Audit Report: Workflow 51af9a22de5b4c01a31c56930f1bd3ce

Overview

- Source JSON: reports\workflow-report-51af9a22de5b4c01a31c56930f1bd3ce.json
- Run ID: 51af9a22de5b4c01a31c56930f1bd3ce
- Papers input: papers\Calorie Posting in Chain Restaurants - Bollinger et al. (2011).pdf
- Started at: 2026-02-16T00:47:01.282461+00:00
- Finished at: 2026-02-16T00:55:32.638748+00:00
- Duration: 0:08:31.356287

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: 0.27966953317781795 / 0.2842173325501089
- Confidence labels: low=74, medium=9, high=0

Final Answer

- Main contribution
  - Documents the impact of NYC calorie posting on consumer calories at the point of purchase, using NYC vs control cities and a Seat-

tle placebo test; supplements with primary survey data (Appendix A). This isolates behavioral responses to posted calories and decomposes margins (extensive vs intensive). (Context and framing: page 4, words 275–576; Appendix note: page 35–36)

- Identification strategy
  - Primary approach: a multi-city, pre-post comparison (NYC vs control cities) to capture the posting effect, with Seattle serving as an external placebo where posting did not occur. This strengthens attribution to labeling rather than concurrent trends. (Seattle placebo test: page 15)
  - Data sources include both store-level outcomes and customer surveys (Appendix A), enabling both aggregate and micro-level insights. (Appendix reference: page 35–36)
- Key results
  - Overall effects
    - Calorie posting reduces calories per transaction and per store-day; average calories per store-day fall 4.6 after posting. Reduction in calories per transaction is larger than the store-day reduction, implying a rise in transactions. Decomposition: ΔC_total = ΔC_extensive + ΔC_intensive, with substantial share from the extensive margin. (page 13)
  - Item-level and margin effects (Table 4)
    - Beverages: ΔBev +0.005 per transaction; log calories per beverage −0.008.
    - Foods: ΔFood −0.029 per transaction; log calories per food item −0.039.
    - Net: total items per transaction declines (log scale −0.027); roughly 75% of total calorie reduction from not purchasing food items (extensive margin). (page 18)
  - Seattle placebo test
    - In Seattle (no posting for food calories), drink calories per transaction fall 4.6; food calories per transaction rise 0.8, supporting the idea that posted calories drive beverage reductions but not a broad food-calorie shift when posting is absent. (page 15)
  - Heterogeneity (Table 5)
    - Effects vary with demographics and settings: significant interactions for median income, percent with college degree, and percent female; high-calorie customers exhibit larger reductions. (page 21)
  - Quantile results (Table 6)
    - Larger calorie reductions at higher quantiles of calories per transaction (75th–99th percentiles): substantial absolute reductions, about 5–6% relative. The 99th percentile shows the largest absolute decrease (−77 calories) with ~5–6% relative change. (page

22)
  - Data and survey instruments
    * Appendix A shows Stanford/Starbucks customer survey forms, indicating primary data collection from customers. (page 35–36)

- Limitations and caveats
  - External validity may be limited to contexts with similar posting policies and city characteristics; results differ across cities and cohorts, as shown by heterogeneity and cross-city comparisons. (inference from themes: page 4; Table 5: page 21)
  - Despite Seattle placebo support, attribution to posting cannot be proven beyond all confounds; some contemporaneous factors could still influence outcomes, though the placebo test mitigates this concern. (Seattle placebo: page 15)
  - Results show substantial heterogeneity (demographics, high- vs low-calorie cohorts) and are more pronounced at higher quantiles, suggesting that effects may not be uniform across all consumers or menu contexts. (Table 5: page 21; Table 6: page 22)

- Quick, figure-ready takeaways (if you want a compact table)
  - Overall: ΔC_tx < 0; ΔC_sd < 0; ΔC_tx magnitude $>$ ΔC_sd; transactions up.
  - Margin decomposition: 75% of calorie reduction via extensive margin (not buying food items).
  - Item-level shifts: beverages up slightly; foods down; net items down.
  - Heterogeneity: larger effects for higher-calorie consumers; income/education/gender interactions present.
  - Quantiles: bigger absolute reductions at the top end (e.g., 99th percentile -77 calories; ~5–6% relative change).

If you'd like, I can produce a compact, figure-ready one-page table of coefficients and percent changes by margin and quantile with exact numbers and page citations.

Sub-Answers

Sub-answer 1

- Question: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')
- Question tokens estimate: `19`
- Answer:
- Key context: NYC calorie posting policy prompted concerns about costs and potential behavioral responses; evidence includes multiple analyses

from NYC vs control cities and Seattle as a control. (page 4, words 275-576)

- Main quantitative finding on overall calories: Calorie posting reduces calories per transaction and per store-day, with a documented drop in average calories per store-day of about 4.6 after posting; the reduction in calories per transaction is larger than the store-day reduction, implying a rise in transactions alongside lower per-transaction calories. (page 13, words 275-528)
- Item-level and margin effects (Table 4):
  - Beverages: number of beverages per transaction increases slightly (ΔBev $+0.005$); calories per beverage fall slightly (log calories per beverage -0.008).
  - Foods: number of food items per transaction falls (ΔFood -0.029); calories per food item fall more (log calories per food item -0.039).
  - Net: total items per transaction decline (log scale -0.027), with roughly three-quarters of the total calorie reduction coming from not purchasing food items (extensive margin). (page 18, words 0-349)
- Seattle check (placebo / external test): In Seattle (where food calories were not posted), drink calories per transaction fell by about 4.6 and food calories per transaction rose about 0.8, consistent with the idea that posted calories drive beverage reductions but not food-calorie changes where posting didn't occur. This supports the NYC finding on beverages and the absence of a large food-calorie shift when posting is absent. (page 15, words 0-254)
- Heterogeneity in impact (Table 5): The effect of posting on log(calories per transaction) varies with characteristics:
  - A few interactions are statistically significant (e.g., median income, percent with college degree, percent female; high- and low-calorie cohorts show differential responses).
  - Notably, high-calorie customers show larger reductions. (page 21, words 0-349)
- Quantile results (Table 6): Calorie reductions are larger at higher quantiles of calories per transaction:
  - 75th–99th percentile reductions are substantial in absolute calories yet roughly 5–6% in percentage terms.
  - The 99th percentile shows the largest absolute decrease (-77 calories) with a still ~5–6% relative change. (page 22, words 0-349)
- Appendix / survey instrument: The paper includes Appendix A with the Stanford University research survey and Starbucks customer survey forms, indicating primary data collection from customers. (page 35-36, words 0-349)

- Additional notes:
  - The results and discussion reference the array of potential consumer responses (extensive vs intensive margins, substitution effects) and discuss the role of different menu postings across cities. (page 4, 275-576; page 10-12 context)

If you'd like, I can extract a compact figure-ready summary (e.g., a one-page table of coefficients and percent changes by margin and quantile) with the exact numbers and cited 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 (0.31591268523688953)`
- Retrieval method: `local`
- Evidence type: `empirical study of calorie posting effects on consumer behavior`
- Data source: Starbucks transaction data and cardholder data (NYC, Boston, Philadelphia)
- Answer:

The main research question is whether mandatory calorie posting in chain restaurants affects consumer behavior, specifically whether it changes calories per transaction (overall calories purchased) and how this effect occurs (learning vs. salience).

- Quote snippet:

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

- Citation anchors:
  - page=3, words=275-564, section=Introduction, note=Contains the claim that whether calorie posting affects 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 (0.29755037116853844)`
- Retrieval method: `local`
- Evidence type: `Author-stated contribution / novelty in methodology and scope`

- Data source: Starbucks transaction data (Boston, NYC, Philadelphia); anonymous cardholder data; in-store consumer surveys
- Answer:

The paper's primary contribution relative to the existing literature is presenting a much larger and broader empirical analysis of mandatory calorie posting, using transaction-level data across multiple cities plus individual-level data and consumer surveys to study not only the overall effect on calories per transaction but also mechanisms (learning vs. salience), time-path, product substitution, heterogeneity, and potential profit implications.

- 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. We also analyze individual-level data (1.5 million transactions) over time, as well as a survey

- Citation anchors:
  - page=page 9, words=0-349, section=Introduction / Contribution, note=Authors claim their dataset is larger and broader than prior work and include transaction-level data, multiple cities, individual-level data, and a consumer knowledge survey to study mechanism and longevity.

A03: What is the central hypothesis being tested?

- Category: `A) Research question / contribution`
- Confidence: `low (0.29008151226673723)`
- Retrieval method: `local`
- Evidence type: `hypothesis statement / research question`
- Data source: Starbucks transaction data and cardholder data
- Table/Figure: Table 9
- Answer:

The central hypothesis is that mandatory calorie posting affects consumer purchasing behavior, specifically reducing calories per transaction, and that this effect may operate through learning and/or salience mechanisms.

- Quote snippet:

  > If learning effects are important (and commuters' memories are not too short), then we might expect these consumers to reduce their calories per transaction everywhere (even outside of NYC) after being exposed to calorie information in NYC.

- Citation anchors:
  - page=29, words=0-349, section=Testing mechanisms (learning vs. salience), note=Sets up hypothesis about learning vs salience and tests with NYC exposure

  - page=30, words=0-349, section=Results and interpretation, note=Finds reductions in calories per transaction and discusses learning and salience support

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.3046419921486621)`
- Retrieval method: `local`
- Evidence type: `textual description of dependent variables from regression specification`
- Data source: American Economic Journal: economic policy february 2011 article by Bollinger et al.
- Table/Figure: Table 3
- Answer:

The main dependent variables are the calories per transaction for beverages, the calories per transaction for food, and the total calories per transaction (beverage + food).

- Quote snippet:

  > In the top row, the dependent variable is log(beverage calories per transaction). In the second row, the dependent variable is log(food calories per transaction). In the bottom row, we report the estimated impact on log(beverage + food calories).

- Citation anchors:
  - page=11, words=275-522, section=Dep Var description (Table 3), note=Defines dependent variables as log(beverage calories per transaction), log(food calories per transaction), and log(beverage + food calories)
  - page=12, words=0-349, section=Regression specification continuation, note=Represents the same dependent variables in the cardholder and transaction data analysis

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.2439181675693594)`
- Retrieval method: `local`
- Evidence type: `Regression specification / variable definition`
- Data source: cardholder and transaction datasets (within-city vs control-city variation)
- Answer:

The key treatment/exposure variable is the calorie-posting indicator POST: a dummy equal to one for NYC stores on every day after April 1, 2008.

- Quote snippet:

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

- Citation anchors:
  - page=page 11 words 0-349, words=0-349, section=Regression specification, note=Defines POST as a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008).

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

- Category: `A) Research question / contribution`
- Confidence: `medium (0.3530092910389292)`
- Retrieval method: `local`
- Evidence type: `textual_context_description`
- Data source: Starbucks transaction data and cardholder data (NYC, Boston, Philadelphia); plus Seattle control surveys
- Answer:

United States; New York City calorie-posting law in chain restaurants (Starbucks), with Boston and Philadelphia as controls; data from NYC, Boston, and Philadelphia collected Jan 1, 2008–Feb 28, 2009 (about 14 months), plus in-store surveys in Seattle for comparison.

- Quote snippet:

  We study the impact of mandatory calorie posting on consumers' purchase decisions using detailed data from Starbucks.

- Citation anchors:
  - page=1, words=0-349, section=Intro/Context, note=NYC calorie-posting policy; study scope with Starbucks data.
  - page=5, words=0-349, section=Data Summary, note=Data: 222 NYC stores, 94 Boston/Philadelphia stores; period Jan 1 2008–Feb 28 2009.
  - page=5, words=275-591, section=Data Summary (continued), note=Period: 3 months before and 11 months after posting (Jan 1 2008–Feb 28 2009); over 100 million transactions.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.2858720961676016)`
- Retrieval method: `local`
- Evidence type: `empirical evidence from surveys and cardholder data demonstrating both learning and salience mechanisms`
- Answer:

A combination of learning and salience effects drives the effect of calorie posting.

