Audit Report: Workflow 69751a2167c847a894f2d5d3c78b8a9d

Overview

- Source JSON: reports\workflow-report-69751a2167c847a894f2d5d3c78b8a9d.json
- Run ID: 69751a2167c847a894f2d5d3c78b8a9d
- Papers input: papers\Calorie Posting in Chain Restaurants - Bollinger et al. (2011).pdf
- Started at: 2026-02-16T04:34:15.281488+00:00
- Finished at: 2026-02-16T04:43:25.512897+00:00
- Duration: 0:09:10.231409

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: 0.27966773663401223 / 0.2842025482871281
- Confidence labels: low=74, medium=9, high=0

Final Answer

- Main contributions
  - Provides empirical evidence on the NYC calorie-posting policy (post-2008) and its impact on calories, item-level purchases, and overall

shopping behavior, with a decomposition into margins and exploration of heterogeneity. (Provenance: Background: page 4, words 275-576)
  - Documents beverage-vs-food dynamics, heterogeneous effects by customer type, and quantile-based differences (stronger effects at higher calories-per-transaction quantiles). (Provenance: Beverage/food results: page 18; Heterogeneity: page 21; Quantiles: page 22)
  - Uses Seattle as a control context and conducts robustness checks with alternative time specifications, supporting the credibility of the identified effects. (Provenance: Seattle control: page 15; Robustness/alternative specs: page 13)

- Identification strategy
  - Natural experiment leveraging a binary Post indicator (POST = 1 for NYC stores after 1 Apr 2008) with store fixed effects, time controls (week/day dummies, weather), and robust specifications; cross-checks with separate weekly dummies. (Provenance: Regression controls and design: page 18; Robustness: page 13)
  - Additional control comparison using Seattle (control context) to gauge target effects and heterogeneity; analyses include separate regressions for transaction vs. cardholder data. (Provenance: Seattle: page 15)

  Compact mathematical summary (key specification and interpretation)
  - Model (for outcome y_it, e.g., calories per store-day or per transaction): y_it = _i + _t + Post_it + X_it' + _it
    - Post_it = 1 if NYC store after 1 Apr 2008; _i are store fixed effects; _t are time controls (week/day, etc.); X_it includes controls (weather, etc.).
  - Mechanism decomposition:
    - Total calorie change decomposes into extensive margin (changes in quantity of food purchases) and intensive margin (calories per item). The findings indicate nearly 75% of the reduction comes from the extensive margin (fewer food purchases). (Provenance: Mechanism: page 18)

- Key results
  - Overall calories and behavior
    - Calorie posting reduced average calories per store-day by about 4.6 calories and coincided with an increase in NYC transaction counts (more visits). Robust to alternative weekly-time specifications. (Provenance: Overall impact: page 13)
  - Beverage vs. food items per transaction
    - Beverages per transaction: +0.005 (significant)
    - Food items per transaction: −0.029 (significant)
    - Beverages + food items: −0.027 (significant)

        * Calories per beverage (log): −0.008 (significant)
        * Calories per food item (log): −0.039 (significant)
        * Controls: store fixed effects, week/day dummies, weather; regressions for transaction and cardholder data. (Provenance: page 18)
    - Mechanism: extensive vs. intensive margins
        * Nearly three quarters of total calorie reduction attributed to people opting not to buy food items (extensive margin). (Provenance: page 18)
    - Heterogeneity across customers
        * Effects vary with demographics and customer types (e.g., income, education; high-frequency vs mid/high-calorie customers). (Provenance: page 21)
    - Quantile-based effects
        * Larger reductions at higher calories-per-transaction quantiles; ~5–6% reduction from the 75th to the 99th percentile; 10th percentile near zero. (Provenance: page 22)
    - Seattle/control context
        * In Seattle, drink calories per transaction fell by 4.6 (SE 0.3) while food calories per transaction rose by 0.8, supporting a targeted effect more pronounced for beverages and suggesting heterogeneity in response. (Provenance: page 15)
    - Robustness and interpretation
        * Time-varying effects and mealtime variation: beverage-calorie effects are slightly stronger around lunchtime but broadly uniform across the day. (Provenance: page 15)
    - Appendix/measurement notes
        * Appendix includes survey instruments for Starbucks customers (data collection methods). (Provenance: page 35-36)

- Limitations and caveats
    - Explicit limitations are not enumerated in the provided excerpts; however, generalizability may be constrained by the control context (Seattle) and the NYC-specific policy environment. (Provenance: Seattle control: page 15)
    - The results rely on observational data with fixed effects and time controls; external validity to other metros or policy designs may be limited. (Inferred from design details: page 18; cross-context discussion: page 15)

If you'd like, I can extract a compact mathematical summary with the exact regression equations and variable definitions.

Sub-Answers

Sub-answer 1

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

verbosity='medium')

- Question tokens estimate: `19`
- Answer:
- Background and motivation
  - Calorie posting is contested: it can be costly to require menus to post calories, with reported compliance costs around $2,000 per Wendy's store, and concerns about indirect costs from city-specific menus. NYC health dept stresses point-of-purchase availability. (Provenance: page 4, words 275-576)
- Overall impact on calories
  - Calorie posting reduced average calories per store-day by about 4.6 calories, and the policy coincided with an increase in transaction counts in NYC (i.e., more visits). The analysis uses a binary POST indicator (POST = 1 for NYC stores after 1 Apr 2008) and robust controls; alternative specs using separate weekly dummies yield similar timing patterns. (Provenance: page 13, words 275-528)
- Beverage vs. food items per transaction
  - Table 4 (regressions on items and calories per item):
    - Number of beverages per transaction: +0.005 (significant)
    - Number of food items per transaction: −0.029 (significant)
    - Beverages + food items: −0.027 (significant)
    - Calories per beverage (log): −0.008 (significant)
    - Calories per food item (log): −0.039 (significant)
    - Notes: regressions include store fixed effects, week/day dummies, weather controls; separate regressions for transaction vs. cardholder data. (Provenance: page 18, words 0-349)
- Mechanism: extensive vs. intensive margins
  - A large share of the total calorie reduction arises from fewer food purchases (the extensive margin) rather than lower calories per item. This is illustrated in the regression results and summarized as "nearly three quarters of the total calorie reduction can be attributed to people opting not to buy food items." (Provenance: page 18, words 0-349)
- Heterogeneity across customers
  - Table 5 shows how effects vary with demographics and customer types; e.g., interactions with income, education, and high-frequency vs. mid/high-calorie customers help explain variation in the posting effect. (Provenance: page 21, words 0-349)
- Quantile-based effects

  - Table 6 reports effects by calories-per-transaction quantile. Reductions are larger in higher quantiles; the percent reduction is roughly 5–6% from the 75th to the 99th percentile, while the 10th percentile shows minimal change (near zero, as expected). (Provenance: page 22, words 0-349)

- Seattle/control context
  - In Seattle (control context where many items lacked posted calories), drink calories per transaction fell by 4.6 with a standard error of 0.3, while food calories per transaction rose by 0.8 (negligible overall impact on food calories). This aligns with a target effect more pronounced for beverages and supports heterogeneity in response. (Provenance: page 15, words 0-254)

- Robustness and interpretation
  - The study also explores time-varying effects and mealtime variation; results show the beverage-calorie effect is slightly stronger around lunchtime but broadly uniform across the day. (Provenance: page 15, words 0-254)

- Appendix/measurement notes
  - Appendix includes survey instruments for Starbucks customers, illustrating data collection methods (not central to the main results). (Provenance: page 35-36, entire Appendix content)

If you'd like, I can extract a compact mathematical summary (e.g., key regression specification and dependent variables) with inline formulas.

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 (0.3159934745743112)`
- Retrieval method: `local`
- Evidence type: `explicit research-question statement`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia)
- Answer:

The main research question is whether mandatory calorie posting at chain restaurants affects consumers' behavior, specifically calories per transaction.

- Quote snippet:

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

- Citation anchors:
  - page=3, words=275-564, section=Introduction, note=Directly states the empirical question of whether calorie posting affects behavior.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.29764919438622633)`
- Retrieval method: `local`
- Evidence type: `data scale and mechanism analysis vs prior literature`
- Data source: Starbucks transaction data in NYC, Boston, Philadelphia; anonymous cardholder data; in-store surveys
- Answer:

The paper's primary contribution relative to the existing literature is establishing, with a much larger and richer dataset than prior NYC/short-run studies, that mandatory calorie posting reduces calories per transaction and that the mechanism includes both learning and salience, while also disaggregating intensive-margin vs composition effects and showing the effects across a broad scope (city, chain, and time) and on profitability.

- Quote snippet:

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

- Citation anchors:
  - page=9, words=0-349, section=Literature and data contributions, note=The authors emphasize their contribution lies in a dataset that is much larger and broader than prior NYC studies, including over 100 million transactions and individual-level data.
  - page=8, words=0-349, section=Literature and data contributions, note=Directly contrasts with prior studies and highlights the scale and breadth of their data as a key contribution.

A03: What is the central hypothesis being tested?

- Category: `A) Research question / contribution`
- Confidence: `low (0.29005770191274693)`
- Retrieval method: `local`
- Evidence type: `hypothesis`
- Answer:

The central hypothesis is that mandatory calorie posting affects consumer behavior by reducing calories per transaction (i.e., posting calories influences what consumers choose, with learning or salience as potential mechanisms).

- Quote snippet:

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

- Citation anchors:
  - page=3, words=275-564, section=Introduction / Hypothesis discussion, note=States the empirical question of whether calorie posting affects behavior.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.3044486824352848)`
- Retrieval method: `local`
- Evidence type: `textual evidence from the study defining dependent variables and reporting results`
- Data source: Transaction data and cardholder data from Starbucks calorie posting study
- Answer:

The main dependent variables are calories per transaction, specifically log(beverage calories per transaction), log(food calories per transaction), and log(beverage + food calories) (total calories per transaction).

- Quote snippet:

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

- Citation anchors:
  - page=11, words=0-349, section=Regression specification, note=Defines y_sct as calories per transaction and the POST dummy; basis for identifying effects within-city.
  - page=12, words=0-349, section=Dependent variables and results, note=Top row: log(beverage calories per transaction); second row: log(food calories per transaction); bottom row: log(beverage + food calories).

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.2439615861309941)`
- Retrieval method: `local`

- Evidence type: `model specification / treatment variable definition`
- Data source: Transaction data and cardholder data from Starbucks stores
- Answer:

The key treatment/exposure variable is the calorie-posting indicator (POST): a dummy equal to one for NYC stores on/after April 1, 2008 when calories were posted.

- Quote snippet:

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

- Citation anchors:
  - page=11, words=POS-posted, section=Regression specification / Equation (1), note=Defines the POST dummy as one if calories were posted (NYC stores after April 1, 2008).
  - page=11, words=y_sct- _sct, section=Equation (1) specification, note=Shows the regression form where POST is included as POS Tc t.

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

- Category: `A) Research question / contribution`
- Confidence: `medium (0.3532150662340793)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia) around NYC calorie-posting policy.
- Answer:

Country: United States; Market: chain-restaurant dining (Starbucks) in New York City, with Boston and Philadelphia as controls; Period: around the NYC calorie-posting policy implemented April 1, 2008, using data from January 2008 through February 2009 (pre- and post-policy).

