# When Search Stops Sending Traffic: The Impact of Google AI Overviews on E-Commerce Traffic and Customer Acquisition*

Qiwei Han[†] Yufei Shen[‡] Maximilian Kaiser[§]

**Abstract**

Search engines have long created value for merchants by sending shoppers to their websites. Google AI Overviews (AIOV) now answer part of a search before any external visit, and organic traffic falls as a result. Whether merchants also lose customers is a separate question, and the answer determines how merchants should read their traffic data. We study it using Google's staggered international rollout of AIOV together with linked Google Search Console and Google Analytics data for 4,658 e-commerce domains in 131 markets. After rollout, organic impressions fall by 6.8%, clicks by 2.6%, and sessions by 2.8%. Transactions attributed to Google Organic fall by only 0.2% and revenue by 0.5%; neither estimate differs from zero, and the confidence interval excludes transaction losses larger than 0.62%. Because purchases hold up while visits decline, transactions per session rise by 2.6% and revenue per session by 2.1%, with no change in order values. The lost clicks come mainly from results ranked 4–10 and most strongly from informational searches. Five difference-in-differences estimators produce the same ordering. We call this pattern *selective referral compression*: generative search removes

*This work was funded by Fundação para a Ciência e a Tecnologia (UID/00124/2025, UID/PRR/124/2025, Nova School of Business and Economics, DOI: https://doi.org/10.54499/UID/00124/2025) and LISBOA2030 (DataLab2030 – LISBOA2030-FEDER-01314200).

[†]Nova School of Business and Economics, Universidade Nova de Lisboa.

[‡]Nova School of Business and Economics, Universidade Nova de Lisboa.

[§]Vienna University of Economics and Business.

visits faster than it removes purchases, so traffic volume no longer measures the commercial value of organic search.

**Keywords:** generative AI, AI search, Google AI Overviews, search engines, digital customer acquisition, e-commerce

# 1 Introduction

Organic search brings shoppers to merchant websites, and merchants have long measured its value by rankings, clicks, and visits (Berman Katona 2013, Baye et al. 2016). Google AI Overviews (AIOV) change how those visits arise. The feature summarizes information, compares alternatives, and links to supporting sources directly on the results page. The search engine therefore no longer only points users toward external sites; it now completes part of the information search that users previously carried out on those sites. For merchants this raises a practical question with large stakes: when AI search removes website traffic, how much of that loss reaches purchases and revenue?

The answer matters because traffic is a means to sales, not an end in itself. A lost visit can be a lost purchase. It can also be an information search that no longer needs an external page, such as a shopper checking a product specification that the AI summary now displays. Both produce the same drop in clicks, yet they have opposite implications for a merchant's business. Telling them apart is difficult because it requires two kinds of data that rarely meet. Search-platform data record impressions and clicks but stop at the moment the user leaves Google. Merchant analytics record visits and purchases but, on their own, cannot isolate the effect of a change in the search interface. A further difficulty is that AI answers appear on some queries and not others, so comparisons based on where AIOV is displayed mix the feature's effect with Google's choice of which queries to answer.

Recent work establishes the traffic side of this question. A randomized field experiment finds fewer organic clicks and more zero-click searches when AI summaries appear (Agarwal Sen 2026), and Wikipedia receives fewer search referrals after AIOV becomes available in a market (Khosravi Yoganarasimhan 2026). More broadly, large language models reduce activity on online knowledge platforms and in traditional search (Burtch et al. 2024, Shan Qiu 2026, Padilla et al. 2026). These studies show convincingly that generative AI substitutes for some external information access. None of them, however, observes whether the visits that disappear would have ended in a purchase, which is the quantity that determines what generative search costs merchants.

We address this gap by linking Google Search Console (GSC) and Google Analytics (GA)

data for the same e-commerce merchants. GSC records what happens on Google: impressions, outbound clicks, query text, and organic position. GA records what happens after the user arrives: sessions, transactions, and revenue. The resulting panel covers 4,658 domains, 97,236 domain–country units, and 131 markets, and it accounts for 99% of pre-treatment Google Organic sessions among eligible linked merchants. To identify causal effects, we use the fact that Google introduced AIOV country by country between May 2024 and May 2025. We compare markets after the feature arrives with markets that have not yet received it, using the doubly robust estimator of Callaway Sant'Anna (2021). The estimator adjusts for each unit's pre-treatment outcome history, and inference is clustered by country because treatment varies at the country level.

Our central finding is that AI search removes visits much faster than it removes purchases. Over weeks 2–8 after rollout, impressions fall by 6.78%, clicks by 2.63%, and Google Organic sessions by 2.77%. Transactions, by contrast, fall by only 0.19% and revenue by 0.54%, and neither estimate differs from zero. The transaction estimate is also precise enough to be informative: its confidence interval excludes losses larger than 0.62%, about one fifth of the session decline. Because purchases hold up while visits fall, transactions per session rise by 2.57% and revenue per session by 2.05%, while average order value does not change. The same ordering appears under five difference-in-differences estimators and after absorbing merchant-wide weekly shocks.

Query-level data show where the lost clicks come from. Merchants' average organic position does not worsen after rollout, so the traffic loss is not a ranking loss. Holding each query's pre-rollout position fixed, however, clicks from positions 4–10 fall relative to clicks from the top three results, and this gap is largest for informational queries. In other words, a merchant can keep its ranking and still receive fewer clicks from it, particularly when the searcher is looking for information that the AI summary can supply. The effects also differ across merchant categories in a revealing way. Across 12 categories, session losses range from 0.74% to 6.29%, yet transaction effects are statistically indistinguishable across categories. A ranking of categories by traffic loss would therefore misstate which merchants AIOV affects commercially.

We interpret these results through a simple referral-selection framework. AIOV changes what

a user knows before deciding whether to visit a merchant. Some searches that used to end in a visit are now resolved on the results page, while others still lead to the merchant's site. If the visits that disappear carry below-average purchase probability, visits fall faster than purchases and conversion among the remaining visitors rises, even if no individual shopper becomes more likely to buy. We call this pattern *selective referral compression*.