- Quote snippet:

  > These results support the role of learning. Hence, we find evidence in support of both learning and salience as part of the mechanism for why calorie posting causes consumers to reduce calories per transaction.

- Citation anchors:
  - page=30, words=0-349, section=IV. Discussion, note=Authors argue the mechanism includes both learning and salience; commuters' and survey results support this.

A08: What alternative mechanisms are discussed?

- Category: `A) Research question / contribution`
- Confidence: `low (0.22573341141410347)`
- Retrieval method: `local`
- Evidence type: `theoretical discussion of mechanisms`
- Answer:

Learning effect and salience effect are discussed as alternative mechanisms; they may act alone or in combination.

- Quote snippet:

  > One reason why calorie posting may affect consumer choice is a learning effect.

- Citation anchors:
  - page=25, words=0-349, section=A. Why is There an Effect?, note=Identifies learning effect and salience effect as alternative mechanisms; notes possible combination.

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

- Category: `A) Research question / contribution`
- Confidence: `medium (0.37642094063911)`
- Retrieval method: `local`
- Evidence type: `policy_implications`
- Data source: Starbucks NYC transaction data; control cities (Boston/Philadelphia); cardholder data; surveys.
- Assumption flag: `True`
- Assumption notes: Assumes similar 6% reduction across all chain restaurants would translate into obesity reductions; uses back-of-the-envelope estimate in Section IVB.
- Answer:

Mandatory calorie posting reduces average calories per transaction (about 6%), mainly by lowering food calories with little effect on beverage calories or visit frequency; profits are largely unaffected on average. The costs of posting are low, and the policy may yield modest public health benefits if extended to all chain restaurants, plus potential long-run effects through menu innovation toward lower-calorie options and public education benefits. The results also imply that voluntary posting by individual chains would likely yield different outcomes.

- 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; Conclusions, note=Summary of policy implications: mandatory calorie posting reduces calories per transaction, low costs, potential for innovation, public education benefits.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.3100922381795354)`
- Retrieval method: `local`
- Evidence type: `Empirical results and discussion from the Starbucks calorie-posting study, covering magnitude, margins (extensive vs intensive), heterogeneity, mechanisms (learning vs salience), and policy implications.`
- Data source: Starbucks transaction and cardholder data from NYC and control cities used to study mandatory calorie posting effects.
- Answer:

The welfare interpretation is that mandatory calorie posting appears to yield a health-related welfare gain by reducing calories per transaction (especially food calories) with little impact on how often people shop, implying the policy changes composition of purchases rather than reducing access or shopping intensity. Three quarters of the total calorie reduction stems from not purchasing a food item (extensive margin) and the rest from switching to lower-calorie items (intensive margin), with larger reductions among high-calorie purchasers. However, obesity outcomes are not measured, effects are heterogeneous (varying by income, education, gender, and purchase frequency), and there is mixed evidence on the mechanisms (learning vs. salience); thus the overall welfare impact is uncertain and depends on whether the calorie reductions translate into meaningful health gains at the population level.

- Quote snippet:

  If the policy goal is to address obesity, the most relevant question may be whether calorie posting disproportionately affects consumers

who make high-calorie purchases.

- Citation anchors:
  - page=20, words=0-349, section=C. Heterogeneity in the Effect of Mandatory Calorie Posting, note=No significant change in visit frequency; changes occur in purchase composition.
  - page=20, words=275-551, section=IV. Discussion, note=If the policy goal is to address obesity, the most relevant question ... discussed in the text about obesity-centered policy interpretation.
  - page=22, words=0-349, section=Table 6—Estimated Effects of Mandatory Calorie Posting at Various Quantiles, note=Calorie reductions tend to be larger in the top half of the distribution; welfare implications for high-calorie purchasers.
  - page=24, words=275-564, section=IV. Discussion, note=14% drop in food calories on average; most of this via the extensive margin (not buying food).
  - page=30, words=0-349, section=IV. Discussion, note=Survey evidence shows learning vs salience as mechanisms; mixed evidence on which drives behavior.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.27591253181064024)`
- Retrieval method: `local`
- Evidence type: `limitations described in the article`
- Data source: Starbucks transaction data; limited to a single chain (Starbucks); obesity effects not measured due to BMI data lag.
- Answer:

Two main limitations: (1) the study does not directly measure calorie posting's effect on obesity because BMI data are not available; (2) the analysis uses data from only one chain (Starbucks), preventing generalization to other chains and not accounting for potential offsetting changes at home.

- Quote snippet:

  There are two important limitations to this research. First, we do not directly measure the effect of calorie posting on obesity itself.

- Citation anchors:
  - page=3, words=275-564, section=Limitations, note=Two important limitations: obesity measurement not available; data from only Starbucks; offset by home consumption unknown.

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

- Category: `A) Research question / contribution`

- Confidence: `low (0.3097686871556125)`
- Retrieval method: `local`
- Evidence type: `Data novelty and identification strategy`
- Data source: Two datasets: transaction data for all Starbucks locations and a panel of 2.7 million anonymous cardholders
- Answer:

The paper claims novelty in its data and identification: it uses two rich data sources—a complete transaction dataset for all Starbucks stores and a large panel of anonymous cardholders—to isolate how calorie posting affects individuals' behavior, separating changes in the intensive margin from changes in consumer composition (i.e., within-individual variation).

- Quote snippet:

  > The Starbucks data we study is rich enough to allow us to distinguish these various responses, as we explain in the next section.

- Citation anchors:
  - page=3, words=275-564, section=Data novelty and identification, note=Contains the claim that detailed transaction data are uniquely well-suited to answering the question about posting effects.
  - page=5, words=0-349, section=Data sources and identification approach, note=Notes the two datasets (transaction data and cardholder data) and that the cardholder data allow isolating 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 (0.27854021384451677)`
- Retrieval method: `local`
- Evidence type: `Methods section description`
- Data source: Starbucks transaction and cardholder data
- Answer:

The identification strategy relies on within-city variation over time, comparing NYC calories-posting data before and after the policy change to control cities (Boston/Philadelphia) using a fixed-effects regression to recover the causal effect.

- Quote snippet:

  > Identification of the effect of calorie posting stems from within-city variation over time.

- Citation anchors:
  - page=11, words=275-522, section=Identification strategy, note=Within-city variation over time used to identify the effect

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3050731891068924)`
- Retrieval method: `local`
- Evidence type: `quasi-experimental (difference-in-differences using treatment and control cities)`
- Data source: transaction data and cardholder data from Starbucks
- Answer:

Quasi-experimental (difference-in-differences) design.

- 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=II. Effect of Mandatory Calorie Posting on Calorie Consumption - A. Calories Per Transaction, note=Regression specification with a posting dummy (POS_t) to compare before/after effects
  - page=7, words=275-566, section=III. Seattle control and NYC treatment, note=Describes use of Seattle and control cities to compare effects; indicates natural experiment framework

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.29850120830444815)`
- Retrieval method: `local`
- Evidence type: `explanation of identification strategy (difference-in-differences using NYC as treated city with control cities; within-city, city-week variation)`
- Data source: Starbucks transaction data and cardholder data from NYC, Boston, Philadelphia (Seattle for robustness)
- Answer:

The exogenous variation comes from the timing of calorie posting in NYC (April 1, 2008), with NYC as the treated city and Boston/Philadelphia as controls, exploiting within-city (city-week) variation over time for identification.

- Quote snippet:

  identification of the effect of calorie posting stems from within-city variation over time.

- Citation anchors:
  - page=11, words=275-522, section=Identification/Methodology, note=Stems from within-city variation over time; policy variation at city-week level.
  - page=12, words=0-349, section=Identification, note=Policy variation for identification is at the city-week level.

B04: What is the treatment definition and timing?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.13193543900923896)`
- Retrieval method: `local`
- Evidence type: `textual evidence from methods describing the treatment and timing`
- Data source: Starbucks transaction and cardholder data for NYC, Boston, and Philadelphia
- Answer:

Treatment: the NYC mandatory calorie posting policy applied to New York City stores; Timing: effective after April 1, 2008.

- 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=0-349, section=Methods/Empirical strategy, note=Definition of treatment: POST equals one for NYC stores after April 1, 2008.

B05: What is the control/comparison group definition?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.18001451699152107)`
- Retrieval method: `local`
- Evidence type: `textual evidence from the data/methods description`
- Data source: NYC transaction data with Boston and Philadelphia as control cities; cardholder data used for robustness checks; Seattle analysis uses Portland and San Francisco as additional controls.
- Assumption flag: `True`
- Assumption notes: Assumes Boston and Philadelphia are valid controls for NYC; differences may reflect unobserved heterogeneity, not solely calorie posting effects.
- Answer:

The primary control/comparison group is the set of stores/cardholders in Boston and Philadelphia (non-NYC cities), used as the control for NYC after calorie

posting. The Seattle analysis uses Portland and San Francisco as additional control cities.

- Quote snippet:

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

- Citation anchors:
  - page=6, words=0-349, section=Control group definition, note=Statement that Boston and Philadelphia serve as reasonable controls for NYC.
  - page=14, words=0-590, section=Seattle control cities, note=Seattle analysis uses Portland and San Francisco as control cities.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3085944213567219)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction and cardholder data used in Bollinger et al. (2011)
- Table/Figure: Table 3
- Answer:

Baseline estimating equation: y_sct = x_sct + POST_Tc t + _sct, where 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 (i.e., NYC stores after April 1, 2008), and x_sct includes week fixed effects (seasonality), day-of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared. In the transaction data, store fixed effects are included; in the cardholder data, individual fixed effects are included.

- Quote snippet:

  y_sct = x_sct + POST_Tc t + _sct, where 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 (i.e., NYC stores after April 1, 2008)

- Citation anchors:
  - page=11, words=275-522, section=Estimating equation (1) baseline regression, note=Equation (1) specification: y_sct = x_sct + POST_Tc t + _sct; POST is NYC post-April 1, 2008; controls described

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.25441762026853965)`
- Retrieval method: `local`
- Evidence type: `methodology description from regression specification`
- Data source: American Economic Journal: economic policy february 2011
- Answer:

Unit fixed effects: store fixed effects (transaction data) and individual fixed effects (cardholder data). Time fixed effects: week fixed effects (to control for seasonality). Additional time controls: day-of-week dummies and holiday dummies. Weather controls (temperature and precipitation and their squares) accompany the time controls. In robustness checks, date fixed effects are used instead of week fixed effects.

- 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. Store fixed effects are used for transaction data; cardholder data use individual fixed effects.