- Quote snippet:

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

- Citation anchors:
  - page=1, words=0-349, section=Introduction, note=Contextualizes NYC calorie-posting policy and the national relevance.
  - page=5, words=0-349, section=Data, note=Describes dataset scope: NYC stores and controls in Boston/Philadelphia, pre/post period.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.2858951228579513)`
- Retrieval method: `local`
- Evidence type: `textual_evidence`
- Data source: Starbucks transaction data and cardholder data (NYC and control cities).
- Answer:

The main mechanism is a combination of learning and salience effects: calorie posting changes purchases because it reduces the cost of learning about calories (learning) and also reminds or highlights calories at the point of purchase (salience).

- Quote snippet:

  > 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=25, words=0-349, section=IV. A. Why is There an Effect?, note=Introduces learning vs. salience as competing explanations and mentions the possibility of a combined mechanism.
  - page=30, words=0-349, section=IV. Discussion, note=States that survey evidence supports salience and commuter evidence supports learning; concludes there is evidence for both mechanisms.

A08: What alternative mechanisms are discussed?

- Category: `A) Research question / contribution`
- Confidence: `low (0.22570611463556833)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: American Economic Journal: Economic Policy, Vol. 3 No. 1, Bollinger et al., calorie posting in chain restaurants
- Answer:

The alternative mechanisms discussed are learning effects and salience effects (and the possibility that both operate together).

- Quote snippet:

  > One reason why calorie posting may affect consumer choice is a learning effect. Another possible explanation is a salience effect; behavior may be driven by a combination of learning and salience effects.

- Citation anchors:

  - page=page 25, words=0-349, section=A. Why is There an Effect?, note=Discusses alternative mechanisms: learning vs salience (and their combination).

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

- Category: `A) Research question / contribution`
- Confidence: `medium (0.37698304853589215)`
- Retrieval method: `local`
- Evidence type: `Policy implications and conclusions`
- Data source: Starbucks NYC transaction data, Boston and Philadelphia control cities, anonymous cardholder data, and in-store surveys
- Assumption flag: `True`
- Assumption notes: Assumes that the detected calorie reductions, learning/salience effects, and potential long-run innovations translate into meaningful public health benefits and that these benefits outweigh the costs of implementing posting; generalizability to other chains is uncertain.
- Answer:

The main policy implications are that mandatory calorie posting yields a modest but persistent reduction in calories per transaction (about 6% at Starbucks), largely through lower food calories, with beverage calories largely unchanged; Starbucks profits are, on average, unchanged; the policy has very low implementation costs; there is potential for long-run benefits via menu innovation and public education, which could be larger than the short-run effects; and a key caveat is that the policy is deemed good only if the benefits outweigh the costs (i.e., the policy is not obviously valuable if costs exceed benefits).

- Quote snippet:

  > Mandatory calorie posting caused average calories per transaction to fall by 6 percent; the impact on Starbucks' profit is negligible on average, and costs are very low.

- Citation anchors:
  - page=34, words=0-349, section=IV. Discussion, note=Policy implications summarized in Discussion

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.31023471695865323)`
- Retrieval method: `local`
- Evidence type: `Discussion/Policy implications`
- Data source: Starbucks transaction data; NYC vs control cities; cardholder and store-level data
- Answer:

Mandatory calorie posting appears to generate welfare-relevant health effects with little cost to consumers or firms: calories per transaction fell (about 14% for food) while transaction frequency and Starbucks profits were largely unchanged, suggesting health benefits could come without major economic costs. However, obesity outcomes are not measured, and generalizability beyond Starbucks is limited.

- Quote snippet:

  > To summarize briefly, the analyses above show that mandatory calorie posting caused food calories per transaction to fall by 14 percent, on average, but had a negligible impact on beverage calories per transaction.

- Citation anchors:
  - page=24, words=275-564, section=IV. Discussion, note=Contains the welfare-relevant summary: calorie reductions with negligible profit impact; discusses margins and welfare implications

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.2760082674201887)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction data
- Answer:

The authors acknowledge two main limitations: (1) they do not directly measure the effect of calorie posting on obesity because BMI data are not available in a timely way; they rely on crude estimates from medical literature. (2) the analysis uses data from only one chain (Starbucks), so it is unknown whether the effects would generalize to other chains, and they cannot determine whether consumers offset calorie reductions at Starbucks by eating differently elsewhere (e.g., 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 and generalizability due to single-chain data.
  - page=3, words=275-564, section=Limitations, note=Offsetting behavior not observed; home eating changes unknown.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.30991989936869746)`
- Retrieval method: `local`
- Evidence type: `data/identification novelty`
- Data source: Starbucks transaction data (NYC, Boston, Philadelphia) and anonymous Starbucks cardholder data across the United States
- Answer:

The paper claims that its data are novel because the detailed transaction data are 'uniquely well-suited' to answer the question, and because it combines two data sources—transaction data and anonymous cardholder data—that let it identify how calorie posting affects individuals and isolate effects on the intensive and extensive margins from changes in consumer composition.

- Quote snippet:

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

- Citation anchors:
  - page=3, words=275-564, section=Data and limitations, note=States that the detailed transaction data are uniquely well-suited to answering this question.
  - page=5, words=0-349, section=Data description, note=Notes that the Starbucks data are rich enough to distinguish various responses and that the cardholder data allow assessment of individual behavioral changes.
  - page=5, words=275-591, section=Data advantages / Identification, note=Highlights the advantage of cardholder data to assess how individuals change behavior and to isolate effects on intensive/extensive margins from consumer composition.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2785943489056746)`
- Retrieval method: `local`
- Evidence type: `Difference-in-differences with city-week fixed effects and store/individual fixed effects to isolate intensive-margin effects.`
- Data source: Starbucks transaction data and cardholder data
- Table/Figure: Table 3; Figure 2
- Assumption flag: `True`
- Assumption notes: Assumes no differential shocks between NYC and control cities around the posting date; identification relies on within-city vari-

ation with appropriate fixed effects to isolate the posting effect.
- Answer:

Identification is achieved by exploiting within-city, time variation in calorie posting (NYC after April 1, 2008) and comparing it to Boston/Philadelphia controls using fixed-effects regressions to isolate the effect on calories per transaction (with store fixed effects for transactions and individual fixed effects for cardholders).

- 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; city-week level identification; NYC vs control cities.
  - page=12, words=0-349, section=Estimation design, note=Cardholder data uses individual fixed effects; transaction data uses store fixed effects; identification via within-city variation.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.30511875764717833)`
- Retrieval method: `local`
- Evidence type: `quasi-experimental`
- Data source: Starbucks transaction data and cardholder data across NYC, Boston and Philadelphia (with Seattle/Portland/SF used as controls in related analyses)
- Answer:

Quasi-experimental (a natural experiment) design, using NYC as the treatment group and control cities with pre/post analysis and regression.

- Quote snippet:

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

- Citation anchors:
  - page=page 11, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Regression specification defines a post-treatment indicator (calorie posting) and treatment/control groups, indicating a quasi-experimental design.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.29846298986247766)`
- Retrieval method: `local`
- Evidence type: `Natural experiment / difference-in-differences using the NYC calorie-posting policy as the treatment and Boston/Philadelphia as controls; within-city time variation identified by city-week effects.`
- Data source: Transaction data and cardholder data (Starbucks stores in NYC, Boston, and Philadelphia).
- Assumption flag: `True`
- Assumption notes: Parallel trends between NYC and control cities prior to posting are supported by a lack of pre-trend differences and are explicitly checked in the analysis.
- Answer:

The exogenous variation comes from the calorie-posting policy implemented in NYC on April 1, 2008, with Boston and Philadelphia serving as control cities; identification relies on within-city time variation around that posting date (store/week fixed effects), i.e., a city-level difference-in-differences design.

- 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=1-25, section=Identification strategy, note=POS Tc is a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008).
  - page=12, words=0-30, section=Identification, note=Identification stems from within-city variation over time.
  - page=6, words=150-210, section=Data Summary, note=Qualitatively, Boston and Philadelphia are reasonable controls for NYC.
  - page=14, words=60-110, section=Results / Robustness, note=The policy variation is at the city-week level; within-city variation identifies the effect.
  - page=14, words=280-320, section=Seattle robustness, note=Seattle provides robustness check with post-law data.

B04: What is the treatment definition and timing?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.13185219280876276)`
- Retrieval method: `local`
- Evidence type: `regression specification description`

- Data source: Starbucks transaction and cardholder data in NYC, Boston, and Philadelphia; Bollinger et al., American Economic Journal: Economic Policy (2011)
- Answer:

The treatment is the mandatory calorie posting in New York City stores, defined by a post-treatment indicator that equals 1 for NYC stores on days after April 1, 2008 (i.e., after calorie posting began).

- Quote snippet:

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

- Citation anchors:
  - page=11, words=0-349, section=Methods: Regression specification, note=Defines POST as a dummy for calories posted after April 1, 2008 in NYC stores.

B05: What is the control/comparison group definition?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.1800030198446593)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Starbucks transaction data and cardholder data
- Answer:

The control/comparison group consists of Boston and Philadelphia (the control cities for NYC); in robustness checks, Seattle is compared with Portland and San Francisco as controls.

- Quote snippet:

  Boston and Philadelphia are reasonable controls for NYC.

- Citation anchors:
  - page=6, words=0-349, section=Table 1: Summary Statistics, note=Boston and Philadelphia are reasonable controls for NYC.
  - page=14, words=275-350, section=Robustness/Seattle comparison, note=Seattle analysis uses control cities Portland and San Francisco.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.308491997900624)`
- Retrieval method: `local`
- Evidence type: `regression equation described in text`

- Data source: Starbucks transaction and cardholder data used in Bollinger et al., 'Calorie Posting in Chain Restaurants' (AEJ Economic Policy, 2011)
- Assumption flag: `True`
- Assumption notes: Identification relies on within-city variation over time; assumes posting date exogenously determines treated period; controls for seasonality/weather and uses fixed effects to address unobserved heterogeneity.
- Answer:

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); x_sct includes week fixed effects, day-of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared. In the transaction data specification, store fixed effects are included (and in the cardholder data specification, individual fixed effects).

- 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

- Citation anchors:
  - page=page 11, words=275-522, section=Methods / Empirical strategy, note=Equation (1) defining the baseline regression

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2544411011640151)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Transaction data and cardholder data (Starbucks calorie posting study)
- Answer:

They include unit fixed effects for stores (store fixed effects) and, for cardholder data, individual fixed effects; time-related fixed effects include week fixed effects, plus day-of-week and holiday dummies (seasonality controls). In robustness checks they also use date fixed effects, which replace the day-of-week/week controls.

- Quote snippet:

  > store fixed effects to control for all time-invariant, store-specific heterogeneity.

- Citation anchors:

  - page=12, words=0-80, section=Store fixed effects, note=Store fixed effects to control for all time-invariant, store-specific heterogeneity; notes on city-week variation.
  - page=11, words=0-60, section=Time fixed effects, note=Week fixed effects to control for seasonality; day-of-week dummies and holiday dummies included.
  - page=12, words=0-60, section=Cardholder fixed effects, note=Individual consumer fixed effects used with cardholder data.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2808639424011974)`
- Retrieval method: `local`
- Evidence type: `table note`
- Data source: American Economic Journal: Economic Policy (Feb 2011)
- Table/Figure: Table 9
- Answer:

Robust standard errors; they are not clustered and are reported at the transaction (observation) level.

- Quote snippet:

  Robust standard errors in parentheses.