The paper contributes to two literatures. It extends research on search intermediation, which shows that rank shapes attention, clicks, conversion, and profitability (Ghose Yang 2009, Yang Ghose 2010, Animesh et al. 2011, Berman Katona 2013, Agarwal et al. 2015, Shi Trusov 2021). Generative search adds a margin that these studies hold fixed: the intermediary itself supplies information that once required a visit, and our position and intent results show that this margin changes the clicks a given ranking produces, most of all for informational search. The paper also contributes to research on digital customer acquisition. Funnel studies separate browsing from purchasing (Goldstein et al. 2022, Zhuang et al. 2021), and recent work shows that referrals from ChatGPT convert differently from those of established channels (Kaiser Schulze 2026). We study a change inside the incumbent channel, where the search engine changes which searches produce visits at all. Tracing one rollout from impressions to revenue shows that a large traffic response can coexist with a small commercial one, and that organic traffic volume alone misstates the commercial return to search visibility.

The remainder of the paper is organized as follows. Section 2 develops the referral-selection framework. Section 3 describes the rollout and data, and Section 4 the empirical design. Sections 5–8 report the funnel effects, the position and intent results, category heterogeneity, and robustness checks. Section 9 discusses implications for merchants and for research on search intermediation.

# 2 Search Intermediation, Referral Selection, and Customer Acquisition

## 2.1 Search Intermediation and the Referral Decision

Search engines stand between consumers and the firms they might visit, organizing information and directing attention across possible destinations (Weber Zheng 2007). Rank is the main instrument of this allocation. In sponsored search, click-through rates, conversion rates, and profitability all vary with position (Ghose Yang 2009), and the effect of rank depends on product differentiation and competition (Animesh et al. 2011). Organic visibility likewise drives traffic to external sites (Berman Katona 2013, Baye et al. 2016). A search result, however, is worth something to a merchant only if it produces a visit and the visit produces a sale, and the design of the results page can move these two margins separately. Organic and sponsored listings interact (Yang Ghose 2010), and stronger organic competition can reduce sponsored clicks while raising conversion and revenue (Agarwal et al. 2015). Intermediaries also shape which alternatives consumers inspect and how deeply they inspect them (Dukes Liu 2016), and information shown before a click changes whether the click is still worthwhile (Gu Wang 2022).

Generative search enlarges the information shown before the click. An AI-generated response synthesizes content that a user would otherwise collect by opening several pages, and it also links to the sources it draws on. It can therefore either lower or raise the value of continuing to a particular merchant's page. A visit, in turn, is only one stage of customer acquisition. Consumers browse, compare, return, and leave without buying, and their browsing patterns reveal how close they are to purchase (Goldstein et al. 2022). Product-list attributes likewise affect clicks and purchases differently (Zhuang et al. 2021). For both reasons, a change in visits need not pass one-for-one to transactions.

## 2.2 Selective Referral Compression

A simple framework makes this distinction precise. Consider a search task $q$ in which a merchant appears at organic position $r$. Let $V_q$ be the value to the user of opening the merchant's page under conventional search, and let $c(r)$ be the cost of continuing to and inspecting that result, which rises as the result appears lower on the page. The user visits if

$$J_q(0) = \mathbf{1}\{V_q \geq c(r)\}. \tag{1}$$

Generative search shifts the value of the visit by $\Delta_q$, which can take either sign. Synthesized information lowers the value of visiting because the user already has part of what the page offers; supporting links and prominence inside the AI response can raise it. Under generative search the user visits if

$$J_q(1) = \mathbf{1}\{V_q + \Delta_q \geq c(r)\}. \tag{2}$$

Let $L_q = \mathbf{1}\{J_q(0) = 1, J_q(1) = 0\}$ indicate a displaced visit and $G_q = \mathbf{1}\{J_q(0) = 0, J_q(1) = 1\}$ an induced one, so the net loss of visits is $L_q - G_q$. Let $p_q$ be the probability that a visit from task $q$ ends in a purchase, held fixed across the two environments. Aggregate visits and purchases are $\mathcal{R}(a) = \mathbb{E}[J_q(a)]$ and $\mathcal{P}(a) = \mathbb{E}[p_q J_q(a)]$ for $a \in \{0, 1\}$, with proportional declines

$$\delta_P = \frac{\mathcal{P}(0) - \mathcal{P}(1)}{\mathcal{P}(0)} \qquad \text{and} \qquad \delta_R = \frac{\mathcal{R}(0) - \mathcal{R}(1)}{\mathcal{R}(0)}.$$

We define *selective referral compression* as

$$\delta_P < \delta_R. \tag{3}$$

The condition has a direct interpretation. When visits decline on net, purchases fall proportionally less than visits whenever the visits that disappear carry a lower purchase probability than the average visit before rollout. In the special case in which generative search only displaces visits,

the condition reduces to displaced visits having below-average purchase probability. Conversion among the remaining visitors then rises through selection alone, with no change in any individual visitor's propensity to buy. The condition can equally fail: if AIOV removes visits from shoppers who are close to purchase, transactions fall as fast as visits or faster. The framework therefore does not presume the answer, and which case holds is an empirical question. Generative search can also change what retained visitors know when they arrive, which shifts $p_q$ itself. Funnel data identify the combined effect of selection and information on the relative movement of visits and purchases.

The empirical counterpart of (3) is the transaction-minus-session outcome

$$\log(1 + \text{Transactions}) - \log(1 + \text{Sessions}).$$

For counts well above one, a positive treatment effect on this outcome means that transactions fall proportionally less than sessions, which is exactly the comparison in (3).[1]

## 2.3 Organic Position and Search Intent

The framework points to two sources of heterogeneity that the data allow us to examine. The first is rank, which sets the cost of continuing to a result, $c(r)$. Results whose pre-rollout value sits just above that cost are the ones most likely to change status when AIOV changes what users see before clicking. The clicks a fixed position produces can therefore change without any change in the position itself. Whether this sensitivity is greater for lower-ranked results is an empirical question, which we answer by fixing position groups before rollout. The second is the search task, which determines the size of $\Delta_q$. Synthesized content substitutes most directly for browsing when the user wants information. Navigational and transactional searches often still require reaching a specific site, although AI answers can also settle product facts and comparisons along the way. We therefore estimate the position gradient separately for informational, navigational, commercial-investigation, and transactional queries.