- Citation anchors:
  - page=page 11, words=275-522, section=methodology, note=week fixed effects, day-of-week dummies, holiday dummies, temperature and precipitation controls; cardholder vs transaction specification mentioned
  - page=page 12, words=0-349, section=methodology, note=store fixed effects described; cardholder data use individual fixed effects
  - page=page 12, words=275-624, section=methodology, note=identification from within-city variation; comparison of transaction and cardholder fixed effects

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2809089960125438)`
- Retrieval method: `local`
- Evidence type: `unspecified`
- Answer:

{ "answer": "Robust standard errors are used; the text notes specifically " + ""Robust standard errors in parentheses"; the level of clustering is not specified in the provided context.", "evidence_type": "Direct note from table 9 notes", "confidence": "high", "citation_anchors": [{ "page": "page 29", "start_word":

0, ”end_word”: 349, ”section”: ”Notes”, ”note”: ”Robust standard errors in parentheses.” }], ”quote_snippet”: ”Robust standard errors in parentheses.”, ”table_figure”: null, ”data_source”: null, ”assumption_flag”: null, ”assumption_notes”: null, ”related_questions”: [] }

- Quote snippet:

  combined 1.51 million transactions for us to study. We refer to the first dataset as the transaction data and the second dataset as the cardholder data. The advantage of the cardholder data is that we

- Citation anchors:
  - page=5, words=275-591
  - page=6, words=0-349
  - page=11, words=275-522
  - page=12, words=0-349
  - page=12, words=275-624
  - page=12, words=550-630
  - page=15, words=0-254
  - page=22, words=0-349
  - page=26, words=0-178
  - page=29, words=0-349

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2723139027820027)`
- Retrieval method: `local`
- Evidence type: `Identification assumption (parallel trends)`
- Data source: Cardholder and transaction data; NYC vs Boston/Philadelphia comparison around Apr 1, 2008
- Assumption flag: `True`
- Assumption notes: Key identifying assumption is parallel trends: in the absence of NYC calorie posting, NYC and control cities would have followed similar trends; pre-treatment trends show no difference.
- Answer:

Parallel trends assumption (difference-in-differences): in the absence of calorie posting in NYC, NYC would have followed the same trend as the control cities (Boston/Philadelphia). The authors provide evidence of no pre-trend differences and immediate post-treatment changes around April 1, 2008.

- Quote snippet:

  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=Identification strategy / pre-trend evidence, note=No evidence of pre-trend differences between NYC and Boston/Philadelphia; post-treatment changes align with calorie posting.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2404398310174015)`
- Retrieval method: `local`
- Evidence type: `parallel trends / pre-treatment trend similarity between NYC and control cities`
- Data source: transaction data and cardholder data
- Table/Figure: Figure 2
- Assumption flag: `True`
- Assumption notes: Evidence shows pre-treatment trends are parallel between NYC and control cities, supporting the identifying assumption.
- Answer:

The identifying assumption is that, in the absence of the calorie-posting treatment, NYC would have followed similar trends to the control cities (Boston/Philadelphia). The evidence provided is that there is no pre-treatment difference in trends between NYC and the controls, observed in both the transaction data and the cardholder data (Figure 2).

- Quote snippet:

  no evidence of pre-trend differences between NYC and Boston/Philadelphia.

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

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3177279378982099)`
- Retrieval method: `local`
- Evidence type: `Event-study / pre-trend analysis`
- Data source: Starbucks transaction and cardholder data across Boston, NYC, and Philadelphia
- Table/Figure: Figure 2
- Answer:

Yes. The study conducts an event-study–style test using weekly fixed effects and a post-posting indicator, and finds no pre-trend differences between NYC and

control cities; calories per transaction drop right after April 1, 2008 and persist through Feb 2009 (with some holiday-period deviations in the transaction data but not in cardholder data).

- Quote snippet:

  > First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia. Second, in both cases it is clear that the drop in calories per transaction occurred right around April 1, 2008, and persisted through February 2009.

- Citation anchors:
  - page=13, words=275-528, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=No evidence of pre-trend differences; post-treatment drop around posting date; discussion of Figure 2.
  - page=14, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Time path shows reduction occurred immediately following posting; persists.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.09974872819763983)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: American Economic Journal: Economic Policy (Bollinger et al., 2011)
- Assumption flag: `False`
- Assumption notes: No instrumental variable is used; identification relies on a policy-change (calorie posting) with city-time variation and fixed effects (diff-in-diff style). The POST variable is a treatment indicator, not an instrument requiring a first stage.
- Answer:

There are no instrumental variables used in this analysis. The study uses a difference-in-differences design around the calorie posting policy, treating NYC stores after April 1, 2008 as the treated group and using separate week effects and weather controls; the POST variable is a simple treatment dummy, not an instrument, so there is no first stage.

- Quote snippet:

  > The results shown in Table 3 are based on specifications in which calorie posting is binary—i.e., the POST variable is simply a dummy equal to one at NYC stores on every day after April 1, 2008.

- Citation anchors:

  - page=13, words=275-528, section=Calorie posting regression specification, note=POST variable is simply a dummy equal to one at NYC stores on every day after April 1, 2008; indicates a treatment indicator rather than an instrument.
  - page=15, words=0-254, section=Figure 2 regression setup, note=Regression uses separate week effects for NYC and control regions, day-of-week effects, and weather controls, illustrating a diff-in-diff style design rather than IV.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2451413327313195)`
- Retrieval method: `local`
- Evidence type: `No first-stage analysis or instrument strength (F-stat, partial R^2) is discussed in the provided sections.`
- Data source: Starbucks transaction and cardholder data
- Answer:

Not reported in the provided context.

- Quote snippet:

  The estimated effect on beverage calories is not significantly different from zero.

- Citation anchors:
  - page=3, words=275-564
  - page=11, words=275-522
  - page=12, words=275-624
  - page=12, words=550-630
  - page=13, words=0-349
  - page=18, words=0-349
  - page=18, words=275-541
  - page=23, words=275-539
  - page=25, words=275-599
  - page=30, words=0-349

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.22352108386975664)`
- Retrieval method: `local`
- Evidence type: `regression specification / study design description`
- Data source: Starbucks transaction and cardholder data for NYC, Boston, and Philadelphia

- Answer:

The study does not use a regression discontinuity design (RDD). It employs a difference-in-differences-like approach with a post-treatment indicator for NYC after April 1, 2008, and does not define a running variable, cutoff, or bandwidth as in an RDD.

- Quote snippet:

  POS Tc 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 dummy (calories posted) and NYC after April 1, 2008
  - page=29, words=275-442, section=Results (Table 9) / post-01April08 interactions, note=Use of post-01April08 dummy in analyses

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2889113158658434)`
- Retrieval method: `local`
- Evidence type: `textual description of the estimator used for staggered adoption in the study's methodology`
- Data source: Transaction data and cardholder data (Starbucks calorie posting study across NYC and control cities)
- Answer:

A difference-in-differences estimator with timing variation implemented by including separate weekly dummies for NYC and the control cities (i.e., a staggered-adoption/event-study style DiD), excluding the POST variable in that specification.

- 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=13, words=An-variable., section=Estimation approach, note=Describes modifying equation (1) to include city-specific weekly dummies to capture timing differences (staggered adoption).

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

- Category: `C) Data, sample, and measurement`

- Confidence: `low (0.2803742622042457)`
- Retrieval method: `local`
- Evidence type: `dataset description`
- Data source: Starbucks transaction data (NYC, Boston, Philadelphia) and anonymous Starbucks cardholder data (US); Seattle in-store survey data for robustness checks.
- Answer:

Transaction data from all Starbucks company-owned stores in NYC (222) and Boston/Philadelphia (94); cardholder data from anonymous Starbucks cardholders across the United States; and in-store Seattle/San Francisco surveys used for robustness checks.

- 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=Intro to datasets and scope of transaction data
  - page=5, words=275-591, section=Data Sources, note=Description of cardholder data and transaction data; sample sizes
  - page=6, words=0-349, section=Data Measures/Overview, note=Table 1 and data characteristics; distinction between datasets
  - page=25, words=275-590, section=Seattle Surveys, note=In-store customer surveys in Seattle and San Francisco
  - page=36, words=0-349, section=Appendix, note=Survey form for Starbucks Customers

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3227127802070161)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Calorie Posting in Chain Restaurants, American Economic Journal: Economic Policy, Vol. 3 No. 1 (2011)
- 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 / Data structure, note=Statement that for each transaction we observe time, date, store location, items purchased, and price of each item.

C03: What is the sample period and geographic coverage?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.23080458900578665)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Starbucks transaction data for NYC, Boston, and Philadelphia; cardholder data used for comparison
- Answer:

Sample period: January 1, 2008 to February 28, 2009; Geographic coverage: Starbucks locations in New York City, Boston, and Philadelphia (NYC has 222 stores; Boston and Philadelphia together have 94 stores).

- 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=A. Data Summary, note=Contains description of sample period and geographic coverage (NYC, Boston, Philadelphia).

C04: What are the sample restrictions / inclusion criteria?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.24518952823710455)`
- Retrieval method: `local`
- Evidence type: `Data description / sample restrictions (Methods)`
- Data source: Starbucks transaction data (NYC, Boston, Philadelphia) and anonymous cardholder data
- Answer:

Sample restrictions / inclusion criteria include two data sets: (1) transaction data covering all Starbucks locations in NYC (222 stores) and Boston/Philadelphia (94 stores) with observations for 3 months before and 11 months after calorie posting (Jan 1, 2008–Feb 28, 2009); and (2) anonymous cardholder data (2.7 million individuals) with a subsample defined as those who averaged at least one transaction per week in NYC, Boston, or Philadelphia in the pre-post NYC period. The cardholder subsample consists of 7,520 NYC individuals and 3,772 individuals in Boston/Philadelphia (1.51 million transactions in total). Exclusions included: stores not open for the entire period (balanced panel) and transactions with more than four units of any one

item (about 2.2% of transactions). Cardholders may not be representative of Starbucks customers in general.

- 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, note=Describes transaction and cardholder data, and introduces the subsample criteria.
  - page=5, words=275-591, section=Data Summary, note=Details the cardholder subsample numbers and representativeness caveat.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.29790605406550574)`
- Retrieval method: `local`
- Evidence type: `Data Summary: cardholder subsample with combined transactions`
- Data source: Starbucks transaction and cardholder data for NYC, Boston, and Philadelphia
- Answer:

1.51 million transactions

- Quote snippet:

  > There are 7,520 such individuals in NYC and 3,772 such individuals in Boston and Philadelphia, generating a combined 1.51 million transactions for us to study.

- Citation anchors:
  - page=5, words=0-349, section=Data Summary, note=Mentions 2.7 million anonymous individuals in the dataset.
  - page=5, words=275-591, section=Data Summary / Cardholder data, note=States 7,520 NYC cardholders and 3,772 others generating a combined 1.51 million transactions.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.32668180401971075)`
- Retrieval method: `local`

- Evidence type: `Direct references to outcome definitions and transformations from the regression specifications and results.`
- Data source: Transaction data and cardholder data from Starbucks study
- Answer:

The key outcome is calories per transaction (e.g., calories per transaction, beverage calories per transaction, and food calories per transaction). The analysis uses both raw calories and log-transformed forms. In particular, regressions model y_sct, a measure of calories per transaction, and there are explicit uses of log transformations (e.g., log(calories per item purchased) as a dependent variable; and in quantile analyses, regressions use the log of the quantile as the dependent variable). The percent changes cited (e.g., a 11.2% decrease) come from log-based coefficients (e.g., a coefficient of -0.119).

- Quote snippet:

  The estimated effect for food calories per transaction is a 11.2 percent decrease (based on the coefficient estimate of -0.119).

- Citation anchors:
  - page=11, words=275-522, section=Regression specification/outcome variable, note=y_sct is a measure of calories per transaction.
  - page=12, words=275-624, section=Results: cardholder data; percentage change, note=The food calories per transaction shows an 11.2% decrease with coefficient -0.119.
  - page=16, words=275-558, section=Transformations; dependent variables, note=Includes log(calories per item purchased) as a dependent variable.
  - page=22, words=0-349, section=Quantile analysis; dependent variable, note=Regressions using the log of the quantile as the dependent variable.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2922025287010622)`
- Retrieval method: `local`
- Evidence type: `methodology description`
- Data source: transaction data and cardholder data
- Answer:

Exposure to calorie posting is measured with a treatment indicator (POST) that equals 1 for NYC stores on days after calories are posted (April 1, 2008). The analysis uses within-city variation (store-day for the transaction data; individual fixed effects for the cardholder data). Constructed variables include POST as the exposure, as well as outcome measures such as log(beverage calories

per transaction), log(food calories per transaction), and log(beverage + food calories); other items-per-transaction metrics are also created.