- Citation anchors:
  - page=29, words=0-349, section=Table 9, note=Robust standard errors in parentheses.
  - page=29, words=0-349, section=Table 9, note=An observation is a transaction.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2723139807629138)`
- Retrieval method: `local`
- Evidence type: `parallel trends / diff-in-differences assumption`
- Data source: Difference-in-differences analysis using NYC stores after April 1, 2008 and Boston/Philadelphia controls (transaction and cardholder data).
- Assumption flag: `True`
- Assumption notes: Identification relies on parallel trends prior to calorie posting; NYC and controls show no differential pre-trends, supporting a valid counterfactual. The analysis uses fixed effects and controls to address other confounds.

- Answer:

The key identifying assumption is the parallel trends (difference-in-differences) assumption: in the absence of calorie posting, NYC and the control cities (Boston/Philadelphia) would have followed similar trends in calories per transaction.

- 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=Figure 2 discussion / pre-trend, note=No evidence of pre-trend differences between NYC and Boston/Philadelphia.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.24036484632117797)`
- Retrieval method: `local`
- Evidence type: `Parallel-trends/diff-in-diff evidence; timing alignment with posting; cross-dataset robustness; within-individual variation`
- Assumption flag: `True`
- Assumption notes: Supports the identifying assumption via parallel trends prior to posting, timing alignment with posting, and cross-dataset corroboration; robust controls and within-individual variation address potential confounds.
- Answer:

Evidence supporting the identifying assumption includes (i) no pre-trend differences between NYC and control cities prior to calorie posting, and (ii) an immediate drop in calories per transaction after posting, corroborated by both transaction and cardholder data, with robustness checks across datasets.

- 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.

- Citation anchors:
  - page=14, words=0-349, section=Identifying assumption / robustness (pre-trends and timing), note=States no pre-trend differences NYC vs controls; drop around posting date; corroborated by two datasets.

- page=12, words=275-624, section=Using cardholder vs transaction data to isolate intensive margin; addresses confounds, note=Cardholder data isolates intensive-margin effects; argues against confounding changes in composition; within-individual variation.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.31798632269597993)`
- Retrieval method: `local`
- Evidence type: `event-study / pre-trend tests described in the paper`
- Data source: Cardholder and transaction data from Starbucks across Boston, NYC, and Philadelphia (prior to NYC calorie posting).
- Table/Figure: Figure 2
- Answer:

Yes. The study includes event-study/pre-trend tests showing no evidence of pre-trend differences between NYC and the control cities (Boston/Philadelphia); and the calorie-per-transaction drop occurs right after the posting date (April 1, 2008), persisting thereafter.

- Quote snippet:

  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.

- Citation anchors:
  - page=11, words=275-522, section=Figure 2 pre-trend discussion, note=No evidence of pre-trend differences between NYC and Boston/Philadelphia.
  - page=13, words=275-528, section=Event timing and impact discussion, note=Vertical line at April 1, 2008; drop in calories per transaction occurs around posting date.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.09979825575888589)`
- Retrieval method: `local`
- Evidence type: `methodology`
- Answer:

No instrumental variables are used. The analysis relies on a difference-in-differences-like design with a POST indicator for NYC after Apr 1, 2008, com-

bined with store fixed effects and other controls; therefore there is no first-stage for an instrument.

- 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=Methods/Results, note=Definition of POST and binary treatment; no instrument mentioned.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.24513026966018944)`
- Retrieval method: `local`
- Evidence type: `not_reported_in_context`
- Data source: Starbucks transaction data and cardholder data
- Answer:

Not reported in the provided context.

- Quote snippet:

  We estimate regressions of the following form: (1) ysct = xsct + POSTct + sct

- 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 (0.22343987037482763)`
- Retrieval method: `local`
- Evidence type: `method_description`
- Data source: Transaction data and cardholder data

- Answer:

The paper does not use a regression discontinuity design. There is no running variable or cutoff for an RDD; instead, the analysis employs a post-treatment indicator (calorie posting after April 1, 2008) within a difference-in-differences framework.

- Quote snippet:

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

- Citation anchors:
  - page=11, words=To-2008), section=A. Calories Per Transaction, note=Describes the regression specification with a post-treatment indicator; defines the NYC posting date of April 1, 2008

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.28889020046602576)`
- Retrieval method: `local`
- Evidence type: `methodology description`
- Data source: Transaction data and cardholder data
- Table/Figure: Figure 2; Table 3
- Assumption flag: `True`
- Assumption notes: There is no evidence of pre-trend differences between NYC and the control cities (First, with both datasets we see no evidence of pre-trend differences between NYC and Boston/Philadelphia), supporting the parallel trends assumption for the DiD design.
- Answer:

Staggered adoption across cities: NYC stores began calorie posting on April 1, 2008, and Seattle began on January 1, 2009. The estimator is a Difference-in-Differences regression with a post-treatment dummy (POST_Tc t) and city/week fixed effects (a two-way FE DiD).

- 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=Model specification (DiD regression), note=Defines the regression y_sct = x_sct + POST_Tc t + _sct and the POST_Tc t dummy (calories posted) for NYC after 2008-04-01.

  - page=16, words=275-558, section=Staggered adoption across cities, note=Notes Seattle's calorie posting law dated 2009-01-01, illustrating staggered adoption vs NYC.
  - page=13, words=275-528, section=Alternative specification, note=Describes an alternative approach with separate week dummies for NYC and controls to capture timing differences.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.28023930128603075)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Starbucks transaction data (NYC, Boston, Philadelphia) and anonymous Starbucks cardholder data; Seattle transaction data; in-store surveys
- Answer:

The study uses two primary datasets: (1) the transaction data from all Starbucks company-owned stores in NYC, Boston, and Philadelphia; and (2) cardholder data consisting of anonymous Starbucks cardholders tracked over time (with NYC pre/post posting). It also references a Seattle transaction dataset used as a robustness check and in-store survey data.

- Quote snippet:

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

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary, note=Definition of the two datasets: transaction data and cardholder data.
  - page=14, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Seattle dataset mentioned as robustness check (post-law period) and within-city analyses.
  - page=25, words=275-599, section=Appendix: Survey, note=In-store customer surveys (792 completed) used as additional data.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3225688568722501)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Calorie posting study using transaction and cardholder data (Bollinger et al., American Economic Journal: Economic Policy, 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, note=Describes observing all transactions at each location; unit of observation is transaction.

C03: What is the sample period and geographic coverage?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.23067771097921516)`
- Retrieval method: `local`
- Evidence type: `data_summary`
- Data source: Transaction data from Starbucks stores in NYC, Boston, and Philadelphia
- Answer:

Sample period: January 1, 2008 to February 28, 2009; geographic coverage: NYC plus Boston and Philadelphia (all Starbucks company-owned stores in NYC and all 94 stores in Boston and Philadelphia).

- Quote snippet:

  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=States sample: 222 NYC stores and 94 stores in Boston/Philadelphia; period = 3 months pre and 11 months post calorie posting (Jan 1, 2008–Feb 28, 2009).

C04: What are the sample restrictions / inclusion criteria?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2451827001199968)`
- Retrieval method: `local`
- Evidence type: `data description / sample selection criteria`
- Data source: Starbucks transaction data and anonymous cardholder data (NYC, Boston, Philadelphia)
- Answer:

- Transaction data: includes all transactions from 222 NYC Starbucks stores and 94 stores in Boston and Philadelphia during the period January 1, 2008 to February 28, 2009; stores not open for the entire period are excluded and transactions with more than four units of any one item are excluded (balanced panel). - Cardholder data: a subsample of anonymous cardholders who averaged at least one transaction per week in one of NYC, Boston, or Philadelphia in the pre-calorie-posting period (7,520 in NYC; 3,772 in Boston/Philadelphia; combined 1.51 million transactions). These cardholders may not be representative of Starbucks customers in general. - Normalization for confidentiality is applied to NYC values in the tables/presentation.
- 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 / Sample scope, note=Transaction data cover all NYC stores and all Boston/Philadelphia stores; period before and after calorie posting; balanced panel.
  - page=5, words=275-591, section=Cardholder data sample, note=Definition of subsample: individuals averaging 1 transaction per week in pre-posting period; counts: 7,520 NYC; 3,772 in Boston/Philadelphia; 1.51 million transactions total.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2978690009143336)`
- Retrieval method: `local`
- Evidence type: `textual_excerpt`
- Data source: transaction data and cardholder data from Starbucks NYC, Boston, and Philadelphia
- Answer:

1.51 million transactions

- Quote snippet:

  > combined 1.51 million transactions for us to study.

- Citation anchors:
  - page=5, words=275-591, section=Data Summary, note=Statement indicating combined 1.51 million transactions in dataset used for main analysis.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3268006721752479)`
- Retrieval method: `local`
- Evidence type: `Description of measurement and transformations of the primary outcome from the study`
- Data source: Transaction data and cardholder data (Starbucks; NYC, Boston, Philadelphia)
- Table/Figure: Table 3
- Answer:

The key outcome is calories per transaction (including beverage calories per transaction, food calories per transaction, and total calories per transaction). It is measured as calories per transaction at store-day observations, and in some specifications is modeled using log transformations (e.g., log(beverage + food calories) or log(calories per transaction)). Coefficients are interpreted as percentage changes (e.g., a 11.2% decrease) when the dependent variable is log-transformed. Additionally, a quantile-based approach regresses the log of the nth quantile of calories per transaction.

- Quote snippet:

  ysct = xsct + POST Tc t + sct, where ysct 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.

- Citation anchors:
  - page=5, words=275-591, section=Data and variables; primary outcome, note=Describes calories per transaction as the key variable and the datasets used.
  - page=11, words=275-522, section=Regression specification, note=Defines ysct as calories per transaction and the POST dummy; shows regression form.
  - page=12, words=275-624, section=Results with logs, note=Reports log specifications (log(beverage + food calories)) and percent-change interpretations (e.g., −0.119 corresponds to an 11.2% decrease).
  - page=22, words=0-349, section=Quantile analysis, note=Explains regressing the log of the nth quantile of calories per transaction.
  - page=6, words=275-517, section=Baseline calorie measures, note=Provides baseline calories per transaction (drink, food, total) prior to posting.

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

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

- Confidence: `low (0.29227057615324575)`
- Retrieval method: `local`
- Evidence type: `Description of treatment/exposure construction (POST dummy) and identification strategy`
- Data source: Transaction dataset and cardholder dataset (Starbucks data)
- Assumption flag: `True`
- Assumption notes: Identification relies on within-city (city-week) variation over time; fixed effects control for seasonality and other factors; assumes NYC post period is exogenous to other changes within city-week.
- Answer:

Treatment is measured with a post-treatment indicator POST that equals 1 for NYC stores on every day after April 1, 2008 when calories were posted. Exposure is identified from within-city time variation (city-week level). An alternative specification uses separate city-week dummies and omits POST to flex timing. Constructed variables include POST, store fixed effects, week fixed effects, day-of-week, holidays, and weather controls; the analysis uses both a transaction dataset (store-day level) and a cardholder dataset (individual fixed effects).

- Quote snippet:

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

- Citation anchors:
  - page=11, words=0-349, section=Regression specification, note=POST dummy defined as a daily indicator for NYC stores after calories are posted
  - page=12, words=0-349, section=Identification strategy, note=Identification relies on within-city variation; policy variation is at the city-week level
  - page=12, words=0-349, section=Robustness/alternative specification, note=Alternative approach uses separate week dummies for NYC and controls and excludes POST

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.28700928116835767)`
- Retrieval method: `local`
- Evidence type: `regression_controls`
- Data source: Transaction data and cardholder data
- Answer:

Yes. The key covariates/controls are week fixed effects, day-of-week dummies, holiday dummies, and weather variables (temperature and temperature squared,

precipitation and precipitation squared). These controls are part of the regression specification (x_sct) and are used consistently across the analyses with both transaction data and cardholder data, i.e., they are always included.