[1] With zero counts, log-plus-one differences approximate proportional changes less closely; we therefore treat the outcome as a transformed funnel margin, not an exact conversion-rate ratio.

The framework connects to a broader body of evidence that technology changes where and at what cost consumers acquire information. When a platform lowers the cost of acquiring part of the relevant information, consumers rely more on that part and change their subsequent decisions (Jiang et al. 2025). Generative AI extends this logic, because it produces answers that substitute for information previously obtained from external sources. ChatGPT has reduced activity and contribution on online knowledge platforms (Burtch et al. 2024, Shan Qiu 2026), and LLM adoption has reduced use of traditional search engines (Padilla et al. 2026). For AIOV specifically, a randomized field experiment finds fewer outbound organic clicks when AI summaries appear (Agarwal Sen 2026), and the geographic rollout reduced search referrals to Wikipedia (Khosravi Yoganarasimhan 2026).

These studies measure what generative search does to traffic, but none observes what the lost traffic was worth. Channel research comes closest: Kaiser Schulze (2026) show that ChatGPT referrals differ from established e-commerce channels in conversion, order value, and revenue per session. Their comparison is across channels and across the users who arrive through them. Our setting holds the channel fixed and changes the intermediary, because the incumbent search engine changes which searches produce visits at all. Linking search exposure to sessions, transactions, and revenue for the same merchants lets us measure how much of the lost traffic carried purchases with it.

# 3 Institutional Setting and Data

## 3.1 AI Overviews and International Rollout

AI Overviews place an AI-generated response above or among conventional search results, together with links to supporting sources. Because the response combines information from several pages before the user opens any of them, it changes both whether a visit occurs and what the user already knows when arriving at a merchant's site. Google introduced the feature in stages. The United States received broad availability in May 2024; the United Kingdom, India, Japan, Indonesia,

Mexico, and Brazil followed in August 2024; more than 100 countries and territories followed later in 2024; additional European markets followed in March 2025; and availability reached more than 200 countries and territories and more than 40 languages in May 2025 (Google 2024b,c,a, 2025b,a). We assign each market the first calendar week in which Google documents general AIOV availability, denoted $G_c$ for market $c$. Because treatment is defined by market-level availability, the estimates measure the effect of entering an AIOV-enabled search environment, an intention-to-treat effect that includes queries on which no AI Overview appears.

## 3.2 Linking Search Exposure to Merchant Outcomes

We build a domain–country–week panel that links each merchant's GSC and GA records. GSC covers the search side of the journey: impressions, outbound clicks, query text, and the position at which the domain appears for each query. Clicks on links inside an AI Overview also count as GSC clicks, so the click effects we estimate measure net changes in outbound referrals. GA covers the merchant side: Google Organic sessions, completed transactions, and attributed revenue. We aggregate GSC to the domain–country–week level to match GA and keep the query-level records for the position and intent analyses in Section 6.

This link lets us follow one rollout through every stage of the acquisition funnel, from impressions and clicks on Google to sessions on the merchant site and finally to transactions and revenue. It also provides a check on the link itself. Because GSC and GA are separate measurement systems, equal click and session effects show that the difference between traffic and purchase effects arises after the user arrives on the site, not at the boundary between the two systems.

## 3.3 Sample Construction

The eligible universe contains 7,083 e-commerce domains with linked GSC and GA data. We rank domain–country units by pre-rollout Google Organic sessions and keep the units that together account for 99% of pre-treatment activity and have enough pre-treatment history to construct outcome profiles. The resulting sample contains 4,658 domains, 97,236 domain–country units, and

131 markets, and its construction uses only pre-treatment information. A domain that operates in several countries appears once per country, and each of these units has its own rollout date. Zeros are recorded only in weeks in which a unit has valid measurement, and missing weeks outside that window remain missing, so that gaps in data collection are not mistaken for weeks without activity. The panel is therefore unbalanced, and the estimator uses the observations each group-time comparison requires; Section 8 shows that balanced and common-support samples give the same results. Appendix A describes the sample and the five rollout cohorts.

## 3.4 Outcome Construction

For domain $d$, country $c$, and week $t$, let

$$S_{dct} = \log(1 + \text{Sessions}_{dct}) \qquad \text{and} \qquad T_{dct} = \log(1 + \text{Transactions}_{dct}),$$

and apply the same transformation to impressions and clicks. Because refunds can make weekly revenue negative, revenue enters as

$$R_{dct} = \text{sign}(\text{Revenue}_{dct}) \log(1 + |\text{Revenue}_{dct}|) .$$

To compare adjacent stages of the funnel, we construct three margins:

$$CVR_{dct} = T_{dct} - S_{dct}, \tag{4}$$

$$AOV_{dct} = R_{dct} - T_{dct}, \tag{5}$$

$$RPS_{dct} = R_{dct} - S_{dct}. \tag{6}$$

$CVR$ measures transactions relative to sessions, $AOV$ revenue relative to transactions, and $RPS$ revenue relative to sessions. Estimating each margin directly as an outcome accounts for the covariance between adjacent stages and keeps weeks with zero transactions in the sample. We report effects as $100[\exp(\widehat{ATT}) - 1]$, a percentage mapping of the effect on the transformed scale;

for revenue this mapping is a display convention, since the underlying outcome is signed log.