- 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=page 11, words=275-522, section=Regression specification; POST dummy, note=Definition of POST and identification via within-city variation.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.28721785956610646)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Starbucks transaction data and cardholder data (pre/post-calorie posting), as described in the study.
- Answer:

Yes. The analyses consistently include week fixed effects to control for seasonality, day-of-week dummies, holiday dummies, and weather controls (temperature and precipitation, with squared terms). In the cardholder-data specification, 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=page 11, words=0-349, section=Methods/Regression specification, note=Controls include week fixed effects, day-of-week dummies, holiday dummies, temperature and temperature squared, precipitation and precipitation squared.
  - page=page 29, words=0-349, section=Table 9 notes, note=Cardholder-data regressions include individual, week, and day-of-week fixed effects, and weather controls.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.27821695977620237)`
- Retrieval method: `local`

- Evidence type: `Data cleaning / exclusion criteria`
- Data source: Transaction data and cardholder data
- Answer:

Missing data are handled by dropping non-eligible observations rather than imputing values; specifically, transactions from stores not open during the full data period are dropped to form a balanced panel, and transactions with more than four units of any one item are dropped. No imputation or weighting is described.

- 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).

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=Excludes stores not open for entire period (balanced panel) and excludes transactions with >4 units of any item; 2.2% excluded.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.29563479335392373)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction and cardholder data; regression analysis in Bollinger et al., American Economic Journal: Economic Policy (2011).
- Answer:

No weights (survey weights or propensity weights) are reported. The analysis uses regression with fixed effects (e.g., week fixed effects to capture seasonality) and various controls, based on two data sources (transaction data and cardholder data) and with aggregation to address serial correlation.

- Quote snippet:

  > y_sct = x_sct + _POS Tc t + _sct, where y_sct is a measure of calories per transaction; POS Tc t is a dummy equal to one if calories were posted; week fixed effects control for seasonality (plus weather controls).

- Citation anchors:
  - page=11, words=0-349, section=Regression specification, note=Shows regression form: y_sct = x_sct + _POS Tc t + _sct; defines y_sct and POS dummy.

  - page=11, words=275-522, section=Model controls, note=Notes week fixed effects and weather controls to address seasonality and other factors.
  - page=12, words=275-624, section=Robustness / aggregation, note=Describes aggregating data to address serial correlation and testing differences before vs after posting.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.22259982351998478)`
- Retrieval method: `local`
- Evidence type: `data linkage description`
- Data source: Starbucks transaction data; anonymous Starbucks cardholder data
- Answer:

Yes. The study links two data sources—the transaction data and anonymous Starbucks cardholder data—by analyzing the same individuals across both datasets. Cardholder data track purchases for individuals and are compared with the transaction data to assess changes in behavior; the analysis focuses on subsamples of cardholders with NYC/Boston/Philadelphia activity and then compares effects across datasets.

- Quote snippet:

  > In addition to the transaction data we have data for a sample of anonymous Starbucks cardholders, tracking their purchases over the same period of time all over the United States. There are 2.7 million anonymous individuals in this dataset.

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=Describes transaction data and cardholder data; 2.7 million anonymous individuals; subsample definition; cross-dataset analysis.
  - page=5, words=275-591, section=A. Data Summary, note=States 'The advantage of the cardholder data is that we can assess how ... isolated effects' and 'we compare ... cardholder data with transaction data'.
  - page=3, words=275-564, section=Main text, note=Discusses data sources (transaction and cardholder data) and empirical approach to calorie posting.

C12: What summary statistics are reported for main variables?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3270992716107185)`

- Retrieval method: `local`
- Evidence type: `table`
- Data source: Transaction data and Cardholder data (prior to calorie posting in NYC)
- Answer:

Table 1 reports the following summary statistics for the Transaction Data and Cardholder Data (prior to calorie posting): average weekly transactions per store; average weekly revenue per store; percent transactions with brewed coffee; percent transactions with beverage; percent transactions with food; average number of items per transaction; average number of drink items per transaction; average number of food items per transaction; food attach rate; average dollars per transaction; average calories per transaction; average drink calories per transaction; and average food calories per transaction (normalized to 1.00).

- Quote snippet:

  Table 1—Summary Statistics for Transaction Data and Cardholder Data (Prior to policy change)

- Citation anchors:
  - page=6, words=0-349, section=Table 1—Summary Statistics for Transaction Data and Cardholder Data (Prior to policy change), note=Contains the list of summary statistics for both transaction and cardholder datasets prior to calorie posting.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2833973846762995)`
- Retrieval method: `local`
- Evidence type: `descriptive figures and table (Figure 1, Figure 2, Table 1)`
- Data source: Starbucks transaction data and cardholder data for NYC, Boston, and Philadelphia
- Table/Figure: Figure 1; Figure 2; Table 1
- Answer:

Yes. The paper includes descriptive figures and a table that establish baseline patterns: Figure 1 shows average calories per transaction by week (NYC vs control cities), Figure 2 shows cross-dataset results with no pre-trend differences, and Table 1 provides summary statistics prior to the policy change.

- 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, note=Figure 1 described as showing average calories per transaction by week, NYC vs controls.
  - page=14, words=0-349, section=II, note=Figure 2 depicts results for each dataset; shows no pre-trend differences.
  - page=6, words=0-349, section=Table 1, note=Table 1 provides summary statistics prior to policy change.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.4277065804689114)`
- Retrieval method: `local`
- Evidence type: `Table 3 results showing effect of calorie posting on calories per transaction`
- Data source: American Economic Journal: Economic Policy, February 2011
- Table/Figure: Table 3
- Answer:

5.8 percent decrease in average calories per transaction (approximately 14.4 calories per transaction).

- Quote snippet:

  > the bottom row, we report the estimated impact on log(beverage + food calories), finding a 5.8 percent decrease in average calories per transaction, equivalent to 14.4 calories.

- Citation anchors:
  - page=page 12, words=0-349, section=Table 3 results; effect on calories per transaction, note=Main headline effect: 5.8% decrease in average calories per transaction (14.4 calories).

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.18141563527717403)`
- Retrieval method: `local`
- Evidence type: `Regression specification description and justification`
- Data source: Starbucks transaction data and cardholder data
- Answer:

The preferred specification is the regression y_sct = x_sct + _POS Tc t + _sct, where y_sct is a measure of calories per transaction, and x_sct includes week fixed effects, day-of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared; estimated separately for transaction data and cardholder data. This specification is pre-

ferred because it controls for seasonality and weather, enabling isolation of the effect of calorie posting, with identification drawn from within-week variation.

- 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 11, words=0-349, section=Regression specification, note=Describes the regression form and the variables included in x_sct.
  - page=page 11, words=275-522, section=Weather/Identification discussion, note=Notes weather controls and week fixed effects; discusses identification from within-week variation.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.45872401362016013)`
- Retrieval method: `local`
- Evidence type: `Empirical estimates from the Starbucks calorie posting study (transaction and cardholder data) reported in Tables 3-6 and the discussion.`
- Data source: NYC Starbucks calorie posting study; transaction and cardholder data
- Table/Figure: Table 3; Table 4; Table 6
- Answer:

The effect is economically meaningful chiefly for food calories, with reductions around 14% per transaction for food calories (11.2% in cardholder data) and about a 5–6% overall decrease in calories per transaction. Beverage calories change negligibly (about a 0.3% decrease). The study also finds prices rose on average, but Starbucks profits were not significantly affected. A back-of-the-envelope calculation is discussed for broader adoption (6% reduction across all chain restaurants), but there is no direct dollar impact estimated in the data.

- Quote snippet:

  the bottom row, we report the estimated impact on log(beverage + food calories), finding a 5.8 percent decrease in average calories per transaction, equivalent to 14.4 calories.

- Citation anchors:
  - page=12, words=0-349, section=Results — calorie effects (Table 3), note=Food calories per transaction down ~13.7%; beverage calories

~0.3%; total calories ~5.8% (14.4 calories).
  - page=12, words=275-624, section=Cardholder data results, note=Cardholder data show beverage calories not significantly different; food calories per transaction down ~11.2%; total calories ~5.0%.
  - page=18, words=275-541, section=Substitution and margins, note=Nearly three quarters of calorie reduction due to extensive margin; percent reduction ~5–6% across high-calorie items.
  - page=24, words=275-564, section=Profit and revenue discussion, note=Prices per item rose; profits largely unchanged (no significant effect on Starbucks profit).
  - page=30, words=0-349, section=Is the effect big enough to matter?, note=Crude calculation discusses a 6% reduction if applied across all chain restaurants; explores potential implications.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2836016351745133)`
- Retrieval method: `local`
- Evidence type: `Robustness checks`
- Data source: American Economic Journal: Economic Policy (February 2011) article
- Answer:

Key robustness checks and outcomes: (1) Date fixed effects robustness check; results barely changed (transaction calories: beverage −0.004, food −0.152, total −0.063). (2) Alternative dependent variable robustness check; using absolute calories yields almost identical results to the baseline. (3) Serial-correlation robustness check by aggregating pre/post calorie posting data; still finds approximately the same effect. (4) Cross-check for pre-trend differences; no evidence of pre-trend differences between NYC and control cities, supporting identification. Overall, the main findings—that calorie posting reduces calories per transaction and that effects persist across specifications—are qualitatively robust across these checks.

- Quote snippet:

  > As a robustness check, we include date fixed effects in the above specification ... The estimates based on the transaction data are barely changed.

- Citation anchors:
  - page=12, words=As-effects, section=Robustness checks, note=Date fixed-effects robustness check; results barely changed.
  - page=12, words=We-identical, section=Robustness checks, note=Alternative dependent variable robustness check; using

absolute calories yields almost identical results.
  - page=12, words=To-effect., section=Robustness checks, note=Serial-correlation robustness check via pre/post aggregation; results similar.
  - page=14, words=First,-differences., section=Robustness checks, note=No pre-trend differences between NYC and controls; robustness of identification.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2122229744352862)`
- Retrieval method: `local`
- Evidence type: `placebo tests: pre-trend check and Seattle placebo test`
- Data source: Starbucks cardholder and transaction data; NYC and control cities (Boston, Philadelphia; Seattle data referenced).
- Assumption flag: `True`
- Assumption notes: No pre-trend differences between NYC and control cities prior to posting; supports parallel trends assumption for placebo tests.
- Answer:

The authors perform placebo-like checks: (1) a pre-trend test showing no evidence of pre-existing differences between NYC and the control cities (Boston/Philadelphia) prior to posting, and (2) a Seattle placebo test where regression analysis shows no significant impact of calorie posting on calories per transaction, supporting that observed NYC effects are due to posting rather than common trends.

- Quote snippet:

  no evidence of pre-trend differences between NYC and Boston/Philadelphia; Seattle analysis shows no significant impact of calorie posting.

- Citation anchors:
  - page=14, words=0-349, section=Figure 2 discussion / pre-trend check, note=No evidence of pre-trend differences between NYC and Boston/Philadelphia.
  - page=29, words=0-349, section=Seattle placebo test / Section III note, note=Seattle (and controls) show no significant impact of calorie posting.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.25703417393733546)`
- Retrieval method: `local`
- Evidence type: `not explicitly stated in the provided context`

- Data source: Starbucks transaction data and cardholder data
- Answer:

No falsification or unaffected outcomes are described 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=11, words=0-349, section=Methods/Regression specification, note=Describes estimating regressions to examine effects of calorie posting; no mention of falsification or unaffected outcomes.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.3746442020132088)`
- Retrieval method: `local`
- Evidence type: `Table 5 (heterogeneity by income, education, gender, baseline exposure) and text describing NYC vs control regions; Table 9 (regional/commuter effects); Table 2 (size-related changes in beverage choices).`
- Data source: Transaction data; Cardholder data
- Table/Figure: Table 5; Table 9; Table 2
- Answer:

Heterogeneity results show: (i) by income and education: larger reductions in calories per transaction in higher-income ZIPs and in areas with more residents with college degrees; (ii) by baseline exposure: high-calorie and low-frequency (baseline exposure) customers exhibit larger reductions, with high-calorie customers averaging >250 calories per transaction showing about a 26% drop; (iii) by region/baseline location: calorie posting reduced beverage choices more in NYC but did not significantly change the frequency of cardholder purchases relative to control cities; (iv) by size (beverage size): shifts occurred across smaller, same, and larger sizes alongside calorie per ounce changes, indicating size-related variation in responses.