- Quote snippet:

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

- Citation anchors:
  - page=11, words=275-522, section=Regression specification, note=Explicitly lists controls: week fixed effects, day-of-week dummies, holiday dummies, temperature and temperature squared, precipitation and precipitation squared.
  - page=12, words=275-624, section=Regression discussion, note=Notes that weather controls are included and that the same controls are used for both transaction and cardholder data.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.27792692287479065)`
- Retrieval method: `local`
- Evidence type: `descriptive data handling in methods`
- Data source: Starbucks transaction data and cardholder data
- Answer:

Missing data are handled by dropping observations to form a balanced panel: observations from stores not open for the entire data period are excluded, and transactions with more than four units of any one item are dropped. There is no imputation or weighting described; about 2.2% of transactions are excluded.

- Quote snippet:

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

- Citation anchors:
  - page=6, words=0-349, section=Data cleaning / Balanced panel, note=States that they exclude stores not open during the entire data period to form a balanced panel and exclude transactions with more than four units of any one item; 2.2% of transactions are excluded.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.29540471022143683)`
- Retrieval method: `local`
- Evidence type: `methodology description`
- Data source: transaction data and cardholder data (Starbucks NYC, Boston, Philadelphia)
- Answer:

No weights are described or used; the analysis relies on regression specifications with fixed effects and controls, without mention of survey weights or propensity weights.

- Quote snippet:

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

- Citation anchors:
  - page=11, words=0-349, section=Methodology: Regression specification, note=States regression form and controls; no weighting mentioned.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.22246639621182168)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Transaction data; Cardholder data (anonymous) from Starbucks across NYC, Boston, Philadelphia
- Answer:

The data are not merged into a single dataset; instead two sources are used—the transaction data (store-level transactions across NYC, Boston, and Philadelphia) and the anonymized cardholder data (individual-level purchase histories). Linkage is performed at the cardholder level by defining a subsample of cardholders who averaged at least one transaction per week in NYC, Boston, or Philadelphia, allowing analysis of within-individual changes across both data sources; results from the two sources are then compared rather than merged.

- Quote snippet:

  There are 2.7 million anonymous individuals in this dataset.

- Citation anchors:

  - page=5, words=0-349, section=Data sources and linkage, note=Introduces two datasets (transaction data and cardholder data) and the existence of 2.7 million anonymous individuals; describes the cardholder data and its use.
  - page=5, words=275-591, section=Data sources and linkage, note=Notes a subsample of cardholders (7,520 NYC and 3,772 Boston/Philadelphia) and 1.51 million transactions; highlights cross-dataset analysis.
  - page=5, words=0-349, section=Data sources and linkage, note=States that the advantage of cardholder data is to assess how calorie information causes particular individuals to change behavior; frames comparison across datasets.

C12: What summary statistics are reported for main variables?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3270899297372985)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Transaction data and Cardholder data (prior to calorie posting in NYC, April 1, 2008)
- Table/Figure: Table 1
- Answer:

Table 1 reports summary statistics for main variables, including: avg. weekly transactions per store, avg. weekly revenue per store, percent transactions with brewed coffee, percent transactions with beverage, percent transactions with food, avg. number of items per transaction, avg. number of drink items per transaction, avg. number of food items per transaction, food attach rate, avg. dollars per transaction, avg. calories per transaction, avg. drink calories per transaction, and avg. food calories per transaction.

- 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)

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2833644669735172)`
- Retrieval method: `local`
- Evidence type: `descriptive figures and table`

- Data source: Starbucks transaction data and cardholder data for NYC, Boston, and Philadelphia
- Table/Figure: Figure 1; Figure 2; Table 1
- Answer:

Yes. The article uses descriptive figures to establish baseline patterns, including Figure 1 showing weekly average calories per transaction by city and Figure 2 comparing NYC to controls (no pre-trend differences), and a Table 1 with summary statistics.

- Quote snippet:

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

- Citation anchors:
  - page=9, words=0-349, section=II.A Calories Per Transaction, note=Figure 1: average calories per transaction per week, NYC vs control cities.
  - page=14, words=0-349, section=II.A Calories Per Transaction, note=Figure 2: results by dataset; discussion of pre-trend differences.
  - page=11, words=275-522, section=Figure 2 discussion, note=No evidence of pre-trend differences between NYC and controls.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.42785311918531443)`
- Retrieval method: `local`
- Evidence type: `Regression estimate from calorie posting study (Table 3; log-beverages + foods calories)`
- Data source: Starbucks store-day transaction data and cardholder data
- Table/Figure: Table 3
- Answer:

-5.8% decrease in average calories per transaction

- Quote snippet:

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

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

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.18122989848182008)`
- Retrieval method: `local`
- Evidence type: `Regression specification (methodology)`
- Data source: Starbucks transaction data and Starbucks cardholder data (NYC, Boston, Philadelphia)
- 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, POS_Tc t is a dummy for when calories were posted (NYC stores after April 1, 2008), and x_sct includes week fixed effects, day-of-week dummies, holiday dummies, temperature and temperature squared, and precipitation and precipitation squared. It is estimated separately for the transaction data and the cardholder data to control for seasonality and other influences and isolate the effect of calorie posting on calories per transaction.

- Quote snippet:

  > we estimate regressions of the following form: (1) y_sct = x_sct + POS Tc t + _sct, where y_sct is a measure of calories per transaction...

- Citation anchors:
  - page=11, words=0-349, section=Regression specification, note=Equation (1) and description of y_sct, POS_Tc t, and controls including week fixed effects, day-of-week, holidays, and weather.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.4587002143915059)`
- Retrieval method: `local`
- Evidence type: `Empirical estimates from transaction and cardholder data in a natural-experiment on calorie posting`
- Data source: Starbucks calorie posting study data (NYC and control cities); includes transaction data and cardholder data
- Table/Figure: Table 3; Table 4; Table 6; Figure 3; Figure 4
- Assumption flag: `True`
- Assumption notes: Extrapolates potential obesity impact using a crude calculation assuming a 6% reduction in calories at all chain restaurants and that 25% of calories come from chain restaurants
- Answer:

The effect is modest overall but economically meaningful in parts. Beverages per transaction fell by about 0.3% (not robust in all samples), while food calories

per transaction declined by 13.7% in the transaction data. Combined (beverage + food), average calories per transaction dropped ~5.8% (14.4 calories). Using cardholder data, beverage calories show no significant change, food calories per transaction decline by about 11.2%, and total calories fall ~5.0%. About three quarters of the total calorie reduction comes from the extensive margin (fewer food purchases), with larger per-item reductions concentrated among higher-calorie purchases (quantile results show roughly 5–6% reductions from the 75th to 99th percentiles; high-calorie purchasers saw around a 26% reduction). The authors also document that average price per item rose after posting and profits were not significantly affected. They present a crude back-of-the-envelope calculation to assess whether a 6% reduction at all chain restaurants could meaningfully affect obesity, but do not give a definitive conclusion in the excerpt.

- Quote snippet:

  calorie posting caused a trivial decrease in beverage calories per transaction of 0.3 percent.

- Citation anchors:
  - page=12, words=0-349, section=Table 3 (transaction data), note=Beverage calories per transaction −0.3%
  - page=12, words=0-349, section=Table 3 (transaction data), note=Food calories per transaction −13.7%
  - page=12, words=0-349, section=Table 3 (transaction data), note=Total calories per transaction (log) −0.063 (~−5.8%)
  - page=12, words=275-624, section=Table 3 (cardholder data), note=Cardholder data: beverage calories not significantly different; food calories per transaction −11.2%; total calories −5.0%
  - page=16, words=275-558, section=Table 4 / discussion, note=Three quarters of total calorie reduction due to extensive margin (not buying food items)
  - page=22, words=0-349, section=Table 6 (quantiles), note=High-calorie purchases show ~26% reduction; percent change roughly 5–6% from 75th to 99th percentile
  - page=24, words=275-564, section=IV. Discussion, note=Price per item purchased increased; profits not significantly affected
  - page=20, words=0-349, section=Results, note=No statistically significant change in purchase frequency in NYC vs control cities
  - page=24, words=275-564, section=IV. Discussion, note=Crude calculation: if calories declined 6% at all chain restaurants, given chain restaurants account for ~25% of calories, the authors discuss potential obesity impact

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

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

- Confidence: `low (0.2836027021129647)`
- Retrieval method: `local`
- Evidence type: `robustness_checks`
- Data source: American Economic Journal: Economic Policy (February 2011)
- Table/Figure: Table 3; Figure 2; Table 5; Table 9
- Answer:

Key robustness checks and whether results survive: (1) Date fixed effects vs week/day controls: adding date fixed effects leaves the transaction-calorie effects virtually unchanged (beverage calories: −0.004; food calories: −0.152; total: −0.063). (2) Absolute calories specification: results are almost identical to the log-specifications. (3) Serial correlation robustness: aggregating pre/post periods yields the same conclusions with significance >99%. (4) Weather controls robustness: findings hold even when weather controls are dropped. (5) No pre-trend evidence: Figure 2 shows no pre-trend differences between NYC and controls. (6) Milk usage data: no significant impact on milk, supporting the calorie-posting story. (7) Heterogeneity checks: larger calorie reductions in higher-income and more-educated areas; female cardholders more responsive. (8) Commuters vs non-commuters: patterns persist outside NYC for some groups, suggesting learning vs salience effects are nuanced but overall calorie reductions remain. Overall, the main findings are robust across these alternative specifications and samples.

- 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=0-60, section=robustness_date_fixed_effects, note=As a robustness check, we include date fixed effects; transaction-data estimates barely changed.
  - page=12, words=61-110, section=robustness_absolute_calories, note=Absolute calories specification yields almost identical results to the log specification.
  - page=12, words=111-170, section=robustness_serial_correlation, note=To address serial correlation, pre/post aggregation still shows similar effects and significance >99%.
  - page=14, words=0-60, section=robustness_pre_trend, note=Figure 2 shows no pre-trend differences between NYC and controls.
  - page=11, words=275-311, section=robustness_weather_controls, note=Weather controls have an insignificant impact; results unchanged if weather controls are excluded.
  - page=20, words=0-60, section=robustness_milk_data, note=Milk usage data show no statistically significant impact, supporting no broad consumption shift.

  - page=25, words=275-320, section=robustness_heterogeneity, note=Heterogeneity: stronger effects in higher-income/educated areas; females more responsive.
  - page=29, words=0-60, section=robustness_commuters, note=Commuter vs non-commuter patterns: differences observed but overall calorie reductions persist in subgroups.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.21232835402329395)`
- Retrieval method: `local`
- Evidence type: `placebo tests (pre-trend check; seasonal/placebo check; Seattle non-posting city test)`
- Data source: Cardholder data and transaction data from Starbucks; Seattle/NYC comparisons
- Table/Figure: Figure 2
- Assumption flag: `True`
- Assumption notes: Placebo tests rely on parallel trends between NYC and control groups prior to posting; lack of pre-trend supports causal interpretation
- Answer:

They run placebo tests to check robustness of the calorie-posting effects: (1) a pre-treatment (pre-trend) test comparing NYC with control cities before posting shows no pre-trend differences, (2) a seasonal/placebo check around Christmas/New Year shows NYC is not different from controls during that period, suggesting the main effect is not driven by seasonality, and (3) a Seattle/non-posting-city test shows no significant impact, supporting that the observed effects arise from posting in NYC rather than other factors.