# 4 Empirical Design

## 4.1 Group-Time Treatment Effects

Because AIOV reached markets at different dates, conventional two-way fixed-effects estimators would compare newly treated markets with markets that were treated earlier. Such comparisons bias the estimates when effects differ across cohorts or change over time (de Chaisemartin D'Haultfoeuille 2020, Goodman-Bacon 2021, Sun Abraham 2021). We therefore use the group-time estimator of Callaway Sant'Anna (2021), which compares each rollout cohort only with markets that have not yet been treated. Let $Y_{it}(0)$ be the outcome of domain–country unit $i$ in week $t$ if its market has not yet received AIOV, and $Y_{it}(g)$ the outcome if the market first receives it in week $g$. For $t \geq g$,

$$ATT(g,t) = \mathbb{E}[Y_{it}(g) - Y_{it}(0) \mid G_i = g]\,. \tag{7}$$

Identification rests on conditional parallel trends. Let $X_i(Y)$ be unit $i$'s pre-treatment profile for outcome $Y$ and $\Delta_g Y_{it}(0) = Y_{it}(0) - Y_{i,g-1}(0)$. With $g-1$ as the base period, we assume

$$\mathbb{E}\big[\Delta_g Y_{it}(0) \mid G_i = g, X_i(Y)\big] = \mathbb{E}\big[\Delta_g Y_{it}(0) \mid G_i > t, X_i(Y)\big]\,. \tag{8}$$

In words, absent AIOV, units in a newly treated market would have followed the same path as not-yet-treated units with similar pre-treatment histories, so the latter identify the counterfactual change for cohort $g$.

## 4.2 Pre-Treatment Adjustment

We estimate $ATT(g,t)$ with the doubly robust procedure, which combines outcome regression and propensity-score weighting and remains consistent if either component is correctly specified. For

each outcome, $X_i(Y)$ contains four summaries of that outcome in the pre-treatment period: its mean, linear trend, standard deviation, and share of zero weeks. The adjustment aligns treated and comparison units on their pre-treatment behavior, while identification still comes from differences in rollout timing. The position and intent analyses apply the same adjustment over their own pre-treatment windows.

### 4.3 Aggregation and Inference

Event-time effects aggregate the cohort-time effects,

$$ATT_e = \sum_{g \in \mathcal{G}_e} \omega_{g,e} \, ATT(g, g+e),$$

where $\omega_{g,e}$ is cohort $g$'s weight at event time $e = t - g$. The headline estimand averages weeks 2 through 8 after rollout,

$$\overline{ATT}_{2:8} = \frac{1}{7} \sum_{e=2}^{8} ATT_e. \tag{9}$$

We exclude weeks 0 and 1 because they contain partial exposure, and we choose the window so that all five rollout cohorts are observed over its full length. Since rollout timing is shared by all domains in a market, standard errors, confidence intervals, and joint tests are clustered by country.

## 5 Main Results

### 5.1 Effects Along the Acquisition Funnel

AIOV reduces every upstream measure of organic reach (Table 1, Panel A). Impressions fall by 6.78% ($\beta = -0.0702$, $p < .01$), which indicates that merchants' links are seen less often once AI summaries occupy part of the results page. Clicks fall by 2.63% ($\beta = -0.0267$, $p < .001$) and Google Organic sessions by 2.77% ($\beta = -0.0281$, $p < .001$). These declines are in line with field evidence that AI summaries reduce outbound clicks (Agarwal  Sen 2026, Khosravi

**Table 1:** Effects of AI Overviews on Organic Search and Customer Acquisition

| | Impressions | Clicks | Sessions | Transactions | Revenue |
|---|---|---|---|---|---|
| *Panel A. Search and commercial outcomes* | | | | | |
| Average ATT, $e = 2, \ldots, 8$ | $-0.0702^{**}$ | $-0.0267^{***}$ | $-0.0281^{***}$ | $-0.0019$ | $-0.0054$ |
| | (0.0227) | (0.0054) | (0.0081) | (0.0022) | (0.0080) |
| Mapped effect (%) | −6.78 | −2.63 | −2.77 | −0.19 | −0.54 |

| | Session–Click | $CVR$ | $AOV$ | $RPS$ |
|---|---|---|---|---|
| *Panel B. Direct contrasts across funnel stages* | | | | |
| Average ATT, $e = 2, \ldots, 8$ | 0.0004 | $0.0254^{***}$ | $-0.0041$ | $0.0203^{***}$ |
| | (0.0037) | (0.0070) | (0.0061) | (0.0050) |
| Mapped effect (%) | +0.04 | +2.57 | −0.41 | +2.05 |

*Notes.* Callaway–Sant'Anna doubly robust group-time average treatment effects averaged over event weeks $e = 2, \ldots, 8$, with not-yet-treated markets as comparisons and adjustment for the pre-treatment outcome mean, slope, standard deviation, and zero share. Panel B outcomes are estimated directly: Session–Click = $\log(1+S) - \log(1+C)$, $CVR = \log(1+T) - \log(1+S)$, $AOV = R - \log(1+T)$, and $RPS = R - \log(1+S)$, where $R$ is signed-log revenue. Country-clustered multiplier-bootstrap standard errors in parentheses. Mapped effects equal $100[\exp(\widehat{ATT}) - 1]$. The sample contains 5,180,452 observations, 97,236 domain–country units, 4,658 domains, and 131 countries. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Yoganarasimhan 2026). Clicks recorded by Google and sessions recorded by merchants also fall by the same amount: the directly estimated session-minus-click contrast is +0.04% ($p = .912$). The traffic loss therefore passes intact from Google to the merchant site, and any divergence between traffic and purchases must arise after the visit begins.

Purchases, by contrast, barely move. Transactions fall by 0.19% and revenue by 0.54%, and neither estimate differs from zero. This is not a case of an imprecise estimate that leaves large effects unresolved: the 95% confidence interval for transactions excludes losses larger than 0.62%, about one fifth of the session decline. The directly estimated transaction-minus-session margin ($CVR$) rises by 2.57% ($p < .001$), confirming that transactions fall significantly less than sessions. Revenue follows the same pattern. Revenue per session ($RPS$) rises by 2.05% ($p < .001$), while average order value ($AOV$) changes by −0.41% and does not differ from zero. The remaining visits thus produce more orders, not larger ones, and revenue holds up because conversion among remaining visitors rises.