- Quote snippet:

  for those who averaged more than 250 calories per transaction, calories per transaction fell by 26 percent.

- Citation anchors:
  - page=21, words=Table 5—Heterogeneity-Notes, section=Heterogeneity in the Impact of Calorie Posting on log (Calories per transaction),

note=Income, education, gender, baseline exposure interactions reported.
  - page=21, words=Posting × median income-0.006), section=Heterogeneity in the Impact of Calorie Posting on log (Calories per transaction), note=Income interaction; negative coefficient indicating larger reductions in higher-income areas.
  - page=21, words=Posting × percent with college degree-0.010), section=Heterogeneity in the Impact of Calorie Posting on log (Calories per transaction), note=Education interaction; negative coefficient.
  - page=21, words=for those who averaged more than 250 calories per transaction-26 percent, section=Heterogeneity in the Impact of Calorie Posting on log (Calories per transaction), note=Baseline exposure finding: high-calorie consumers show larger reductions.
  - page=20, words=The milk order data includes Boston and Philadelphia-usage., section=Data sources and regional variation, note=Control cities used to identify region/seasonality effects.
  - page=20, words=We found that female cardholders were more responsive-males., section=Heterogeneity in the Impact (gender), note=Gender interaction observed in cardholder data.
  - page=29, words=Table 9—Effects of Calorie Posting on Commuters' Log-Notes:, section=Regional heterogeneity (region/commuters), note=Commuter vs non-commuter effects described.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.14262418580547825)`
- Retrieval method: `local`
- Evidence type: `Mechanism tests of salience vs. learning; margin analysis (extensive vs intensive); substitution effects; commuter vs non-commuter analysis; survey evidence on knowledge; quantile analysis of calories per transaction.`
- Data source: Starbucks calorie-posting study using transaction data and cardholder data across NYC, Seattle, and control cities (Boston and Philadelphia).
- Answer:

Tests differentiate mechanisms by (i) comparing effects inside NYC posting zones to non-posting contexts and between commuters and non-commuters, (ii) analyzing substitution versus consumption frequency to separate intensive and extensive margins, and (iii) using survey data and quantile analyses to assess how calories are perceived and which purchasers are affected. Overall, the results support a learning mechanism (calorie information affects choices beyond posting sites and among commuters) rather than a salience-only mechanism, and show that reductions come mainly from the extensive margin (fewer food items) with additional substitution toward lower-calorie options and some mod-

est changes in calories per item.

- 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=275-564, section=Learning vs salience mechanism testing, note=States that learning effects are observed beyond posting locations and are inconsistent with pure salience effects.
  - page=7, words=275-566, section=Survey-based mechanism evidence, note=Describes in-store surveys before/after posting to gauge knowledge and impact of posting on awareness.
  - page=11, words=0-349, section=Baseline changes and regression specification, note=Introduces the regression framework to control for seasonality and identify posting effects on calories per transaction.
  - page=16, words=0-349, section=Substitution effects and margins, note=Analyzes substitution, intensive vs extensive margins, and item-level sales to identify mechanism of calorie posting impact.
  - page=21, words=275-421, section=Quantile analysis of calories per transaction, note=Examines effects at different quantiles of calories per transaction to see differential impacts across purchasers.
  - page=29, words=275-442, section=Commuter vs non-commuter interaction results, note=Presents the interaction results showing commuter and non-commuter effects in NYC vs outside NYC.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.35924071902791654)`
- Retrieval method: `local`
- Evidence type: `Robustness checks across samples, time bandwidths, and controls (weather, serial correlation) demonstrating consistent effects.`
- Data source: Transaction data and cardholder data
- Table/Figure: Figure 2; Table 3
- Answer:

The results are robust to alternative samples, bandwidths, and controls. The decline in calories per transaction occurs immediately after calorie posting (around April 1, 2008) with no pre-trend differences between NYC and control cities; findings hold across transaction data and cardholder data, and persist for up to 11 months. Robustness checks show that weather controls have little effect, and excluding them leaves results unchanged; a conservative pre/post aggregation test for serial correlation yields essentially the same effect; within-individual

variation in cardholder data confirms that the observed change is primarily an intensive-margin effect rather than composition changes.

- Quote snippet:

  Our findings are unchanged if we exclude the weather controls.

- Citation anchors:
  - page=page 14, words=0-349, section=Robustness: pre-trends and timing, note=No pre-trend differences between NYC and controls; timing around Apr 1, 2008
  - page=page 12, words=0-274, section=Internal validity: cardholder data identification, note=Cardholder data isolates intensive margin; within-individual variation; similar estimates
  - page=page 12, words=275-624, section=Robustness: serial correlation checks, note=Conservative aggregation pre/post yields same effect; significance remains
  - page=page 11, words=0-349, section=Robustness: weather controls, note=Weather controls included; results unchanged if weather controls are excluded

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2696840155152815)`
- Retrieval method: `local`
- Evidence type: `Conclusions section of the study (summary statements about effects, mechanisms, and implications).`
- Data source: Starbucks transaction data; NYC calorie posting policy study (includes commuter and survey analyses).
- Answer:
- Mandatory calorie posting reduces average calories per transaction by 6% at Starbucks (long-lasting).
- The effect is driven primarily by changes in food choices rather than beverages.
- The impact is larger for high-calorie consumers.
- Both learning and salience contribute to the observed reductions; survey data show ignorance about calories but increased sensitivity, and commuter behavior suggests learning effects in non-posting stores.
- Starbucks profits are largely unchanged on average; near competitors' stores, revenue may even rise.

- Policy implications: low costs with potential for further reductions (and greater long-run effects if chains offer more low-calorie options); results may differ for other chains.
- 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=Conclusion, note=Concluding statements summarizing main findings (6% drop, long-lasting, beverage vs food, learning and salience, profits, policy implications).
  - page=28, words=275-556, section=Discussion/Conclusion, note=Survey results indicating salience vs learning mechanism; supports dual explanation.
  - page=30, words=0-349, section=Commuters analysis, note=Commuter results show learning and salience with mixed evidence; supports dual mechanism.
  - page=3, words=275-564, section=Intro to mechanisms, note=Discussion of salience vs learning effects and limitations; supports framing in conclusions.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.27332796759951955)`
- Retrieval method: `local`
- Evidence type: `Literature review / empirical and theoretical prior work on nutrition information, labeling, and consumer response.`
- Answer:

The most important prior papers are those that (a) document how calorie/nutrition information affects actual purchasing and weight outcomes, and (b) lay out the information/labeling literature and its implications for consumer choice. Key references include Bassett et al. 2008 on NYC calorie information and purchasing behavior, Currie et al. 2010 on the effect of fast-food exposure on obesity, foundational work on information processing and labeling by Jacoby et al. (1977) and Ippolito & Mathios (1990, 1995), and papers on strategies and framing of nutrition information by Downs, Loewenstein, and Wisdom (2009) and Kiesel & Villas-Boas (2008). These studies are central because they provide empirical benchmarks for labeling effects, theoretical underpinnings on information effects, and context for evaluating calorie posting in chain restaurants like Starbucks.

- Quote snippet:

Currie et al. (2010) find that teenagers whose schools are located within 0.1 miles of a fast food chain have significantly higher obesity rates.

- Citation anchors:
  - page=37, words=1-18, section=References, note=Foundational studies on nutrition information and labeling (Jacoby 1977; Ippolito & Mathios 1990, 1995).
  - page=25, words=1-10, section=References, note=Bassett et al. 2008 NYC calorie information study.
  - page=36, words=1-12, section=References, note=Currie et al. 2010: The Effect of Fast Food Restaurants on Obesity and Weight Gain.
  - page=7, words=1-16, section=References, note=Downs, Loewenstein, and Wisdom (2009): Strategies for Promoting Healthier Choices.
  - page=37, words=550-624, section=References, note=Kiesel & Villas-Boas (2008): Nutrition labeling experiments.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.23832736218425818)`
- Retrieval method: `local`
- Evidence type: `In-text citations of prior NYC calorie-posting studies discussed as direct predecessors.`
- Data source: Receipt-based purchase data from NYC studies (Downs et al. 2009; Elbel et al. 2009).
- Answer:

Downs, Loewenstein, and Wisdom (2009) and Elbel, Kersh, Brescoll, and Dixon (2009) are the papers this work most directly builds on or extends.

- Quote snippet:

  Two subsequent papers compare purchase data before and after calorie posting in NYC. Downs, Loewenstein, and Wisdom (2009) collected a total of 1,354 receipts from patrons.

- Citation anchors:
  - page=8, words=275-554, section=Discussion of prior NYC calorie posting studies, note=Mentions two NYC-based prior 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 (0.2601991366407673)`
- Retrieval method: `local`

- Evidence type: `textual`
- Data source: NYC calorie posting literature; Downs et al. 2009; Elbel et al. 2009.
- Answer:

The papers by Downs, Loewenstein, and Wisdom (2009) and Elbel et al. (2009) are used as benchmarks/comparisons.

- Quote snippet:

  > Two subsequent papers compare purchase data before and after calorie posting in NYC—Downs, Loewenstein, and Wisdom (2009) and Elbel et al. (2009).

- Citation anchors:
  - page=8, words=0-60, section=Prior literature / comparison studies, note=States that two subsequent papers compare purchase data before and after calorie posting in NYC (Downs et al. 2009; Elbel et al. 2009).
  - page=8, words=61-120, section=Prior literature / comparison studies, note=Describes the two cited prior studies as comparisons/benchmarks for the current analysis.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.28915800382619017)`
- Retrieval method: `local`
- Evidence type: `data_sources`
- Data source: Transaction data (Starbucks NYC, Boston, Philadelphia purchases); Cardholder data (anonymous cardholders); Milk order data; In-store surveys (Seattle and San Francisco); prior comparative dataset (Downs, Lowenstein, Wisdom 2009).
- Table/Figure: Table 1
- Answer:

The study cites four main data sources used to analyze the impact of calorie posting: (1) Transaction data covering Starbucks purchases across NYC, Boston, and Philadelphia (222 NYC stores and 94 in the other cities) for 3 months before and 11 months after posting, including time, store, items, price, and calories; (2) Cardholder data consisting of anonymous Starbucks cardholder purchases over the same period, enabling within-person analysis and comparison with the transaction data; (3) Starbucks milk order data capturing daily replenishment of different milk types to assess potential calorie contributions from milk choice; and (4) In-store consumer surveys conducted in Seattle and San Francisco (before and after posting) to gauge consumer knowledge and control for time trends. A prior dataset from Downs, Lowenstein, and Wisdom (2009) is also referenced for comparison with a much smaller dataset of receipts.

- Quote snippet:

  At each location we observe all trans- actions for a period of time 3 months before and 11 months after calorie posting commenced (i.e., January 1, 2008–February 28, 2009).