- 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=Results / Figure 2 robustness checks, note=No evidence of pre-trend differences between NYC and control cities; post-post effect aligns with posting date.
  - page=14, words=0-349, section=Results / Figure 2 robustness checks, note=Seasonal period (Christmas/New Year) shows NYC behavior not different from controls, indicating results are not driven by seasonality.
  - page=25, words=0-349, section=Seattle placebo test, note=Seattle (no posting) shows no significant impact, supporting causal interpretation.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2570038258376549)`
- Retrieval method: `local`
- Evidence type: `falsification_outcome`
- Answer:

The falsification outcomes tested (unaffected outcomes) include the frequency of cardholders' purchases in NYC relative to control cities, which showed no significant change, and the milk usage data, which also showed no statistically significant impact.

- Quote snippet:

  > We found no statistically significant change in the frequency of cardholders' purchases in NYC relative to the control cities; indeed, the point estimate of the coefficient on calorie posting was very close to zero.

- Citation anchors:
  - page=20, words=0-349, section=B. Substitution Effects, note=Tests the outcome that purchases frequency did not change.
  - page=20, words=0-349, section=B. Substitution Effects, note=Milk usage data showed no statistically significant impact.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.37464915727731324)`
- Retrieval method: `local`
- Evidence type: `Heterogeneity results from Tables 5 and related text discussing income, baseline calories, and regional controls; plus Table 2 examples of size-related changes.`
- Data source: Transaction data and cardholder data from the Starbucks calorie posting study
- Answer:

Heterogeneity results show: (i) by income, the calorie-posting effect is larger in higher-income zip codes (Posting $\times$ median income $= -0.012$, significant); (ii) by baseline exposure, higher-baseline-calorie customers exhibit larger reductions (medium-calorie: -0.298; high-calorie: -0.444; and higher-percentile analyses indicate bigger absolute drops for those with higher pre-posting calories); (iii) by size, the data indicate size-related shifts in beverage choices (e.g., 8.4% in NYC switched to a smaller size while 11.6% switched to a larger size); (iv) by region, using NYC as the treatment with Boston/Philadelphia as controls, there is no statistically significant change in purchase frequency in NYC relative to

controls, though beverage-choice effects were observed, with female cardholders more responsive than males.

- Quote snippet:

  Posting × median income (in $100,000) −0.012**; Posting × high calorie customer −0.444***; Philadelphia as controls. We found no statistically significant change in the frequency of cardholders' purchases in NYC relative to the control cities.

- Citation anchors:
  - page=21, words=Table 5—Heterogeneity-Notes, section=Heterogeneity in the Impact of Calorie Posting, note=Intro to heterogeneity results including income and baseline exposure interactions
  - page=21, words=Posting × median income-0.012**, section=Table 5 results, note=Income interaction with posting
  - page=21, words=Posting × high calorie customer-0.007, section=Table 5 results, note=Baseline exposure interaction showing larger reductions for high-calorie customers
  - page=21, words=Posting × high calorie customer-−0.444***, section=Table 5 results, note=Largest negative interaction for high-baseline-calorie customers
  - page=11, words=Smaller size-drink, section=Table 2, note=Illustrates size-related changes in beverage choices (NYC vs control)
  - page=11, words=8.4 percent-size drink, section=Table 2, note=NYC share switching to smaller size; contrasts with larger-size switching
  - page=20, words=Philadelphia as controls-control cities, section=Region analysis, note=Region-specific comparison shows no statistically significant change in purchase frequency in NYC relative to controls

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.14262024086930683)`
- Retrieval method: `local`
- Evidence type: `Regression-based mechanism test using commuter status and location interactions to distinguish salience vs learning (Table 9; post-period interactions).`
- Data source: Starbucks transaction and cardholder data; NYC stores and control cities (Boston, Philadelphia).
- Table/Figure: Table 9
- Answer:

The authors test whether the calorie-posting mechanism operates through salience or through learning by examining commuters (who shop both in NYC and outside NYC) and non-commuters. They estimate a regression with post-period indicators and interactions by purchase location (NYC vs

outside NYC) and commuter status. The key finding is that commuters reduce calories per transaction in NYC stores by about 7.7% and also show a sizable (12.0%) reduction outside NYC (though less precise). Since the effect extends to non-posted locations, the results support a learning mechanism rather than a pure salience effect.

- Quote snippet:

  > The estimates indicate that commuters reduced their calories per transaction in NYC stores by roughly 7.7 percent, which is similar to the reduction for non-commuters in NYC (6.0 percent). We estimate the effect on commuters' non- NYC transactions to be even larger (12.0 percent reduction).

- Citation anchors:
  - page=29, words=275-442, section=IV, note=Mechanism test: Table 9 results show commuter effects in NYC and outside NYC after calorie posting; supports learning over salience.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.3591894683289965)`
- Retrieval method: `local`
- Evidence type: `Robustness checks / sensitivity analyses`
- Data source: Transaction and cardholder data from Starbucks stores in NYC, Boston, and Philadelphia
- Answer:

Robustness checks show the results are not sensitive to alternative samples, bandwidths, or controls: analyses using both transaction data and cardholder data; tests with absolute calories; aggregation pre/post to address serial correlation; and excluding weather controls. No pre-trend differences between NYC and controls were found, and the main effects persist across samples, though magnitudes differ (beverage calories often not significant in cardholder data; food calories -11.2%; total calories -5%).

- Quote snippet:

  > The cardholder data allows us to isolate the effect due to changes in the intensive margin from the changes in consumer composition. The estimated effect on beverage calories is not significantly different from zero.

- Citation anchors:
  - page=page 11, words=0-349, section=Robustness checks (sample/data), note=Versions estimated separately with the transaction data and the cardholder data.

- page=page 12, words=275-624, section=Robustness, note=Aggregating pre/post calorie posting to address serial correlation; using absolute calories yields similar results.
- page=page 12, words=0-349, section=Robustness - beverage/food calories, note=Cardholder data: beverage calories not significantly different; food calories −11.2%.
- page=page 12, words=275-624, section=Robustness - alternative dependent variable, note=We repeated the analysis using absolute calories; results almost identical.
- page=page 11, words=0-349, section=Weather controls robustness, note=Findings unchanged if we exclude weather controls.
- page=page 14, words=0-349, section=Pre-trend & robustness, note=No evidence of pre-trend differences; estimates are similar across samples.
- page=page 14, words=0-349, section=General robustness, note=We obtain similar estimates of the effect of calorie posting.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2695900717709611)`
- Retrieval method: `local`
- Evidence type: `empirical analysis of transaction data and survey evidence; policy implications`
- Data source: Starbucks transaction data; cardholder purchase records (NYC vs non-NYC; commuters)
- Answer:
- Mandatory calorie posting reduces average calories per transaction by 6% at Starbucks, with the effect lasting over time and driven mainly by changes in food purchases rather than beverages.
- The effect is larger for high-calorie consumers.
- Learning and salience both contribute to the effect: surveys show increased sensitivity to calories, and some reductions occur in non-posting stores after NYC exposure, especially for commuters who had prior NYC encounters.
- Profit impact is negligible on average; in some nearby-location contexts (e.g., near Dunkin Donuts), revenue may even rise.
- Crude estimates suggest only modest potential effects on obesity, but long-run effects could be larger if restaurants offer more low-calorie options and public education benefits exist.

- Policy costs are very low, so even small benefits could outweigh the costs; findings may not generalize beyond Starbucks since the study covers only one chain.
- Overall conclusion: calorie posting has measurable but modest effects with mixed mechanisms and important implications for future menu Innovation and public awareness.
- 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/Discussion, note=Core conclusion: 6% drop in calories per transaction, long-lasting, learning and salience both contribute.
  - page=31, words=0-349, section=Is the Effect Big Enough to Matter?, note=Back-of-the-envelope calculation about potential obesity impact; overall small effect.
  - page=30, words=0-349, section=Commuters analysis, note=Learning vs. salience evidence from commuters; NYC exposure effects on non-NYC stores.
  - page=25, words=0-349, section=Why is There an Effect?, note=Explains learning and salience explanations; survey approach.
  - page=24, words=275-564, section=IV. Discussion, note=Overall discussion and summary of results.
  - page=38, words=275-448, section=Policy implications, note=Long-run potential benefits and public education considerations.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.27341031138679683)`
- Retrieval method: `local`
- Evidence type: `literature_review`
- Answer:

The most important prior papers cited are those that established and tested how nutrition/calorie information at the point of purchase affects consumer choices and, in some studies, health outcomes. Key works include Bassett et al. (2008), which provides NYC fast-food calorie information data; Currie et al. (2010), which links fast-food exposure to obesity outcomes; Downs, Loewenstein, and Wisdom (2009), which discuss strategies for promoting healthier choices; Ippolito and Mathios (1990, 1995), foundational analyses on information and health choices; Variyam and Cawley (2006), a classic treatment of nutrition labels and obesity; Kiesel and Villas-Boas (2008), which experiments with nu-

tritional shelf labels; Jin and Leslie (2003, 2009), which study information disclosures such as hygiene grades; and Jacoby, Chestnut, and Silberman (1977), an early examination of consumer use of nutrition information. Together, these works undergird the article's exploration of learning vs. salience mechanisms, the potential for information to influence behavior beyond posting locations, and the broader policy relevance of calorie labeling.

- Quote snippet:

  > The notion that increasing the provision of nutrition information may stimulate people to adopt healthier eating habits is an old idea.

- Citation anchors:
  - page=37, words=0-349, section=References, note=Bassett et al. (2008) NYC calorie information at fast-food chains; central empirical baseline.
  - page=36, words=275-422, section=References, note=Currie et al. (2010) The Effect of Fast Food Restaurants on Obesity and Weight Gain.
  - page=37, words=0-349, section=References, note=Downs, Loewenstein, Wisdom (2009) Strategies for Promoting Healthier Food Choices.
  - page=37, words=0-349, section=References, note=Ippolito & Mathios (1990, 1995) Information and Health Choices.
  - page=37, words=550-649, section=References, note=Variyam & Cawley (2006) Nutrition Labels and Obesity.
  - page=37, words=550-649, section=References, note=Kiesel & Villas-Boas (2008) Another Nutritional Label: Shelf Labels and Consumer Choice.
  - page=37, words=0-349, section=References, note=Jacoby, Chestnut, and Silberman (1977) Consumer Use and Comprehension of Nutrition Information.
  - page=37, words=0-349, section=References, note=Jin & Leslie (2003) The Effect of Information on Product Quality: Restaurant Hygiene Grade Cards.
  - page=37, words=0-349, section=References, note=Elbel et al. (2009) Calorie Labeling and Food Choices: A First Look in NYC.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.2387836351559931)`
- Retrieval method: `local`
- Evidence type: `textual_evidence`
- Answer:

It builds directly on the NYC calorie-posting literature, citing Bassett et al. (2008); Downs, Loewenstein & Wisdom (2009); Elbel et al. (2009); Currie et

al. (2010); and Anderson & Matsa (2011) as key prior work, i.e., the studies examining calorie labeling/posting effects at chain restaurants (with background from earlier nutrition-information research).