The magnitudes can be expressed as a pass-through rate. The point estimates imply that a 1%

loss of organic sessions is accompanied by roughly a 0.07% loss of transactions, and even the upper end of the transaction confidence interval implies a pass-through of at most about 0.2%. Most of the visits AIOV removes were therefore not going to end in a purchase during the period we study. This separates two quantities that organic search measurement usually treats as one: website traffic and realized commerce. A merchant that tracked only organic sessions would read the rollout as a loss of customers, whereas a merchant that tracked organic transactions would see little change. Section 6 examines which visits disappear.

### 5.2 Dynamic Effects

The weekly estimates in Figure 1 reproduce the ordering of Table 1 over time. Session effects are negative throughout weeks 2–8, transaction and revenue effects stay close to zero, and $CVR$ rises, so the gap between traffic and purchases is present in every week of the window and not driven by a single period. Coefficients in the weeks immediately before rollout are jointly indistinguishable from zero for sessions ($p = .391$), transactions ($p = .511$), revenue ($p = .425$), and $CVR$ ($p = .468$), which supports the comparability of treated and not-yet-treated markets just before the feature arrives. Appendix C reports tests over a longer pre-treatment window and placebo rollout dates.

## 6 Where Referral Losses Are Concentrated

The main results show that AIOV removes visits that carry few purchases. Query-level GSC data allow us to examine which visits these are. For each domain, query, and week, GSC reports the query text and the position at which the domain appeared, so we can ask whether the decline in clicks reflects worse rankings or a change in which positions and search tasks still generate clicks.

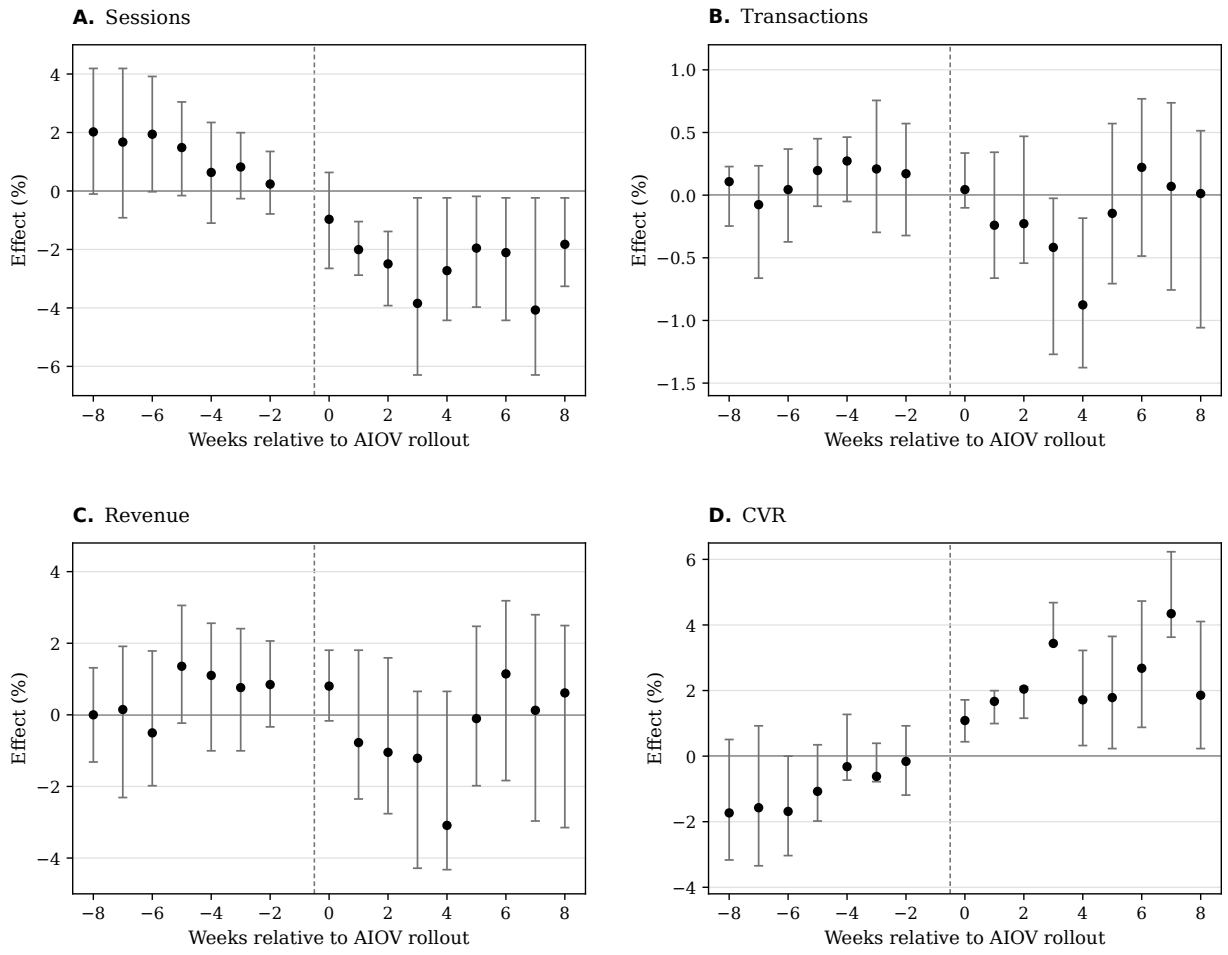



**Figure 1:** Dynamic Effects of Google AI Overviews. Event-time effects under the preferred Callaway–Sant'Anna specification. The headline estimate averages event weeks $e = 2, \ldots, 8$. Whiskers show 95% confidence intervals.