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=Describes the transaction dataset and the cardholder dataset, including scope (NYC, Boston, Philadelphia) and the two datasets used.
  - page=5, words=275-591, section=A. Data Summary, note=Details the cardholder data, its representativeness caveats, and its use to isolate effects on intensive vs extensive margins.
  - page=6, words=0-349, section=Table 1—Summary Statistics for Transaction Data and Cardholder Data (Prior to policy change), note=Introduces the primary data tables summarizing both datasets prior to the calorie posting policy change.
  - page=6, words=0-349, section=Data Sources, note=Mentions milk order data as an additional data source that informs calorie analysis via milk selections.
  - page=9, words=0-349, section=Literature/Background, note=References a prior dataset (Downs, Lowenstein, Wisdom 2009) used for comparison with a smaller receipts dataset.
  - page=14, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Describes how results from the transaction and cardholder datasets are presented (Figure 2) and how within- vs between-person variation informs analysis.
  - page=25, words=275-599, section=Seattle/San Francisco Surveys, note=Details the in-store consumer surveys conducted in Seattle and San Francisco (before/after posting) to test knowledge and control for time trends.
  - page=6, words=0-349, section=Milk Data, note=Notes obtaining daily milk-order data (regular, skim, nonfat) for all stores in NYC, Boston, and Philadelphia to assess dairy-related calories.

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

- Category: `E) Citations and related literature`
- Confidence: `medium (0.3864347264898625)`
- Retrieval method: `local`
- Evidence type: `Methodology description and references`
- Data source: Starbucks transaction data and cardholder data
- Table/Figure: Figure 1; Figure 2
- Assumption flag: `True`
- Assumption notes: Assumes parallel trends or no differential shocks across NYC vs control cities; authors argue pre-trend differences are not present

and rely on within-city variation over time for identification.
- Answer:

The study relies on a fixed-effects panel design with within-city variation (a diff-in-diff–style identification using city-week variation and store fixed effects), estimating regressions with a calorie-posting indicator and various fixed effects; it also cites randomized field experiments as methodological references (e.g., Cai, Chen, and Fang 2009) and related econometric work (e.g., Chetty, Looney, and Kroft 2009; DellaVigna 2009).

- Quote snippet:

  In both cases (transaction data and cardholder data), identification of the effect of calorie posting stems from within-city variation over time.

- Citation anchors:
  - page=11, words=275-522, section=Methodology - Regression specification, note=Identification stems from within-city variation over time; store fixed effects included
  - page=12, words=0-349, section=Identification, note=Identification of the calorie-posting effect stems from within-city variation over time for both datasets
  - page=37, words=0-349, section=References - Methodology, note=Cites Cai, Chen, and Fang (2009) randomized natural field experiment; Chetty, Looney, Kroft (2009); DellaVigna (2009)

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.24675499392198427)`
- Retrieval method: `local`
- Evidence type: `references to seminal literature cited in the references section`
- Answer:

Yes. The paper cites classic nutrition-information literature and seminal works such as Jacoby et al. (1977) on consumer use and comprehension of nutrition information, as well as earlier works like Jacoby (1974) and Ippolito & Mathios (1990, 1995).

- Quote snippet:

  Jacoby, Jacob, Robert W. Chestnut, and William Silberman. 1977. "Consumer Use and Comprehension of Nutrition Information."

- Citation anchors:
  - page=37, words=0-349, section=References, note=Classic nutrition information literature; includes Jacoby et al. (1977) 'Consumer Use

and Comprehension of Nutrition Information'.
  - page=37, words=275-624, section=References, note=Includes Ippolito & Mathios (1990, 1995) and Jacoby (1974); foundational nutrition labeling studies.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.29702986581922664)`
- Retrieval method: `local`
- Evidence type: `appendix and data description`
- Data source: Transaction data and cardholder data
- Answer:

Yes. The paper cites an Appendix where the survey questionnaire is shown, and describes two primary data sources (transaction data and cardholder data) that support its claims.

- Quote snippet:

  The questionnaire is shown in the Appendix.

- Citation anchors:
  - page=7, words=The-Appendix., section=Appendix (survey questionnaire), note=Sentence indicates the 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 (0.29240311641257244)`
- Retrieval method: `local`
- Evidence type: `Author-stated limitations / gaps in prior literature`
- Data source: Page 3 limitations discussion
- Answer:

The literature does not directly measure obesity effects of calorie posting; BMI data from the CDC are not yet available; studies have limited generalizability because data cover only one chain (Starbucks) and do not address whether effects transfer to other chains or whether consumers offset calorie reductions at Starbucks by eating differently at home.

- Quote snippet:

  First, we do not directly measure the effect of calorie posting on obesity itself. Current lags in the availability of BMI data from the

Centers for Disease Control (CDC) suggest this will not be addressed for a few more years.

- Citation anchors:
  - page=3, words=275-564, section=Limitations of prior research, note=Gaps include lack of direct obesity impact measurement, unavailability of BMI data, and generalizability from a single chain.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.16771906052772012)`
- Retrieval method: `local`
- Evidence type: `absence of information in provided context`
- Answer:

No replication files or code are mentioned in the provided context.

- Quote snippet:

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

- Citation anchors:
  - page=5, words=0-349
  - page=5, words=275-591
  - page=7, words=0-349
  - page=15, words=0-254
  - page=23, words=0-349
  - 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 (0.25837690434210736)`
- Retrieval method: `local`
- Evidence type: `textual analysis of Methods/Data sections; no mention of preregistration`
- Data source: Transaction data and cardholder data; in-store surveys in Seattle and San Francisco
- Answer:

No explicit 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) ysct = xsct + POSTct + sct,

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=No mention of a pre-analysis plan or registered trial; describes data sources.
  - page=7, words=275-566, section=B. Study Design / Data, note=Describes surveys and data collection; no registration noted.
  - page=9, words=0-349, section=Introduction / Background, note=No discussion of preregistration or pre-analysis plan in this section.
  - page=11, words=275-522, section=Estimation Strategy, note=Outlines regression specification; no preregistration referenced.
  - page=25, words=275-599, section=Appendix: Survey Methods, note=Survey methods described; no preregistration mentioned.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2749485981300251)`
- Retrieval method: `local`
- Evidence type: `confidentiality/restricted access`
- Data source: Starbucks transaction data and cardholder data
- Answer:

Yes. Data access is restricted due to confidentiality requirements, and details cannot be disclosed.

- Quote snippet:

  Due to confidentiality requirements, we are unable to reveal any more details about these differences.

- Citation anchors:
  - page=6, words=275-517, section=Data confidentiality, note=Direct statement that confidentiality requirements prevent revealing details.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2185788399744389)`
- Retrieval method: `local`
- Evidence type: `methodology description`
- Data source: Starbucks transaction data in NYC, Boston, Philadelphia; and cardholder data
- Answer:

Yes. The paper documents key data-cleaning steps, including excluding stores not open for the full data period to form a balanced panel, and excluding transactions with more than four units; it also describes data construction details such as dataset definitions and a normalization step to preserve confidentiality.

- 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=5, words=0-349, section=Data Description, note=Transaction and cardholder data described; data include time, store, items, prices, and calories.
  - page=5, words=275-591, section=Data Cleaning / Confidentiality, note=Normalization of NYC values to one for confidentiality.
  - page=6, words=0-349, section=Data Cleaning, note=Explicit cleaning steps: exclude stores not open during entire period and exclude transactions with >4 units; 2.2% excluded.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2842173325501089)`
- Retrieval method: `local`
- Evidence type: `Mention of a robustness check (excluding weather controls) but no full set of robustness/sensitivity analyses described.`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia).
- Answer:

Partially reported.

- Quote snippet:

  "Our findings are unchanged if we exclude the weather controls."

- Citation anchors:
  - page=11, words=275-522, section=Analysis, note=States that findings are unchanged if weather controls are excluded (a robustness check).

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.285164312231853)`
- Retrieval method: `local`

- Evidence type: `empirical results`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia, Seattle; U.S. locations)
- Table/Figure: Table 5; Table 9
- Answer:

Generalizable to urban, chain-restaurant customers who are regularly exposed to calorie labeling—especially higher-income, more educated populations and female customers who tend to purchase higher-calorie items; and to commuters who are exposed to labeling in one city and then shop in other locations.

- Quote snippet:

  "the decrease in calories per transaction was larger in zips with higher income and in zips with more education (i.e., more people with college degrees)."

- Citation anchors:
  - page=21, words=0-349, section=Heterogeneity in the Impact, note=Income and education amplify the effect; female cardholders more responsive.
  - page=20, words=0-349, section=Heterogeneity in the Impact, note=Gender differences in responsiveness.
  - page=29, words=0-349, section=Commuters, note=Commuters' post-April non-NYC transactions show effects consistent with learning; testing outside NYC after NYC exposure.
  - page=30, words=0-349, section=Commuters, note=Learning vs salience interpretation; non-NYC effects depend on prior NYC posting.
  - page=5, words=275-591, section=Data Summary, note=Cardholder data may not be representative; above-average Starbucks loyalty.

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.21434355896608506)`
- Retrieval method: `local`
- Evidence type: `Limitations on generalizability`
- Data source: Bollinger et al., American Economic Journal: Economic Policy, 2011
- Assumption flag: `False`
- Answer:

Populations/settings least likely to generalize are: (1) Starbucks cardholders in the sample who are above-average in loyalty and may not represent all Starbucks customers; (2) Starbucks data limited to company-owned stores (excluding independent locations); (3) applicability to other restaurant chains or contexts beyond Starbucks; and (4) contexts with differing policy configurations (e.g.,

Seattle's pastry exemption) which may not generalize to beverages-only policies.

- Quote snippet:

  A 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=Second limitation: data for only one chain (Starbucks); cannot know if effects generalize to other chains.
  - page=5, words=0-349, section=Data coverage, note=Transaction data cover all NYC Starbucks locations and all Boston/Philadelphia locations; independent locations excluded.
  - page=5, words=275-591, section=Cardholder representativeness, note=Cardholders may be above-average in loyalty; not representative of Starbucks customers in general.
  - page=14, words=275-590, section=Contextual generalizability, note=Seattle policy context with pastry exemption; limits generalization to other contexts/policy designs.

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.20192774300854582)`
- Retrieval method: `local`
- Evidence type: `limitations and boundary conditions`
- Data source: cardholder dataset; Starbucks transaction data
- Answer:

Yes. The authors discuss boundary conditions and scope limits, noting two main limitations: they do not directly measure obesity effects, and the data cover only one chain (Starbucks), limiting generalizability and raising questions about offsetting behaviors.

- 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/Scope, note=Authors describe two major limitations: no direct obesity measurement; data from a single chain.

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

- Category: `G) External validity and generalization`
- Confidence: `medium (0.3574931243293877)`
- Retrieval method: `local`
- Evidence type: `empirical study findings (calorie posting effects across time and markets)`
- Data source: Bollinger et al., Calorie Posting in Chain Restaurants, American Economic Journal: Economic Policy (2011)
- Answer:

The results could change across different time periods or markets due to how posting effects interact with memory, seasonality, and consumer inflows, as well as local market characteristics. In this study, the calorie posting effect appears immediately after posting (April 1, 2008) and persisted through February 2009, with some holiday-related deviations observed in the transaction data, and a different pattern around Christmas in the cardholder data. Commuter-heavy patterns show a larger reduction in calories per transaction than non-commuters, and the magnitude of substitution between food/beverage items varies by dataset and market. The analysis relies on NYC stores as the treatment and Boston/Philadelphia as controls, and the authors caution that results may not generalize to other chains (only Starbucks was observed) or to markets with different competitive dynamics (e.g., proximity of Dunkin Donuts). Therefore, in other time periods or markets, one could see: (i) similar immediate post-posting drops but with larger or smaller magnitudes depending on commuting shares or holiday effects; (ii) different patterns of substitution (more or less reduction in food calories vs beverage calories) depending on local menus and consumer behavior; (iii) altered effects if calorie posting occurs in more or fewer nearby locations or if other chains' posting policies are in place; (iv) potential differences in profitability effects if local competition shifts margins. In short, the direction—calorie reductions following posting—may persist but the size and composition of effects are likely to vary with time period and market structure.

