence type: `validation and robustness checks`
- Data source: Transaction data (universe of Starbucks transactions) and cardholder data; milk order data for supply context; Seattle data for robustness

- Table/Figure: Figure 1; Figure 2; Table 1; Table 3; Table 5
- Answer:

Yes. The paper describes validation-like steps rather than a formal data audit, including data cleaning, use of two datasets for cross-checks, pre-trend tests, and multiple robustness checks.

- Quote snippet:

  First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia.

- Citation anchors:
  - page=5, words=0-60, section=Data sources and data handling, note=Intro of transaction and cardholder data and their use to separate intensive/extensive margins.
  - page=6, words=0-80, section=Data cleaning and data handling, note=Balance panel, exclusion of non-open stores, and large-bulk purchases.
  - page=6, words=230-270, section=Data sources, note=Two datasets used for cross-validation.
  - page=14, words=0-80, section=Validation/Robustness checks, note=Pre-trend differences check and cross-dataset consistency; within-individual variation.
  - page=14, words=200-260, section=Robustness checks, note=Controls for day-of-week and holidays; within-city robustness.
  - page=20, words=0-60, section=Heterogeneity and robustness, note=Gender and purchase-frequency heterogeneity; cardholder data used.
  - page=25, words=0-70, section=Supplementary data collection, note=Milk order data and in-store surveys as supplementary checks.
  - page=14, words=275-350, section=Seattle robustness, note=Seattle data used to validate findings.

J04: Is there evidence of reporting bias or selective sample inclusion?

- Category: J) Data quality and integrity
- Confidence: low (0.3050681117858262)
- Retrieval method: local
- Evidence type: Sampling bias and limited representativeness
- Data source: Combination of transaction data and anonymous cardholder data from Starbucks datasets
- Answer:

Yes. The study acknowledges sampling limitations: the cardholder data may not be representative of Starbucks customers in general because the sampled individuals are above-average loyal to Starbucks; the deck relies on a subsample of cardholders who averaged at least one transaction per week in NYC, Boston

or Philadelphia, which may introduce selection bias. The authors also note that the survey sample in Seattle likely under-represents customers who order regular coffee, indicating potential sampling bias in reporting.

- Quote snippet:

  these cardholders may not be representative of Starbucks customers more generally, as we expect these individuals are above average in their loyalty to Starbucks.

- Citation anchors:
  - page=5, words=275-591, section=A. Data Summary, note=Cardholder sample may not be representative; high loyalty to Starbucks.
  - page=25, words=275-599, section=Survey methodology, note=Survey sample under-represents consumers who ordered regular coffee; potential reporting bias.

K01: What goodness-of-fit or diagnostic metrics are reported?

- Category: K) Model fit and diagnostics
- Confidence: low (0.24714003186543324)
- Retrieval method: local
- Evidence type: statistical_metrics
- Data source: American Economic Journal: Economic Policy, Bollinger et al. 2011
- Answer:

They report R-squared (R^2) values as a measure of fit and provide regression standard errors (in parentheses) with significance stars indicating statistical significance. Specifically, R^2's range from 0.27 to 0.82 for the transaction data and 0.26 to 0.37 for the cardholder data; standard errors accompany coefficient estimates, and significance stars denote p-values.

- Quote snippet:

  R^2's ranging from 0.27 to 0.82. The regressions using the cardholder data in the top panel are based on 1,511,516 observations, and the R^2 vary between 0.26 and 0.37.

- Citation anchors:
  - page=page 18, words=0-349, section=Notes, note=R^2's ranging from 0.27 to 0.82 for the transaction data; R^2 vary between 0.26 and 0.37 for the cardholder data.
  - page=page 18, words=0-349, section=Notes, note=Standard errors are reported in parentheses for coefficient estimates; significance stars indicate statistical significance.
  - page=page 18, words=0-349, section=Notes, note=Significance stars denote p-values (e.g., *** for 1%).