## 6.1 Predetermined Organic Position

The decline in clicks is not a decline in rankings. GSC reports the position of a domain's topmost link on the results page, averaged across impressions, and paid ads do not occupy a position in this measure. Average position changes by $-0.283$ ($p = .226$) after rollout; since lower values indicate higher placement, merchants' rankings do not deteriorate. Clicks from a given position nonetheless fall. To measure this, we assign each query to a position group using pre-rollout data and hold the assignment fixed, so that the comparison does not depend on how Google records positions once AIOV arrives. Top-3 clicks change by 0.57% and do not differ from zero (Figure 2, Panel A), while clicks from positions 4–10 fall by 1.99% and clicks from positions 11–20 by 0.93%. The directly

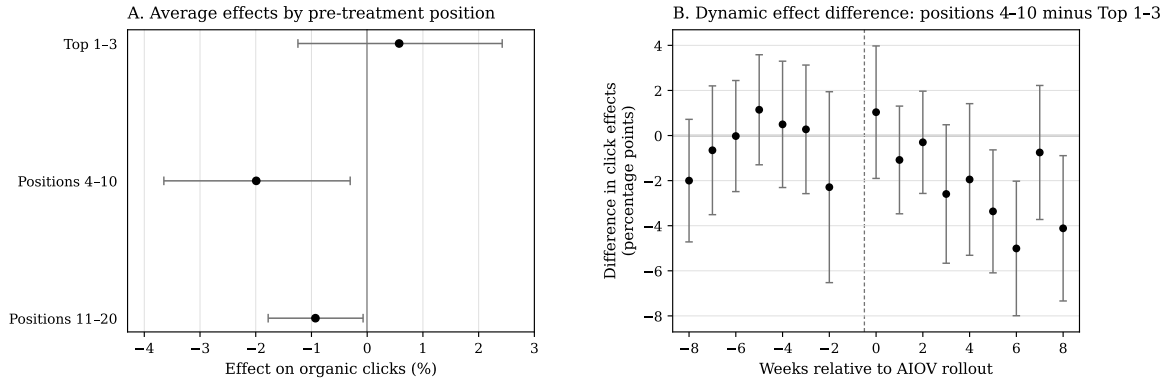



**Figure 2:** AIOV Effects on Organic Clicks by Predetermined Search Position. Panel A reports average effects by pre-rollout position group. Panel B reports the event-time difference between positions 4–10 and the top three. Position groups are fixed using pre-rollout GSC data. Whiskers show 95% confidence intervals. The average positions-4–10-minus-Top-3 effect over $e = 2, \ldots, 8$ is $-0.0258$ ($p = .020$); pre-rollout coefficients at $e = -4, -3, -2$ are jointly indistinguishable from zero ($p = .393$).

estimated positions-4–10-minus-Top-3 contrast is $-0.0258$ ($p = .020$). Panel B shows that this gap opens after rollout and widens in later weeks, while coefficients in the three weeks before rollout are jointly indistinguishable from zero ($p = .393$).

AIOV therefore lowers the clicks that a fixed ranking produces, and it does so mainly below the top three results. This is the pattern the framework predicts when the AI summary gives users enough information to stop before reaching lower-ranked results, whose inspection cost was already higher before rollout.

## 6.2 Search Intent

Position alone does not reveal what the user was trying to do. We classify each query's intent with a multilingual XLM-RoBERTa model (Conneau et al. 2020) into four nonexclusive labels: commercial investigation, informational, navigational, and transactional.[2] For query $q$, intent $k$, position group $b$, unit $i$, and week $t$, let $p_{qk}$ be the predicted probability of intent $k$ and $C_{qit}$ the

[2]The model is trained on an external labeled query corpus after removing queries that overlap with the GSC data. On held-out validation it reaches a macro ROC-AUC of 0.825 and a macro PR-AUC of 0.640. Label-level performance and calibration diagnostics are available from the authors; Appendix G describes the position and intent measures.

observed clicks. We allocate clicks to intents in proportion to these probabilities,

$$C_{it}^{k,b} = \sum_{q \in b} p_{qk} C_{qit}, \tag{10}$$

and estimate effects on $Y_{it}^{k,b} = \log(1 + C_{it}^{k,b})$. Because a query can carry several intents, these outcomes overlap and do not add up to total clicks.

Informational search loses the most clicks below the top three (Table 2). The positions-4–10-minus-Top-3 contrast is −0.0285 ($p = .0077$) for informational queries and −0.0209 ($p = .0114$) for transactional queries, while it is close to zero for navigational and commercial-investigation queries. None of the Top-3 effects differs from zero, so the gradients reflect losses in positions 4–10: clicks there fall by 2.15% for informational queries, 1.94% for transactional queries, and 0.94% for navigational queries. Panel B tests whether these gradients differ. The informational gradient is steeper than every other gradient: it is more negative than the navigational gradient by 0.0263 ($p = .0027$), the commercial-investigation gradient by 0.0275 ($p = .0014$), and the transactional gradient by 0.0071 ($p = .0408$).

These results match the second source of heterogeneity in the framework. When users seek information, the AI summary substitutes most directly for visiting a page, and the loss of clicks below the top three is largest. Lower-ranked transactional results also lose clicks, which indicates that AI summaries settle part of the comparison and product research that precedes a purchase. Navigational searches, where the user wants a specific site, are least affected.