- 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=which-IV., section=Context and limitations, note=Notes limitations: only one chain (Starbucks) and other general caveats.
  - page=6, words=96-significant, section=Table 1: Summary Statistics, note=Pre-post characteristics and control comparisons across cities.
  - page=11, words=likely-seasonality)., section=Methods and controls,

note=Controls for seasonality and weather; within-city variation.
  - page=12, words=to-0.119), section=Cardholder data results, note=Cardholder-level estimates of calorie changes (food vs beverage).
  - page=14, words=104-199, section=Post-posting trend and robustness, note=Immediate post-posting drop and persistence; holiday deviations observed in some data.
  - page=16, words=106-table, section=Substitution effects, note=Product-level substitution analysis; margins and pricing considerations.
  - page=29, words=Vol.-calorie, section=Commuters table (Table 9), note=Commuter vs non-commuter effects on calories per transaction.
  - page=24, words=stemmed-policy, section=Discussion of margins and profits, note=Revenue and profit implications; context for market differences.

H01: Are key variables measured directly or via proxies?

- Category: `H) Measurement validity`
- Confidence: `low (0.30790171154554274)`
- Retrieval method: `local`
- Evidence type: `Direct measurement of calories per transaction from item-level nutrition data; proxy measurement via milk order data for milk usage`
- Data source: Transaction data; Cardholder data; Milk order data
- Answer:

Key variables related to calories are measured directly from item-level nutrition data (calories per beverage, food item, and total calories per transaction). Some related measures, such as milk usage, are proxied using auxiliary data (milk order data) rather than direct caloric intake.

- Quote snippet:

  An important variable of interest is calories per transaction. Based on the transaction data, we compute that, prior to calorie posting, in NYC: average drink calories per transaction were 143; total calories per transaction were 247.

- Citation anchors:
  - page=6, words=0-349, section=Milk order data, note=Milk order data used to assess milk quantities (proxy for milk-related calories)
  - page=6, words=275-517, section=Calories per transaction, note=Direct measurement: calories per transaction computed from transaction data

H02: What measurement error risks are acknowledged or likely?

- Category: `H) Measurement validity`

- Confidence: `low (0.3324416287995888)`
- Retrieval method: `local`
- Evidence type: `limitations/measurement error risks`
- Data source: Starbucks transaction data and cardholder data; Seattle/San Francisco in-store surveys
- Answer:

Key measurement error risks acknowledged or likely include: (1) the study does not directly measure obesity outcomes, only calorie posting effects on purchases; (2) data are drawn from a single chain (Starbucks), limiting generalizability and potentially misrepresenting effects at other chains; (3) potential offsetting changes at home are unobserved; (4) survey-based knowledge measures rely on self-reports and are prone to error (calorie knowledge found to be very inaccurate); (5) survey sampling bias, such as under-representation of customers who order regular coffee, could distort results; (6) confidentiality restrictions limit disclosure of data differences, complicating interpretation of measurement properties.

- Quote snippet:

  First, we do not directly measure the effect of calorie posting on obesity itself.

- Citation anchors:
  - page=3, words=275-564, section=Limitations, note=Two important limitations: no direct obesity measurement; data from only one chain.
  - page=3, words=275-564, section=Scope limitations, note=Data availability restricted to Starbucks; generalizability limited.
  - page=25, words=275-599, section=Survey sampling bias, note=In Seattle/San Francisco surveys, sample under-represents consumers who ordered regular coffee.
  - page=26, words=0-178, section=Survey measurement error, note=Figure 5 shows respondents' calorie knowledge is very inaccurate; measurement error in knowledge.
  - page=27, words=0-349, section=Survey results and measurement error, note=Respondents overestimate beverage calories; variability indicates measurement challenges in self-reported estimates.
  - page=6, words=275-517, section=Data limitations, note=Confidentiality prevents full detail; differences across data sources complicate interpretation.

H03: Are there validation checks for key measures?

- Category: `H) Measurement validity`
- Confidence: `low (0.24861300785380763)`
- Retrieval method: `local`
- Evidence type: `robustness checks and cross-dataset validation`
- Data source: Transaction data and Cardholder data

- Answer:

Yes. The article reports multiple validation checks for key measures, including robustness checks (e.g., day-of-week and holiday dummies), cross-dataset validation showing similar effects in both transaction and cardholder data, borough-level robustness (Manhattan vs other boroughs), Seattle post-law validation, and distributional (quantile) analysis to verify effects across calories per transaction.

- Quote snippet:

  > as a robustness check, 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=5, words=275-591, section=Data sources and sample, note=Describes two datasets (transaction data and cardholder data) and representativeness; comparison of effects.
  - page=6, words=0-349, section=Data description, note=Table 1 with summary statistics and cross-city controls.
  - page=11, words=275-522, section=Model controls and specification, note=Regression specification with fixed effects; weather as controls.
  - page=12, words=550-630, section=Weather controls, note=Weather controls are insignificant; within-week variation identified.
  - page=14, words=275-590, section=Robustness checks, note=Robustness checks with day-of-week/holiday dummies; Seattle post-law; boroughs.
  - page=16, words=275-558, section=Cross-dataset validation, note=Note that both the transaction data and cardholder data yield similar estimates.
  - page=21, words=275-421, section=Quantile analysis, note=Validation across calories per transaction quantiles.
  - page=22, words=0-349, section=Quantiles results, note=Table 6 presents quantile-based estimates.

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

- Category: `H) Measurement validity`
- Confidence: `low (0.20298002823158073)`
- Retrieval method: `local`
- Evidence type: `limitations and measurement/methods discussion`
- Data source: Cardholder transaction data; store/NYC Seattle transaction data
- Answer:

No. The provided excerpts do not include an explicit discussion of construct validity for core outcomes; they discuss limitations (e.g., not measuring obesity

directly) and how outcomes are measured, but do not address construct validity.

- Quote snippet:

  There are two important limitations to this research. First, we do not directly measure the effect of calorie posting on obesity itself.

- Citation anchors:
  - page=3, words=275-564, section=Limitations and empirical design (page 3), note=Mentions two important limitations and that obesity effects are not directly measured; no explicit construct validity discussion.
  - page=11, words=0-349, section=Methods (page 11), note=Defines outcome y_sct (calories per transaction) and regression specification; shows measurement approach but not validity discussion.

I01: What policy counterfactuals are considered or implied?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.34057657058974905)`
- Retrieval method: `local`
- Evidence type: `Explicitly described and implicitly relied-upon policy counterfactuals in study design and discussion`
- Answer:

Policy counterfactuals considered/implied: (1) a no-posting counterfactual in NYC using Boston and Philadelphia as control cities to estimate what would have happened absent calorie posting; (2) a universal disclosure counterfactual (information unraveling) where all firms voluntarily disclose calorie information; (3) an obesity-outcome counterfactual about whether observed calorie reductions would translate into reductions in obesity (acknowledged but not directly measured); and (4) a menu-offerings counterfactual about whether calorie posting would spur restaurants to offer more low-calorie options.

- 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=20, words=0-349, section=B. Substitution Effects, note=Uses control cities (Boston and Philadelphia) to identify the policy effect (no posting in control cities).
  - page=2, words=0-349, section=Section IV, note=Mentions information unravelling as a counterfactual scenario where all firms disclose calories.
  - page=32, words=0-349, section=Is the Effect Big Enough to Matter?, note=Discusses whether calorie reductions could translate into

obesity reductions (obesity counterfactual).
  - page=32, words=0-349, section=Manager Survey / Menu Offerings, note=Considers whether posting drives more low-calorie menu options (menu-offerings counterfactual).
  - page=11, words=275-522, section=Method/Identification, note=Policy variation identified at city-week level; supports the no-posting counterfactual framework.

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

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3101645282666716)`
- Retrieval method: `local`
- Evidence type: `empirical results on welfare and distributional effects (extensive vs. intensive margins, profits, and heterogeneous effects by income, education, gender, and calorie quantiles).`
- Data source: transaction data and cardholder data
- Answer:

Mandatory calorie posting reduced calories primarily through the extensive margin (people forgoing food items), with about three-quarters of the total reduction coming from not buying food and about one-quarter from substituting toward lower-calorie options. Beverage calories were largely unaffected and there was no change in transaction frequency. Welfare implications were mixed: profits were not significantly affected on average, but the average price per item rose, suggesting that purchases may have shifted toward higher-margin products; the profit effect may depend on local competition (e.g., whether a nearby Dunkin Donuts is present). Distributionally, responses were larger in higher-income and more-educated ZIPs, female cardholders were more responsive, and high-calorie purchasers showed greater reductions; quantile results show larger absolute reductions at higher calories, while percent reductions are roughly 5–6% from the 75th to 99th percentile. Overall, the welfare impact is not uniform across consumers or stores, but calorie posting achieved meaningful reductions in food calories without a broad profit decline.

- Quote snippet:

  > 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=page 16, words=0-349, section=B. Substitution Effects, note=Highlights that most calorie reduction is due to the extensive margin (not buying food) and substitution to lower-calorie items.

  - page=page 18, words=0-349, section=Table 4 and discussion, note=Presents coefficients for items per transaction and calories per item; relates to substitution/extensive margins and calorie effects.
  - page=page 20, words=0-349, section=Heterogeneity in the Effect of Mandatory Calorie Posting, note=Reports larger decreases in calories per transaction in higher-income and more-educated ZIPs; female cardholders more responsive; other subgroup findings.
  - page=page 22, words=0-349, section=Table 6 and quantile analysis, note=Calorie reductions larger at top quantiles; absolute changes higher for higher quantiles, percent change roughly 5–6% from 75th to 99th percentile.
  - page=page 24, words=275-564, section=III. Effect of Mandatory Calorie Posting on Profit, note=Discusses costs of policy and impact on revenues/profits; notes price per item rose; profits may depend on nearby competitors.
  - page=page 24, words=275-564, section=IV. Discussion, note=Summarizes welfare implications: 3/4 of calorie reduction from not buying food; 1/4 from substitutions; overall effects on profits and scale of response.

I03: Are cost-benefit or incidence analyses provided?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3128035665280039)`
- Retrieval method: `local`
- Evidence type: `qualitative discussion`
- Answer:

No formal cost-benefit or incidence analyses are provided. The paper discusses costs of compliance and indirect costs (e.g., about $2,000 per Wendy's store) and notes potential indirect costs and qualitative potential benefits, but does not present a quantified cost-benefit or incidence analysis.

- Quote snippet:

  One news report indicated the cost of compliance for the Wendy's chain was about $2,000 per store. However, the law may have generated some additional indirect costs for chains.

- Citation anchors:
  - page=4, words=275-576, section=IV. Discussion, note=Notes costs of compliance and potential indirect costs; no formal cost-benefit or incidence analysis provided.

I04: What policy recommendations are stated or implied?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.33657103878918027)`

- Retrieval method: `local`
- Evidence type: `policy_recommendation`
- Data source: Starbucks transaction data in NYC with controls in Boston and Philadelphia; cardholder data; in-store surveys.
- Answer:

Adopt mandatory calorie labeling on menus (calorie posting) and expand it beyond NYC, as a policy with low costs and potential long-run benefits. The authors show that mandatory posting reduces calories per transaction (about 6%), is unlikely to harm profits, may drive low-calorie menu innovation, and offers public-education benefits; thus policy should be pursued nationwide rather than relying on voluntary 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=We-Starbucks., section=Conclusion, note=Stated 6% decline in calories per transaction.
  - page=34, words=The-costs., section=Conclusion, note=Costs of posting described as very low; benefits could outweigh costs.
  - page=33, words=Many-writing., section=Conclusion, note=Policy expansion nationwide referenced.