- Quote snippet:

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

- Citation anchors:
  - page=8, words=0-349, section=NYC calorie labeling literature, note=Bassett et al. 2008 mentioned in NYC calorie information study background
  - page=8, words=0-349, section=NYC calorie labeling literature, note=Ippolito & Mathios (1990, 1995) cited as nutrition information impact studies
  - page=8, words=0-349, section=NYC calorie labeling literature, note=Kiesel & Villas-Boas (2008) cited regarding nutrition labeling experiments
  - page=38, words=0-349, section=Background, note=Mathios (2000) on the impact of mandatory disclosure laws
  - page=7, words=275-566, section=Earlier nutrition information studies, note=Jacoby, Chestnut, Silberman (1977) early nutrition information study
  - page=8, words=275-554, section=NYC calorie labeling literature, note=Downs, Loewenstein, Wisdom (2009) receipts study in NYC context
  - page=8, words=275-554, section=NYC calorie labeling literature, note=Elbel et al. (2009) calorie labeling study outside NYC restaurants
  - page=8, words=275-554, section=NYC calorie labeling literature, note=Currie et al. (2010) study on fast-food restaurants and obesity
  - page=8, words=275-554, section=NYC calorie labeling literature, note=Anderson & Matsa (2011) related work on restaurant effects
  - page=36, words=275-422, section=References, note=Anderson & Matsa 2011 (reference in article)
  - page=37, words=0-349, section=References, note=Bassett et al. 2008 (NYC calorie information at fast-food chains)

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.2604222121786964)`
- Retrieval method: `local`
- Evidence type: `Literature benchmark / comparison cited in the results section.`

- Answer:

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

- Quote snippet:

  > Two recent papers by Michael L. Anderson and David A. Matsa (2011) and Janet Currie et al. (2010). Neither paper finds that fast food restaurants have a significant effect on obesity in general.

- Citation anchors:
  - page=8, words=0-349, section=Literature review / Results comparison, note=Two recent papers used as benchmarks; none finds a general effect of fast food on obesity.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.28898710712254194)`
- Retrieval method: `local`
- Evidence type: `Data sources and usage within the study`
- Data source: Transaction data; Cardholder data; Milk order data; In-store surveys
- Answer:

The cited datasets are: (1) Transaction data for all Starbucks locations in NYC (222 stores) and Boston/Philadelphia (94 stores) covering 3 months before to 11 months after calorie posting, used to measure calories per transaction and to analyze changes in behavior across cities; (2) Cardholder data consisting of anonymous Starbucks cardholders (about 2.7 million total, with a subsample of 7,520 NYC and 3,772 Boston/Philadelphia individuals) used to isolate effects on intensive vs extensive margins and to control for changes in consumer composition; (3) Milk order data (daily quantities of regular, skim, nonfat milk) used to assess sources of calories and potential substitutions; (4) In-store customer surveys conducted in Seattle and San Francisco (792 completed surveys) used to test knowledge of calories and corroborate behavioral findings; (5) Cross-city comparisons and control cities (Boston/Philadelphia) to validate causal interpretation and time-path analyses of calorie posting effects.

- 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=Definition of the two datasets: transaction data and cardholder data; NYC/Boston/Philadelphia; 3 months before and 11 months after posting.

  - page=5, words=275-591, section=Data Summary, note=Cardholder data used to assess individual-level behavioral changes and compare with transaction data.
  - page=9, words=0-349, section=Context / Data scale, note=Dataset described as large and broad: over 100 million transactions across Boston, NYC, and Philadelphia.
  - page=14, words=0-349, section=Results/Methods, note=Transaction data and cardholder data used to show pre/post trends and to contrast effects controlling for consumer composition.
  - page=25, words=275-599, section=Survey data, note=In-store surveys in Seattle and San Francisco (792 completed) to test consumer knowledge; supports robustness of findings.

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

- Category: `E) Citations and related literature`
- Confidence: `medium (0.38555090495305183)`
- Retrieval method: `local`
- Evidence type: `Econometric/methodology references (RCTs, field experiments, salience).`
- Data source: Starbucks transaction and cardholder data
- Table/Figure: Figure 2
- Answer:

The paper cites methodological references including a randomized natural field experiment (Cai, Chen, Fang 2009) as an example of an RCT, and broader field-experiment/behavioral economics references such as DellaVigna (2009) and Chetty, Looney, and Kroft (2009) on salience; identification relies on within-city variation with fixed effects (a difference-in-differences–type approach).

- Quote snippet:

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

- Citation anchors:
  - page=page 11, words=275-522, section=Methodology (Identification), note=States identification stems from within-city variation over time.
  - page=page 12, words=0-349, section=Methodology, note=Describes store fixed effects and within-city variation for identification.
  - page=page 14, words=0-349, section=Context/Methodology, note=Mentions within-city variation and fixed effects; cites within references to methodological approach.
  - page=page 37, words=0-349, section=References, note=Cai, Hongbin, Yuyu Chen, and Hanming Fang (2009) randomized natural field experiment.

- page=page 14, words=0-349, section=References, note=Chetty, Raj, Adam Looney, and Kory Kroft (2009) Salience and Taxation.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.24683243268777266)`
- Retrieval method: `local`
- Evidence type: `Literature framing / related research`
- Answer:

Yes. The paper references seminal work such as Jacoby, Chestnut, and Silberman (1977) on consumer use of nutrition information, and it positions against the view that fast-food calorie labeling substantially affects obesity, citing Currie et al. (2010) and Anderson & Matsa (2011) as not finding strong obesity effects.

- Quote snippet:

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

- Citation anchors:
  - page=7, words=275-566, section=Related Research, note=Cites Jacoby, Chestnut, and Silberman (1977) as an early study on nutrition information usage.
  - page=8, words=0-349, section=Related Research, note=Mentions Currie et al. (2010) and Anderson & Matsa (2011) arguing no significant obesity effect from fast-food labeling.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.2969280517961589)`
- Retrieval method: `local`
- Evidence type: `Appendix reference (questionnaire)`
- Data source: Appendix (two-page questionnaire)
- Answer:

Yes. The claims reference an Appendix containing the in-store survey questionnaire (The actual two-page questionnaire is shown in the Appendix). There is no citation of code or data repositories in the provided context.

- Quote snippet:

  The questionnaire is shown in the Appendix.

- Citation anchors:
  - page=7, words=275-566, section=Appendix, note=Appendix contains the questionnaire referenced in the survey section
  - page=25, words=275-599, section=Appendix, note=Appendix shows the actual two-page questionnaire and links to survey methods

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.2925432110580015)`
- Retrieval method: `local`
- Evidence type: `literature gaps`
- Data source: Starbucks transaction data; NYC and control cities referenced in related literature.
- Answer:

The cited literature largely focuses on nutrition labeling of packaged foods, with NYC studies often small and inconclusive; gaps remain regarding long-run effects of calorie posting, generalizability beyond a single chain, obesity outcomes, heterogeneity in consumer responses, substitution patterns (e.g., smaller sizes, lower-calorie items, fewer items), and effects on restaurant profits.

- 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=7, words=Two-transaction., section=NYC literature gaps, note=NYC studies show no significant impact; limited sample and scope imply gaps in long-run effects and generalizability.
  - page=9, words=Since-profitability., section=Gaps addressed by this study, note=Authors highlight need to study time-path, longevity, heterogeneity, substitution, and profits; these are gaps in prior work.
  - page=3, words=However,-research., section=Limitations in prior research, note=Two main limitations discussed: obesity impact not measured and data from only one chain; gaps in generalizability.
  - page=8, words=Neither-general., section=Obesity impact gap, note=Prior work finds no significant obesity impact; gap remains on obesity outcomes.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.1675737647641879)`
- Retrieval method: `local`

- Evidence type: `absence_of_information`
- Data source: Starbucks transaction data in NYC, Boston, and Philadelphia; cardholder data
- Answer:

The provided excerpt does not indicate that replication files or code are provided, nor any location for them.

- 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=No replication files or code mentioned in this section.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.25846153994205967)`
- Retrieval method: `local`
- Evidence type: `No preregistration or trial registration is stated; methods describe regression specifications and data sources but do not reference preregistration.`
- Data source: Transaction data for all Starbucks locations in NYC, Boston, and Philadelphia; anonymized cardholder data for a subsample.
- Answer:

There is no mention of a pre-analysis plan or registered trial in the provided text.

- Quote snippet:

  We estimate regressions of the following form: (1) y_sct = x_sct + _POS Tc t + _sct

- Citation anchors:
  - page=9, words=275-587, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Contains regression specification (y_sct = x_sct + _POS Tc t + _sct); no preregistration mention.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2748035588323431)`
- Retrieval method: `local`

- Evidence type: `Confidentiality / data access constraints`
- Data source: Starbucks transaction and cardholder data (calorie posting study)
- Answer:

Yes — the context discloses data-access constraints due to confidentiality requirements and restrictions on sharing details.

- Quote snippet:

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

- Citation anchors:
  - page=5, words=0-349, section=Data access constraints / confidentiality, note=To preserve confidentiality of competitively sensitive information.
  - page=6, words=275-517, section=Confidentiality, note=Due to confidentiality requirements, we are unable to reveal any more details about these differences.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.21844943666575672)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Starbucks transaction data and cardholder data for NYC, Boston, and Philadelphia (balanced panel after exclusions).
- Answer:

Yes, key steps in data cleaning and construction are documented.

- 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=0-349, section=Data cleaning and sample construction, note=Excludes stores not open during the entire data period (balanced panel) and transactions with more than four units; 2.2% excluded.
  - page=6, words=0-349, section=Data processing and confidentiality, note=Normalization of NYC values to one for confidentiality; data sources include transaction and cardholder data.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2842025482871281)`
- Retrieval method: `local`
- Evidence type: `robustness_check_reported`
- Data source: Starbucks transaction and cardholder data
- Answer:

Partially reported.

- Quote snippet:

  Our findings are unchanged if we exclude the weather controls.

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

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.28505401415743975)`
- Retrieval method: `local`
- Evidence type: `empirical results with heterogeneity analyses and generalization caveats`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia; with Seattle controls)
- Table/Figure: Table 9
- Answer:

Populations/settings most likely to generalize from this study are similar urban chain-restaurant customers exposed to calorie labeling, particularly: (1) residents of higher-income and more educated ZIP codes; (2) female cardholders; (3) high-calorie purchasers and high-frequency customers; and (4) commuters who travel between NYC and other cities. Generalization to other restaurant chains remains uncertain, as the study covers only Starbucks.

- Quote snippet:

  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=page 3, words=275-564, section=Limitations and scope, note=Notes two limitations and that data are from only one chain (Starbucks).

  - page=page 5, words=275-591, section=Data summary and limitations, note=States that cardholders may not be representative of Starbucks customers more generally.
  - page=page 20, words=0-349, section=Heterogeneity in the Effect of Calorie Posting, note=Finds larger calorie reductions in higher-income and more-educated ZIPs; female cardholders more responsive.
  - page=page 21, words=0-349, section=Heterogeneity in the Effect of Calorie Posting, note=Details on heterogeneous effects, including high- vs. low-calorie purchasers.
  - page=page 29, words=0-349, section=Commuters (Table 9), note=Examines commuters and non-commuters to assess learning vs salience.
  - page=page 30, words=0-349, section=Is the Effect Big Enough to Matter?, note=Discusses learning vs salience; supports both mechanisms.

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.21431272991336786)`
- Retrieval method: `local`
- Evidence type: `limitation`
- Data source: Starbucks transaction and cardholder data (NYC, Boston, Philadelphia; Seattle data discussed separately)
- Assumption flag: `True`
- Assumption notes: The answer relies on explicit study limitations that cardholders are not representative and the data cover only one chain; assuming generalization would require evidence across other chains or populations.
- Answer:

Cardholder data from a single-chain study (Starbucks) and a sample of loyal cardholders are least likely to generalize to broader populations or other chains.