K02: Are functional form choices tested (logs, levels, nonlinearities)?

- Category: K) Model fit and diagnostics
- Confidence: low (0.267823785887138)
- Retrieval method: local
- Evidence type: empirical model specification
- Data source: Starbucks cardholder and transaction data
- Table/Figure: Table 9
- Answer:

Yes. The authors use log transformations for key dependent variables (e.g., log(calories + 1) and log daily sales) and include nonlinear terms via squared weather variables, indicating testing of logs, levels, and nonlinear forms.

- Quote snippet:

  "The regression also includes individual, week and day-of-week fixed effects, holiday dummies and weather controls,

- Citation anchors:
  - page=16, words=0-349, section=B. Substitution Effects, note=Regressions include log daily sales and log measures of calories (e.g., log(calories per transaction), log(calories per item purchased)).
  - page=11, words=0-349, section=Regression specification with weather controls, note=Weather controls include temperature and temperature squared, precipitation and precipitation squared.
  - page=29, words=0-349, section=Table 9, note=Dependent variable is log(calories + 1) for commuters; tests post-April-08 effects.
  - page=16, words=275-558, section=B. Substitution Effects – additional forms, note=Dependent variables include log calories per beverage purchased and log calories per item purchased.

K03: Are residual checks or specification tests reported?

- Category: K) Model fit and diagnostics
- Confidence: low (0.23563782282195578)
- Retrieval method: local
- Evidence type: robustness/specification tests in methodology
- Data source: Bollinger et al., Calorie Posting in Chain Restaurants (AEJ Economic Policy)
- Answer:

Yes. They report specification/robustness tests via alternative model specifications (e.g., separate NYC week effects; different control cities), use of fixed effects and weather controls, and robust standard errors with confidence intervals.

- Quote snippet:

An alternative approach is to modify equation (1) to include separate week dummies for NYC and the control cities, and to exclude the POST variable.

- Citation anchors:
  - page=11, words=0-349, section=Regression specification and controls, note=Describes the regression form with controls including week fixed effects and weather controls.
  - page=13, words=275-528, section=Robustness/Alternative specifications, note=States an alternative specification with separate week dummies for NYC and controls to test timing.
  - page=15, words=0-254, section=Figure 2 and regression results, note=Discusses regression with NYC vs control week effects and 95% confidence intervals.
  - page=29, words=0-349, section=Table 9 robustness, note=Notes robust standard errors in parentheses (robustness to heteroskedasticity).

K04: How sensitive are results to alternative specifications or estimators?

- Category: `K) Model fit and diagnostics`
- Confidence: `medium (0.37672306038408554)`
- Retrieval method: `local`
- Evidence type: `robustness checks across specifications and estimators`
- Data source: Transaction data and cardholder data from Starbucks calorie posting study
- Answer:

Results are robust to alternative specifications and estimators. Including date fixed effects barely changes the estimates; removing weather controls does not alter conclusions; aggregating pre/post posting to address serial correlation yields similar effects; using cardholder data versus transaction data gives broadly consistent findings, with beverage calories in the cardholder data remaining largely insignificant in one specification. Overall, the inference that calorie posting reduces calories per transaction remains supported.

- Quote snippet:

  As a robustness check, we include date fixed effects in the above specification (and therefore drop the day-of-week and week fixed effects). The estimates based on the transaction data are barely changed.

- Citation anchors:
  - page=12, words=0-349, section=Robustness checks: date fixed effects, note=As a robustness check, we include date fixed effects ...

The estimates based on the transaction data are barely changed.
- page=12, words=275-624, section=Robustness checks: cardholder vs transaction data, note=The cardholder data allows us to isolate the effect ... estimates reported in the second column.
- page=20, words=0-349, section=Robustness checks: extensive margin, note=No statistically significant change in the frequency of cardholders' purchases in NYC relative to control cities.
- page=12, words=550-630, section=Robustness checks: standard errors, note=Notes: In each case the standard errors are similar to the results in Table 3.