# 7 Heterogeneity Across Merchant Categories

We assign domains to categories using a predetermined Semrush top-level taxonomy and keep the 12 categories with at least 100 domains in the sample; every category is present in every rollout and comparison cohort. Figure 3 shows that categories differ sharply in how much traffic they lose. Session effects range from −0.74% in Health to −6.29% in Vehicles, and equality

**Table 2:** AIOV Effects by Query Intent and Predetermined Organic Position

| Query intent | Top 1–3 | Positions 4–10 | 4–10 minus Top 1–3 |
|---|---|---|---|
| *Panel A. Position-specific treatment effects* | | | |
| Commercial investigation | −0.0047 | −0.0048 | 0.0004 |
| | (0.0044) | (0.0048) | (0.0042) |
| Informational | 0.0055 | −0.0217* | −0.0285** |
| | (0.0080) | (0.0095) | (0.0107) |
| Navigational | −0.0083 | −0.0094* | −0.0020 |
| | (0.0059) | (0.0042) | (0.0046) |
| Transactional | 0.0013 | −0.0196* | −0.0209* |
| | (0.0068) | (0.0089) | (0.0083) |
| *Panel B. Cross-intent differences in the position gradient* | | | |
| Informational minus Navigational | | | −0.0263** |
| | | | (0.0088) |
| Informational minus Commercial investigation | | | −0.0275** |
| | | | (0.0086) |
| Informational minus Transactional | | | −0.0071* |
| | | | (0.0035) |

*Notes.* Position groups are fixed using pre-rollout desktop GSC data. Intent probabilities come from the multilingual XLM-RoBERTa multi-label classifier, and clicks are allocated fractionally by Equation (10) without hard classification or renormalization. Each cell reports the preferred Callaway–Sant'Anna effect averaged over $e = 2, \ldots, 8$. "4–10 minus Top 1–3" is estimated directly at the unit-week level. Country-clustered multiplier-bootstrap standard errors in parentheses. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

across categories is rejected ($p = .0004$). The transaction-minus-session margin also differs across categories ($p = .0002$), ranging from +0.85 to +6.35 on the mapped scale. Vehicles has both the largest session decline and the largest margin, and its deviation from the cross-category mean survives multiplicity adjustment.

Purchase effects, in contrast, do not differ across categories. Transaction effects range only from −0.92% to +0.39%, all 12 simultaneous confidence intervals include zero, and equality across categories is not rejected ($p = .3865$). Categories that lose more traffic therefore do not lose more purchases. The variation across categories lies in how strongly each category's referral traffic is compressed before purchase, not in how many purchases it loses. This matters for judging which merchants AIOV affects. A ranking of categories by session loss would place Vehicles far ahead of Health and suggest large differences in commercial exposure, whereas the transaction estimates are tightly bunched around zero. Selective referral compression varies across categories because

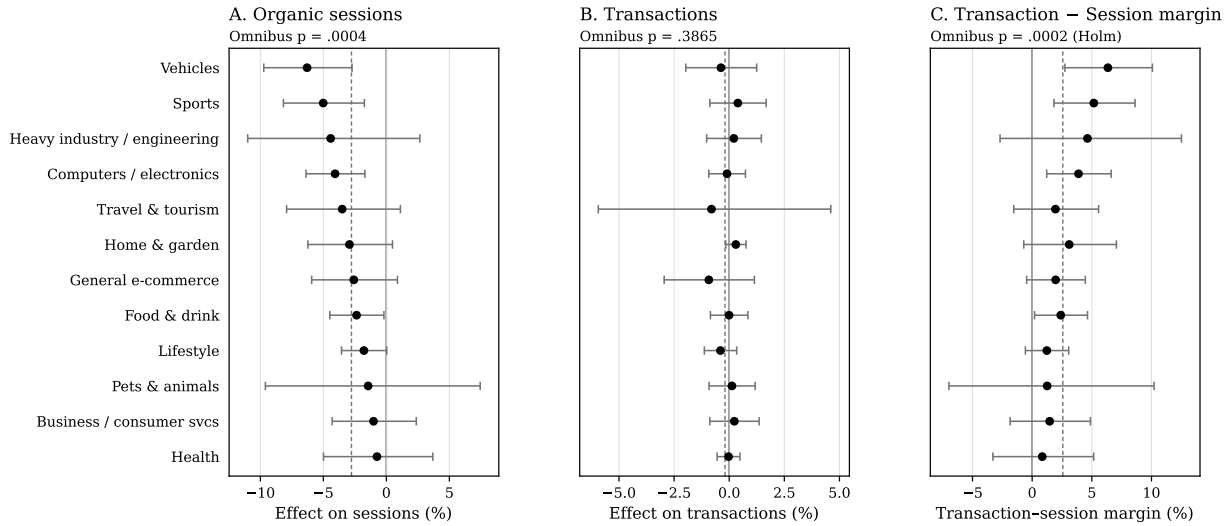



**Figure 3:** Heterogeneity in AIOV Effects Across Merchant Categories. Category-specific effects on organic sessions, transactions, and the transaction-minus-session margin over $e = 2, \ldots, 8$, ordered by the session effect. Whiskers show simultaneous 95% confidence intervals from common country-level multiplier-bootstrap draws; dashed lines mark pooled estimates. Equality across categories is rejected for sessions ($p = .0004$) and the transaction-minus-session margin ($p = .0002$), not for transactions ($p = .3865$). Effects on the mapped percentage scale.

some categories lose more traffic, not because they lose detectably more sales.

# 8 Robustness and Identification Checks

## 8.1 Alternative Estimators

The funnel ordering does not depend on the choice of estimator. Figure 4 compares the preferred specification with four alternatives: Two-Stage DiD, Stacked DiD, the same group-time estimator without pre-treatment adjustment, and two-way fixed effects. Session losses range from 2.47% to 2.91% across the five estimators, and $CVR$ rises by 1.99% to 2.62%; every confidence interval for both outcomes excludes zero. Downstream estimates vary more in precision, as expected for outcomes with many zero weeks, but transaction losses stay well below session losses in every specification. Transaction point estimates range from −0.19% to −0.53% and revenue estimates from −0.54% to −1.21%. The largest come from Two-Stage DiD, at −0.53% for transactions (95% CI −1.16% to 0.10%) and −1.21% for revenue (95% CI −2.63% to 0.23%), both with $p = .10$. Even