- 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=3, words=275-564, section=Limitations – chain generalizability, note=The study notes data are from only one chain (Starbucks), questioning generalization to other chains.
  - page=5, words=275-591, section=Limitations – sample representativeness, note=Cardholders may not be representative of Starbucks customers; they are described as above-average loyal.

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.20195364741755326)`
- Retrieval method: `local`
- Evidence type: `limitations and scope boundaries`
- Data source: Starbucks cardholder dataset (limited to Starbucks).
- Assumption flag: `False`
- Answer:

Yes. They explicitly discuss boundary conditions and scope limits, noting that they do not measure obesity directly and that the data cover only one chain (Starbucks).

- Quote snippet:

  > However, 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 key limitations: no obesity measurement; data from a single chain (Starbucks); potential home behavior offset not known.

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

- Category: `G) External validity and generalization`
- Confidence: `medium (0.35769051116993333)`
- Retrieval method: `local`
- Evidence type: `empirical_findings`
- Data source: Starbucks transaction and cardholder data described in the study
- Answer:

Results may differ in other time periods or markets due to timing relative to calorie posting (effects observed immediately after posting on April 1, 2008 and persisting through Feb 2009), seasonal patterns and changing shopper composition, and the underlying mechanism (learning vs. salience). If learning effects dominate, commuters exposed to calorie information in NYC could reduce calories per transaction elsewhere as memory persists; if salience dominates, effects may be localized to NYC and decline outside it. Market characteristics (e.g., presence of nearby competitors like Dunkin Donuts) and the choice of control markets affect estimated effects, and cross-market comparability matters since NYC results rely on Boston/Philadelphia as controls. Therefore, in different time periods (e.g., holidays) or in different markets, the magnitude, persistence, and geographic reach of the calorie-posting effects could vary.

- Quote snippet:

  the drop in calories per transaction occurred right around April 1, 2008, and persisted through February 2009

- Citation anchors:
  - page=11, words=275-522, section=Regression design and pre-trend checks, note=No pre-trend differences between NYC and controls; timing around posting date used to identify effects.
  - page=14, words=0-349, section=Results: timing and persistence of effects, note=Calorie drop aligns with posting date and persists; supports potential cross-market learning vs salience considerations.
  - page=29, words=0-349, section=Discussion on learning vs salience across markets, note=Commuter-learning mechanism implies potential effects beyond NYC.
  - page=21, words=275-421, section=Heterogeneity across consumers and margins, note=Analyzes effects across different categories and proportions; illustrates behavioral heterogeneity.

H01: Are key variables measured directly or via proxies?

- Category: `H) Measurement validity`
- Confidence: `low (0.3077010504969669)`
- Retrieval method: `local`
- Evidence type: `Direct measurement from transaction/cardholder datasets`
- Data source: Transaction data; Cardholder data
- Answer:

Key variables are measured directly from transaction and cardholder data using item-level calorie information (not proxies). For example, calories per transaction and calories per beverage/food item are computed from actual items purchased, based on published calorie counts, and the number of items per transaction is directly observed.

- Quote snippet:

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

- Citation anchors:
  - page=page 6, words=Based-247., section=Calories per transaction computed from transaction data, note=States that calories per transaction are computed from transaction data prior to calorie posting.
  - page=page 12, words=The cardholder data allows-time., section=Data: cardholder vs transaction differences, note=Notes that cardholder data isolates intensive-margin changes by tracking the same individuals over time.

  - page=page 16, words=To examine the relative importance-items)., section=Measurement of items and calories per item, note=Describes using number of items per transaction and calories per item purchased as dependent variables.

H02: What measurement error risks are acknowledged or likely?

- Category: `H) Measurement validity`
- Confidence: `low (0.33249153366803075)`
- Retrieval method: `local`
- Evidence type: `Limitations and measurement-error risks discussed in the study`
- Data source: Starbucks transaction and cardholder data; Seattle/San Francisco surveys
- Table/Figure: Figure 5
- Answer:

The study acknowledges several measurement-error risks: (1) milk additions to beverages generate extra calories that are not observed in the data; (2) data come from a single chain (Starbucks), limiting measurement validity for other chains and obesity-related outcomes; (3) survey-based calorie knowledge data are prone to measurement error and potential sampling biases (e.g., underrepresentation of regular coffee drinkers and possible wave-to-wave sampling differences); (4) self-reported calorie estimates are inherently noisy and may not reflect actual consumption.

- Quote snippet:

  > Consumers frequently add milk to their beverages at the self-service counter, which is a source of additional calories. Neither the transaction data nor cardholder data provide any information about this behavior.

- Citation anchors:
  - page=3, words=275-564, section=Limitations, note=Data for only one chain (Starbucks); cannot infer effects at other chains or obesity outcomes.
  - page=6, words=275-517, section=Measurement details, note=Milk additions are a source of additional calories not captured by the data.
  - page=25, words=275-599, section=Survey limitations, note=Sample under-represents consumers who ordered regular coffee; potential biases in survey sample.
  - page=27, words=275-577, section=Survey limitations, note=Possible inadvertent sampling of a more informed mix across the second survey wave.
  - page=26, words=0-178, section=Survey results, note=Figure- and text describe errors in respondents' calorie estimates, indicating self-reported data are error-prone.

  - page=3, words=275-564, section=Obesity link, note=No direct measurement of obesity effects.

H03: Are there validation checks for key measures?

- Category: `H) Measurement validity`
- Confidence: `low (0.24859403773494546)`
- Retrieval method: `local`
- Evidence type: `Robustness checks and cross-dataset validation`
- Data source: Transaction data; Cardholder data
- Answer:

Yes. The study performs validation checks on key measures via cross-dataset replication and robustness tests: estimates from transaction data and cardholder data are similar; weather controls have insignificant impact and results hold if removed; day-of-week and holiday dummies are included as robustness checks; sample restrictions ensure a balanced panel (excluding stores not open and bulk purchases).

- Quote snippet:

  Note also that we obtain similar estimates of the effect of calorie posting using both the transaction data and cardholder data.

- Citation anchors:
  - page=11, words=275-522, section=Methods across datasets, note=We estimate regressions separately with the transaction data and the cardholder data.
  - page=11, words=275-522, section=Weather robustness, note=The weather controls are included; weather variables have an insignificant impact on sales; results unchanged if weather controls are excluded.
  - page=14, words=275-590, section=Robustness checks, note=Day-of-week and holiday dummies are included; robustness checks show no difference in estimates; Seattle post-law robustness discussed.
  - page=5, words=275-591, section=Data quality and sample restrictions, note=To preserve confidentiality and ensure comparability, transactions are filtered (e.g., exclude stores not open during the entire period; exclude bulk purchases). Excluded transactions represent 2.2% of all transactions.

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

- Category: `H) Measurement validity`
- Confidence: `low (0.20310078750608626)`
- Retrieval method: `local`
- Evidence type: `not_discussed`
- Data source: Starbucks calorie-posting study transaction data (NYC/Seattle and control cities)

- Answer:

No. The passages do not include an explicit discussion of construct validity for core outcomes.

- Quote snippet:

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

- Citation anchors:
  - page=3, words=275-564, section=Limitations / Scope, note=Mentions that they do not directly measure obesity and that data are limited to one chain; discussion of measurement limitations but not construct validity.
  - page=11, words=0-349, section=Methods / Measurement, note=Describes regression specification to measure calories per transaction; demonstrates measurement approach but not validity discussion.

I01: What policy counterfactuals are considered or implied?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3409389970773805)`
- Retrieval method: `local`
- Evidence type: `counterfactual policy analysis (control-city comparison, commuter/within-city variation, and pre/post natural experiments)`
- Data source: Starbucks transaction data; anonymous cardholder data; Seattle store surveys
- Table/Figure: Table 3; Figure 3
- Assumption flag: `True`
- Assumption notes: Assumes Boston/Philadelphia are appropriate counterfactuals for NYC in the absence of posting; Seattle pre/post provides an additional counterfactual for policy effects.
- Answer:

Policy counterfactuals discussed or implied include (1) NYC's outcomes juxtaposed with control cities (Boston and Philadelphia) where calorie posting did not occur, to estimate what would have happened in NYC absent posting; (2) within-cardholder analyses and commuter patterns that serve as counterfactuals for behavior in non-posting locations; and (3) Seattle's pre/post calorie posting law as a separate counterfactual to triangulate the mechanism (learning vs. salience).

- Quote snippet:

  To control for other factors affecting transactions, we also observe every transaction at Starbucks company stores in Boston and Philadel-

phia, where there was no calorie posting.

- Citation anchors:
  - page=2, words=0-40, section=Methods/Data, note=Boston and Philadelphia stores used as controls where calorie posting did not occur, providing a counterfactual for NYC.
  - page=3, words=275-290, section=Methods/Data, note=Seattle calorie-posting law used as a pre/post counterfactual to study policy effects.

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

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3100505768530898)`
- Retrieval method: `local`
- Evidence type: `empirical findings on welfare and distributional effects of mandatory calorie posting.`
- Data source: Calorie posting study data (NYC Starbucks, control cities; transaction and cardholder data).
- Answer:

The policy reduced calories per transaction mainly through the extensive margin (people opting not to buy food items), with little change in beverage calories or visit frequency. Profits were largely unaffected, though average prices per item rose, implying a shift toward higher-margin products. Distributionally, calorie reductions were larger among higher-income and more educated ZIPs, were greater for female cardholders, and were concentrated at the higher end of the calorie distribution (larger absolute reductions at higher quantiles but similar percent reductions). Overall welfare effects thus concentrated health gains among higher-income/educated groups and high-calorie purchasers, with modest or neutral profit implications and some potential equity considerations.

- Quote snippet:

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

- Citation anchors:
  - page=18, words=1-26, section=Results: calories per item, note=Three quarters of calorie reduction attributed to not buying food items (extensive margin).
  - page=20, words=1-20, section=Heterogeneity in effects, note=Decrease in calories per transaction larger in higher-income and more-educated ZIPs; female cardholders more responsive.
  - page=22, words=1-20, section=Quantile results, note=Calorie reductions larger in the top half of the distribution; percent change roughly

5–6% at higher quantiles.
  - page=24, words=260-310, section=Profit implications, note=No significant effect on Starbucks profit; prices per item purchased increased.
  - page=18, words=260-320, section=Beverage vs food calories, note=Beverage calories per transaction largely unaffected.

I03: Are cost-benefit or incidence analyses provided?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3129039136460898)`
- Retrieval method: `local`
- Evidence type: `Qualitative discussion of costs; absence of formal analysis`
- Data source: Starbucks NYC transaction data and control city data; in-store surveys (Seattle and controls)
- Answer:

No formal cost-benefit or incidence analyses are provided. The paper mentions costs of posting (e.g., compliance costs) and potential indirect costs, but it does not present a quantified cost-benefit or incidence analysis.

- Quote snippet:

  One natural argument against calorie posting is that forcing restaurants to put the information on menus is costly.

- Citation anchors:
  - page=4, words=275-576, section=Costs discussion, note=Mentions costs of compliance and indirect costs of calorie posting

I04: What policy recommendations are stated or implied?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3368107264591931)`
- Retrieval method: `local`
- Evidence type: `policy_recommendation`
- Data source: Starbucks NYC transaction data (with calorie posting starting Apr 1, 2008; NYC stores vs comparator locations) and related survey/cardholder data
- Assumption flag: `True`
- Assumption notes: Assumes that the benefits of calorie posting outweigh the costs and that low compliance and public education effects contribute to policy value; relies on the paper's cost-benefit interpretation.
- Answer:

Mandatory calorie posting is potentially beneficial if its benefits outweigh its costs; costs are very low, and it could reduce calories per transaction while encouraging low-calorie menu innovation and offering public education benefits.