P01: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')

- Category: P) Previous questions

- Confidence: n/a

- Retrieval method: n/a

- Evidence type: n/a

- Answer:

- Scope and data (context for the study)

  - NYC calorie posting law required at point of purchase; discussion of costs and counterarguments. See context describing the policy and related views (page 4, words 275-576).
  - Dataset: universe of over 100 million Starbucks transactions in Boston, NYC, and Philadelphia over 14 months; 1.5 million cardholder transactions; Seattle used as a placebo/contrast. See description of data and design (page 9, words 0-349).

- Key empirical setup

  - Primary specification uses a binary calorie-posting dummy POST = 1 for NYC stores after Apr 1, 2008; regression controls include fixed effects and weather, with week-specific effects. See discussion of Table 3/regression setup (page 13, words 275-528).

- Main findings

  - Calories per transaction declined in NYC after posting, but the decline is not large in absolute terms; NYC vs control comparison shows a drop in log(calories per transaction) around posting time. See Figure 1 and accompanying discussion (page 9, words 0-349; page 10, 0-186).
  - Overall calories per store-day fell by about 4.6 calories, reflecting that higher transaction volume in NYC offset part of the per-transaction

63

decrease. See a key sentence tying the store-day figure to the per-transaction effect (page 13, words 275-528).

    – Seattle results: beverage calories per transaction fell by 4.6 calories; food calories per transaction rose by 0.8 calories, illustrating heterogeneous effects by item type when calories aren't posted for all items in that city (page 10, 0-186).

    – Beverages vs. foods (Table 4):

        ∗ Number of beverages and foods per transaction; but calorie effects are item-specific:

            · log(beverage calories per beverage) -0.008 (significant)

            · log(food calories per food item) -0.039 (larger, significant)

            · Overall, substantial portion of calorie reductions come from consuming fewer food items (extensive margin). See Table 4 and associated notes (page 18, 0-349).

    – Extensive margin effect (dominant mechanism): nearly three quarters of total calorie reduction due to opting not to buy food items. See the explicit statement in Table 4 discussion (page 18, 0-349).

    – Distributional/quantile results: larger reductions at higher quantiles; the percent change is roughly 5–6% across 75th–99th percentiles, with absolute reductions rising in the upper tail (e.g., about -77 calories at the 99th percentile). See Table 6 and related text (page 22, 0-349).

- Heterogeneity and implications

    – Time path and longevity: the study tracks effects up to 11 months post-posting to assess persistence; results discuss potential short-run vs. long-run effects (page 9, words 0-349; page 13, 275-528).

    – Heterogeneity in consumer response and potential profit implications are analyzed; the data allow testing for product substitution patterns and consumer knowledge (page 9, 0-349; page 15, 0-254; page 18, 0-349).

- Methodological notes

    – Prior related findings sometimes show no statistically significant effect due to large standard errors; Bollinger et al. position their approach as larger-scale and with more granular data (page 9, 0-349).

    – Confidentiality: exact values of some internal counts (N) are not disclosed; the Appendix notes this confidentiality constraint (page 18, 0-349).

- Policy context and cost considerations

    – The NYC calorie posting requirement is argued to have in-store/online calorie access; counterarguments cite implementation costs (e.g., about $2,000 per Wendy's store) and potential indirect costs from multi-city menus (page 4, words 275-576).

- Quick formula/objective summary

- Primary model (conceptual): log(Calories_per_transaction) = _store + _week + · POST + X +
- POST = 1 if NYC store after 1 Apr 2008; otherwise 0
- captures the effect of calorie posting on calories per transaction; same framework extended to items per transaction and per-item calories (as reported in Tables 3–6) (page 13, 275-528; page 18, 0-349).

If you want, I can extract specific numbers for a compact table (e.g., key estimates, standard errors, and qualitative interpretation) with precise page references.