J01: What missingness or attrition patterns are reported?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.31588907534780863)`
- Retrieval method: `local`
- Evidence type: `Data limitations, sampling and attrition`
- Data source: Starbucks transaction data and anonymous cardholder data (NYC, Boston, Philadelphia)
- Assumption flag: `True`
- Assumption notes: Authors assume cardholders are above-average in Starbucks loyalty, which may limit generalizability to all customers.
- Answer:

The study reports several missingness/attrition patterns: (1) BMI data are not available, limiting obesity-related analysis; (2) data cover only Starbucks, limiting generalization to other chains; (3) to form a balanced panel, some stores/transactions are excluded (e.g., stores not open for the full period and purchases of more than four units), comprising about 2.2% of transactions; (4) a subset of cardholders is used (7,520 NYC; 3,772 Boston/Philadelphia) from a pool of 2.7 million anonymous individuals, and these cardholders may not be representative of overall Starbucks customers; (5) the fraction of excluded transactions is unknown but believed to be well under 5%; and (6) confidentiality

restrictions limit disclosure of additional details about the data.

- Quote snippet:

  First, we do not directly measure the effect of calorie posting on obesity itself. Current lags in the availability of BMI data from the Centers for Disease Control (CDC) suggest this will not be addressed for a few more years.

- Citation anchors:
  - page=3, words=275-564, section=Limitations, note=BMI data unavailable; study limited to one chain; empirical question remains
  - page=5, words=0-349, section=Data Summary, note=Data summary; 2.7 million anonymous cardholders; 7,520 NYC cardholders; 3,772 Boston/Philadelphia cardholders; representativeness concerns
  - page=5, words=0-349, section=Data Summary, note=Fraction of excluded transactions unknown but believed well under 5 percent
  - page=5, words=0-349, section=Data Summary, note=Excluded transactions represent 2.2% of all transactions
  - page=5, words=0-349, section=Data Summary, note=Due to confidentiality requirements, unable to reveal more details about these
  - page=5, words=0-349, section=Data Summary, note=Cardholders may not be representative of Starbucks customers; LOA in loyalty

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

- Category: `J) Data quality and integrity`
- Confidence: `low (0.3112636229771498)`
- Retrieval method: `local`
- Evidence type: `exclusion criteria (outlier handling)`
- Data source: Transaction data and cardholder data (prior to calorie posting).
- Answer:

Outliers are handled by excluding certain transactions from the analysis: transactions at stores not open for the entire data period (to analyze a balanced panel) and transactions that included more than four units of any one item; these exclusions account for a small fraction of transactions (about 2.2%).

- 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.

- Citation anchors:
  - page=5, words=275-591, section=Data exclusions and balanced panel, note=Excludes stores not open during entire data period and

excludes purchases with more than four units; excludes about 2.2% of transactions.

J03: Are there data audits or validation steps described?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.2433617621771917)`
- Retrieval method: `local`
- Evidence type: `robustness checks and data validation steps in methods/results`
- Data source: Transaction data and cardholder data from Starbucks stores in NYC, Boston, Philadelphia; Seattle post-law data for robustness.
- Assumption flag: `True`
- Assumption notes: Assumes Boston and Philadelphia are reasonable control cities for NYC; supported by pre-trend checks and robustness analyses.
- Answer:

Yes. The study describes data cleaning/validation steps (balanced panel construction, exclusions of certain transactions, confidentiality-preserving normalization), uses two data sources for cross-validation (transaction data and cardholder data), and conducts robustness tests (day-of-week/holiday controls, pre-trend checks showing no differences), plus a Seattle post-law robustness check and corroborating control-city comparisons.

- 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 handling and confidentiality; data cleaning, note=Normalization to protect confidentiality and exclusion of some transactions; use of balanced panel.
  - page=14, words=0-349, section=Validation/Robustness checks, note=No evidence of pre-trend differences; robustness checks with day-of-week/holiday dummies; within-individual variation.
  - page=9, words=0-349, section=Introduction to methodology and data sources, note=Two datasets: transaction data and cardholder data; cross-validation of effects.
  - page=25, words=275-599, section=Data collection/surveys and validation data, note=In-store customer surveys; data across locations/waves; supports validation of consumer responses.

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

- Category: `J) Data quality and integrity`
- Confidence: `low (0.3050038923260564)`

- Retrieval method: `local`
- Evidence type: `selective sample inclusion / sampling bias`
- Data source: Cardholder data and transaction data from Starbucks (NYC, Boston, Philadelphia)
- Assumption flag: `True`
- Assumption notes: The analysis assumes the cardholders are above-average loyal to Starbucks, which may bias representativeness of the sample.
- Answer:

Yes. The study notes selective sample inclusion: the cardholder data may not be representative of Starbucks customers generally, since the analyzed cardholders are expected to be above-average in loyalty to Starbucks. This limits generalizability.

- Quote snippet:

  > however, 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=page 5, words=however,-Starbucks., section=Data limitations – representativeness of cardholder data, note=Explicit statement that cardholders may not be representative; cardholder sample is chosen as a subsample.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.2470820443992412)`
- Retrieval method: `local`
- Evidence type: `regression_diagnostic_metrics`
- Answer:

The paper reports regression goodness-of-fit and diagnostic metrics: (i) R-squared values (R^2) ranging from 0.27 to 0.82 for the transaction data and 0.26 to 0.37 for the cardholder data; (ii) a Pearson chi-square test with p = 0.11 when comparing NYC vs control cities; (iii) 95% confidence intervals in the regression figures; and (iv) robust standard errors used in Table 9.

- Quote snippet:

  > R^2's ranging from 0.27 to 0.82.

- Citation anchors:
  - page=18, words=0-349, section=Table 4, note=R^2's ranging from 0.27 to 0.82.
  - page=12, words=0-349, section=Notes, note=R^2 vary between 0.26 and 0.37 for cardholder data.

  - page=11, words=0-349, section=Notes, note=Pearson's chi-square test p = 0.11.
  - page=15, words=0-254, section=Figure 2, note=Dashed lines represent 95 percent confidence intervals.
  - page=29, words=0-349, section=Table 9, note=Robust standard errors in parentheses.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.2678571462170702)`
- Retrieval method: `local`
- Evidence type: `Regression specification descriptions and variable transformations (levels and logs) with nonlinear weather controls.`
- Data source: Cardholder transaction data and store-level transaction data from Starbucks.
- Answer:

Yes. The study tests multiple functional forms: level (linear) specifications for calories per transaction, and log specifications for variables like daily sales and calories per item; they also include nonlinear controls (e.g., weather variables with squared terms).

- 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 t + _sct.

- Citation anchors:
  - page=11, words=0-349, section=Methods – regression specification, note=Describes a linear level specification for calories per transaction.
  - page=16, words=0-349, section=Substitution/empirical methods, note=Uses log daily sales and log calories per item purchased as dependent variables.
  - page=11, words=0-349, section=Controls, note=Weather controls include temperature and temperature squared, precipitation and precipitation squared (nonlinearities).
  - page=29, words=0-349, section=Table 9 – commuters, note=Dependent variable shown as log(calories + 1) in regression results.

K03: Are residual checks or specification tests reported?

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.23550474337660648)`
- Retrieval method: `local`

- Evidence type: `Contains references to regression specs and a chi-square test; no residual diagnostics or specification tests are described.`
- Answer:

No. There is no mention of residual checks or formal specification tests in the provided material; only regression specifications with fixed effects and robustness via robust standard errors are described, along with a Pearson chi-square test in Table 2.

- Quote snippet:

  Pearson's chi-square test fails to reject that the cell proportions for NYC are equal to those for Boston and Philadelphia (p-value of 0.11).

- Citation anchors:
  - page=11, words=0-349, section=Table 2 notes, note=Mentions Pearson's chi-square test comparing NYC with control cities.
  - page=29, words=0-349, section=Table 9 notes, note=Notes that robust standard errors are used.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `medium (0.37651280310773627)`
- Retrieval method: `local`
- Evidence type: `robustness checks and alternative specifications`
- Data source: Transaction data and cardholder data from the Starbucks calorie posting study (NYC and control cities).
- Answer:

Results are fairly robust to alternative specifications and estimators. Robustness checks show that adding date fixed effects (instead of week/day controls) barely changes the estimated effects of calorie posting; weather controls have little to no impact, and findings hold if weather controls are dropped. Using cardholder data to isolate intensive-margin effects yields qualitatively similar conclusions, though magnitudes and significance can differ by dataset (e.g., beverage calories may be not significant in cardholder data, while food calories show a sizable decrease). Overall, the qualitative conclusions about calorie posting effects are robust to several alternative specifications and estimators.

- Quote snippet:

  The estimates based on the transaction data are barely changed.

- Citation anchors:
  - page=11, words=275-522, section=Robustness and weather controls, note=weather controls included; findings unchanged if excluded

  - page=12, words=0-349, section=Robustness checks and specification robustness, note=date fixed effects; results barely changed; weather controls insignificant; findings unchanged if weather controls excluded

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

- Category: `P) Previous questions`
- Confidence: `n/a`
- Retrieval method: `n/a`
- Evidence type: `n/a`
- Answer:
- Key context: NYC calorie posting policy prompted concerns about costs and potential behavioral responses; evidence includes multiple analyses from NYC vs control cities and Seattle as a control. (page 4, words 275-576)
- Main quantitative finding on overall calories: Calorie posting reduces calories per transaction and per store-day, with a documented drop in average calories per store-day of about 4.6 after posting; the reduction in calories per transaction is larger than the store-day reduction, implying a rise in transactions alongside lower per-transaction calories. (page 13, words 275-528)
- Item-level and margin effects (Table 4):
  - Beverages: number of beverages per transaction increases slightly (ΔBev $+0.005$); calories per beverage fall slightly (log calories per beverage -0.008).
  - Foods: number of food items per transaction falls (ΔFood -0.029); calories per food item fall more (log calories per food item -0.039).
  - Net: total items per transaction decline (log scale -0.027), with roughly three-quarters of the total calorie reduction coming from not purchasing food items (extensive margin). (page 18, words 0-349)
- Seattle check (placebo / external test): In Seattle (where food calories were not posted), drink calories per transaction fell by about 4.6 and food calories per transaction rose about 0.8, consistent with the idea that posted calories drive beverage reductions but not food-calorie changes where posting didn't occur. This supports the NYC finding on beverages and the absence of a large food-calorie shift when posting is absent. (page 15, words 0-254)
- Heterogeneity in impact (Table 5): The effect of posting on log(calories per transaction) varies with characteristics:

  - A few interactions are statistically significant (e.g., median income, percent with college degree, percent female; high- and low-calorie cohorts show differential responses).
  - Notably, high-calorie customers show larger reductions. (page 21, words 0-349)

- Quantile results (Table 6): Calorie reductions are larger at higher quantiles of calories per transaction:
  - 75th–99th percentile reductions are substantial in absolute calories yet roughly 5–6% in percentage terms.
  - The 99th percentile shows the largest absolute decrease (−77 calories) with a still ~5–6% relative change. (page 22, words 0-349)

- Appendix / survey instrument: The paper includes Appendix A with the Stanford University research survey and Starbucks customer survey forms, indicating primary data collection from customers. (page 35-36, words 0-349)

- Additional notes:
  - The results and discussion reference the array of potential consumer responses (extensive vs intensive margins, substitution effects) and discuss the role of different menu postings across cities. (page 4, 275-576; page 10-12 context)

If you'd like, I can extract a compact figure-ready summary (e.g., a one-page table of coefficients and percent changes by margin and quantile) with the exact numbers and cited page references.