- Quote snippet:

  Ultimately, mandatory calorie posting is only a good policy if the benefits outweigh the costs.

- Citation anchors:
  - page=34, words=0-349, section=Conclusion, note=Policy implication: benefits vs costs; low costs; potential for long-run innovation and education.
  - page=33, words=275-553, section=Conclusion, note=Continuation of conclusion with policy implications.

J01: What missingness or attrition patterns are reported?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.3156222240583413)`
- Retrieval method: `local`
- Evidence type: `textual description of missingness/attrition in Data Summary and footnotes`
- Data source: Starbucks transaction data and cardholder data
- Answer:

Missingness/attrition patterns reported include (1) exclusion of transactions from stores not open for the entire data period (balanced panel), (2) a known but small portion of transactions excluded overall (the excluded transactions represent only 2.2 percent of all transactions; the fraction of excluded transactions is unknown but believed to be well under 5 percent), and (3) data gaps due to independent, non–Starbucks-owned locations not covered by the data, implying missing data for those outlets. There is also an acknowledgment that the cardholder sample may not be representative of all Starbucks customers.

- Quote snippet:

  The fraction of excluded transactions is unknown, but we believe it to be well under 5 percent. The excluded transactions represent only 2.2 percent of all transactions.

- Citation anchors:
  - page=5, words=0-349, section=A. Data Summary / Data exclusions, note=Excludes stores not open during the entire data period (balanced panel); fraction of excluded transactions unknown but under 5%; 2.2% of transactions excluded.
  - page=5, words=0-349, section=Data Summary (Footnotes), note=12: data cover all Starbucks company-owned stores; 13: independent locations without data; fraction of excluded transactions unknown but under 5%.
  - page=5, words=275-591, section=A. Data Summary / Data exclusions (continued), note=Cardholders may not be representative of Starbucks customers; potential selection bias.

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

- Category: `J) Data quality and integrity`
- Confidence: `low (0.3111494639257416)`
- Retrieval method: `local`
- Evidence type: `data cleaning / exclusions (outlier handling)`
- Answer:

Outliers are handled by exclusions rather than statistical trimming: they remove transactions at stores not open for the full data period (balanced panel) and any transaction with more than four units of a single item; total exclusions amount to about 2.2% of transactions.

- 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=We-transactions., section=Data cleaning / exclusions, note=Exclude non-open stores and bulk purchases; 2.2% of transactions excluded.

J03: Are there data audits or validation steps described?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.24329738080305616)`
- Retrieval method: `local`
- Evidence type: `Data validation and robustness checks; data cleaning procedures.`
- Data source: Starbucks transaction and cardholder datasets (NYC, Boston, Philadelphia).
- Table/Figure: Figure 2
- Answer:

Yes—the authors describe data validation and data-cleaning steps and perform robustness checks across two datasets to validate their findings. They exclude certain transactions, normalize data for confidentiality, use pre-trend tests, control for holidays and day-of-week, compare NYC to Boston/Philadelphia, and use Seattle as an additional robustness check, including within-individual analysis via the cardholder data.

- Quote snippet:

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

- Citation anchors:
  - page=5, words=0-349, section=Data Summary / Data Cleaning, note=Data describe transaction and cardholder datasets; confidentiality-related exclusions.
  - page=14, words=0-349, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=No evidence of pre-trend differences; cross-dataset robustness checks described.
  - page=14, words=275-590, section=II. Effect of Mandatory Calorie Posting on Calorie Consumption, note=Robustness checks including day-of-week/holiday controls; Seattle robustness; within-individual variation.

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

- Category: `J) Data quality and integrity`
- Confidence: `low (0.3050009509577364)`
- Retrieval method: `local`
- Evidence type: `limitations acknowledgment of sampling bias`
- Data source: Cardholder data (anonymous Starbucks cardholders) and transaction data from Starbucks locations in NYC, Boston, and Philadelphia; Seattle/other controls mentioned.
- Assumption flag: `True`
- Assumption notes: Authors acknowledge that cardholders may not be representative; potential sampling bias affecting generalizability.
- Answer:

Yes. The authors acknowledge potential selective sample inclusion: the cardholder sample may not be representative of all Starbucks customers, being above-average in loyalty, which can bias 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=5, words=combined-Starbucks., section=Data Summary / Cardholder data limitations, note=Notes that cardholders may not be representative of Starbucks customers generally; sample bias.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.24747909226599218)`
- Retrieval method: `local`
- Evidence type: `regression results summary (R-squared values)`

- Data source: Regression results from Bollinger et al. (2011) Table 4 notes (transaction data and cardholder data)
- Answer:

R-squared metrics (R2 and R2') are reported as goodness-of-fit for the regressions. Specifically, R2' ranges from 0.27 to 0.82 for item-per-transaction regressions, and R2 ranges from 0.26 to 0.37 for cardholder-data regressions.

- Quote snippet:

  R2' ranging from 0.27 to 0.82. The regressions using the cardholder data in the top panel are based on 1,511,516 observations, and the R2 vary between 0.26 and 0.37.

- Citation anchors:
  - page=18, words=0-349, section=Table 4 notes, note=R2' ranges 0.27–0.82; R2 ranges 0.26–0.37 for regression models.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.2677824279050766)`
- Retrieval method: `local`
- Evidence type: `empirical regression specifications`
- Data source: cardholder data and transaction data from Starbucks study
- Table/Figure: Table 9; Figure 3
- Answer:

Yes. They test multiple functional forms by using both level and log specifications, including log(calories + 1) and log daily sales, as well as regressions with calories-per-transaction in levels.

- Quote snippet:

  the dependent variable is log(calories + 1)

- Citation anchors:
  - page=page 11 words 0-349, words=0-349, section=Methods/Empirical strategy, note=Regression form y_sct = x_sct + POS_t c t + _sct; discusses the empirical setup
  - page=page 16 words 0-349, words=0-349, section=B. Substitution Effects, note=Describes regressing log daily sales and log calories per item purchased
  - page=page 29 words 0-349, words=0-349, section=Commuters' analysis, note=Dependent variable is log(calories + 1); discusses regression with logs and interactions

K03: Are residual checks or specification tests reported?

- Category: `K) Model fit and diagnostics`

- Confidence: `low (0.2356104236657303)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Transaction data and Cardholder data
- Table/Figure: Table 9
- Answer:

No explicit residual checks or formal specification tests are reported. The authors discuss robustness via alternative specifications (e.g., using separate week dummies, different dependent variables) and report robust standard errors (Table 9).

- Quote snippet:

  "Robust standard errors in parentheses."

- Citation anchors:
  - page=11, words=0-349, section=Model specification, note=Regression specification with fixed effects and controls; baseline model.
  - page=13, words=275-528, section=Robustness/Alternative specifications, note=Describes alternative specification (separate week dummies) and robustness checks.
  - page=16, words=275-558, section=Results/Robustness, note=Discussion of alternative dependent variables and robustness of findings (Table 4).
  - page=29, words=0-349, section=Results/Table notes, note=Table 9 note indicates robust standard errors in parentheses.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `medium (0.376466386118914)`
- Retrieval method: `local`
- Evidence type: `Robustness checks and estimator comparisons`
- Data source: American Economic Journal: Economic Policy, Bollinger et al., calorie posting study (transaction and cardholder data).
- Answer:

Results are robust to alternative specifications and estimators. Robustness checks with date fixed effects show the transaction-based estimates barely change (beverage calories -0.004, food calories -0.152, total calories -0.063); cardholder data confirm the intensive-margin effects but with some differences (beverage calories not significantly different from zero; food calories -11.2% per transaction; total calories -5.0%). Weather controls are insignificant, and identification relies on within-city variation. Quantile analyses indicate the size of effects can vary across the calorie distribution but remain nontrivial, with larger absolute reductions at higher quantiles.

- Quote snippet:

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

- Citation anchors:
  - page=12, words=0-349, section=Robustness checks and estimator comparisons, note=Date fixed effects robustness; transaction results: beverage calories −0.004, food calories −0.152, total calories −0.063.
  - page=11, words=275-522, section=Controls and seasonality, note=Weather variables have an insignificant impact; fixed effects control for seasonality.
  - page=12, words=275-624, section=Cardholder vs transaction data, note=Cardholder data isolate intensive-margin effects; six regressions; within-city variation; extensive margin not captured in these estimates.
  - page=22, words=0-349, section=Quantile analysis, note=Calorie reductions tend to be larger in higher quantiles; effects vary across the distribution.

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

- Category: `P) Previous questions`
- Confidence: `n/a`
- Retrieval method: `n/a`
- Evidence type: `n/a`
- Answer:
- Background and motivation
  - Calorie posting is contested: it can be costly to require menus to post calories, with reported compliance costs around $2,000 per Wendy's store, and concerns about indirect costs from city-specific menus. NYC health dept stresses point-of-purchase availability. (Provenance: page 4, words 275-576)
- Overall impact on calories
  - Calorie posting reduced average calories per store-day by about 4.6 calories, and the policy coincided with an increase in transaction counts in NYC (i.e., more visits). The analysis uses a binary POST indicator (POST = 1 for NYC stores after 1 Apr 2008) and robust controls; alternative specs using separate weekly dummies yield similar timing patterns. (Provenance: page 13, words 275-528)

- Beverage vs. food items per transaction
  - Table 4 (regressions on items and calories per item):
    - Number of beverages per transaction: +0.005 (significant)
    - Number of food items per transaction: −0.029 (significant)
    - Beverages + food items: −0.027 (significant)
    - Calories per beverage (log): −0.008 (significant)
    - Calories per food item (log): −0.039 (significant)
    - Notes: regressions include store fixed effects, week/day dummies, weather controls; separate regressions for transaction vs. cardholder data. (Provenance: page 18, words 0-349)
- Mechanism: extensive vs. intensive margins
  - A large share of the total calorie reduction arises from fewer food purchases (the extensive margin) rather than lower calories per item. This is illustrated in the regression results and summarized as "nearly three quarters of the total calorie reduction can be attributed to people opting not to buy food items." (Provenance: page 18, words 0-349)
- Heterogeneity across customers
  - Table 5 shows how effects vary with demographics and customer types; e.g., interactions with income, education, and high-frequency vs. mid/high-calorie customers help explain variation in the posting effect. (Provenance: page 21, words 0-349)
- Quantile-based effects
  - Table 6 reports effects by calories-per-transaction quantile. Reductions are larger in higher quantiles; the percent reduction is roughly 5–6% from the 75th to the 99th percentile, while the 10th percentile shows minimal change (near zero, as expected). (Provenance: page 22, words 0-349)
- Seattle/control context
  - In Seattle (control context where many items lacked posted calories), drink calories per transaction fell by 4.6 with a standard error of 0.3, while food calories per transaction rose by 0.8 (negligible overall impact on food calories). This aligns with a target effect more pronounced for beverages and supports heterogeneity in response. (Provenance: page 15, words 0-254)
- Robustness and interpretation
  - The study also explores time-varying effects and mealtime variation; results show the beverage-calorie effect is slightly stronger around lunchtime but broadly uniform across the day. (Provenance: page 15, words 0-254)

- Appendix/measurement notes
  - Appendix includes survey instruments for Starbucks customers, illustrating data collection methods (not central to the main results). (Provenance: page 35-36, entire Appendix content)

If you'd like, I can extract a compact mathematical summary (e.g., key regression specification and dependent variables) with inline formulas.