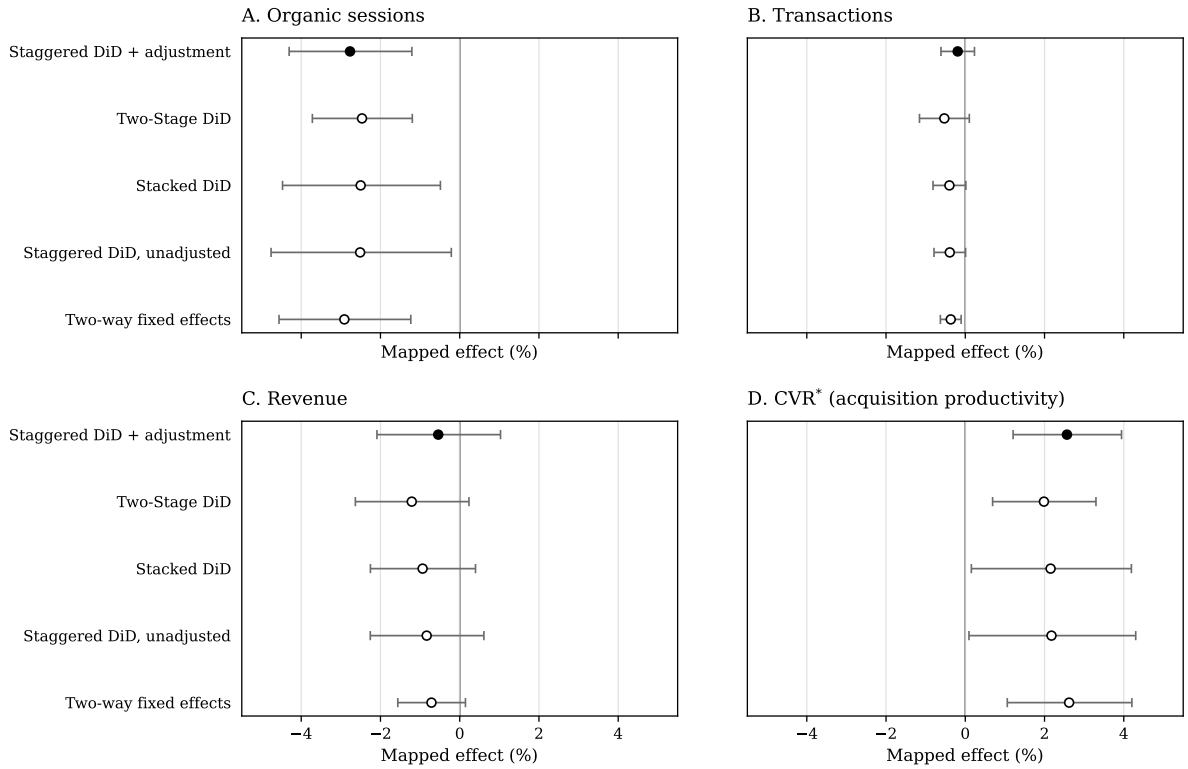



**Figure 4:** Estimates Across Alternative Difference-in-Differences Specifications. Points report estimated effects and whiskers 95% confidence intervals. The staggered-adoption estimators average event weeks $e = 2, \ldots, 8$; two-way fixed effects averages the full post-rollout period. Effects on the mapped percentage scale; revenue uses a signed-log outcome.

the largest downstream estimate is about one fifth of the corresponding session loss. Appendix D reports estimates by rollout cohort, and Appendix E reports count-scale estimates.[3]

## 8.2 Merchant-Wide Concurrent Shocks

Merchants face shocks that hit all their countries at once, such as product launches, inventory changes, site redesigns, brand campaigns, and changes in measurement. If such shocks coincided with rollout dates, they could be mistaken for AIOV effects. Domain-by-week fixed effects, together with domain–country fixed effects, absorb them, so that the remaining variation comes only from differences in AIOV availability across one merchant's countries in the same week. Under this

[3] In the balanced two-way fixed-effects benchmark, a Goodman–Bacon decomposition reconstructs the coefficient exactly. All comparison weights are nonnegative, and 31.2% of the weight comes from comparisons that use already-treated units as controls.

specification sessions fall by 2.32% ($p = .0048$), transactions by 0.42% ($p = .0014$), and the transaction-minus-session margin rises by 1.94% ($p = .0095$). The funnel ordering therefore holds within merchants, comparing the same firm's markets in the same week.

## 8.3 Cross-Market Exposure Within Merchants

AIOV in one country could change a merchant's behavior in its other countries before they receive the feature, for example if the merchant revises its content strategy globally. Such spillovers would contaminate the not-yet-treated comparisons. We test for them among observations whose own market is not yet treated, about 3.25 million observations from 4,603 multi-market domains in 131 countries, by relating outcomes to the share of the merchant's other markets that have already received AIOV. Raising that share from zero to one third changes sessions by +0.36% (95% CI −2.7% to +3.5%) and leaves transactions essentially unchanged. A negative control based on other markets scheduled for treatment within the next four weeks also yields small estimates. We find no systematic evidence of pre-rollout spillovers across a merchant's markets.

## 8.4 Support Restrictions and Placebo Timing

The estimates are similar under stricter sample definitions. A balanced sample covering 99% of pre-treatment activity yields a 2.82% session decline and a +2.64% transaction-minus-session margin, and a common-support sample yields −2.99% and +2.69%. We also assign placebo rollout dates four and eight weeks before the documented dates and re-estimate the three funnel margins. None of the six placebo estimates for $CVR$, $AOV$, and $RPS$ is individually significant, and a joint maximum-$t$ test gives $p = .403$. The rise in conversion among remaining visitors therefore does not appear before AIOV arrives.

# 9 Discussion

## 9.1 Generative Search and Organic Rank

AIOV changes the traffic a ranking produces without changing the ranking itself. Average position does not worsen after rollout, yet positions 4–10 lose clicks relative to the top three, so the traffic a result receives now depends on what the search engine shows above it and not only on where the result appears. This extends the conventional relationship between rank and attention. In conventional search, lower-ranked results are less likely to be inspected and selected because reaching them costs time and effort. A generative interface adds information before that inspection decision. The stronger decline at positions 4–10 is consistent with users stopping once the summary answers their query, and the effect is strongest for informational queries, where synthesized content most directly replaces browsing. Lower-ranked transactional results also lose clicks, so the change reaches beyond purely informational tasks.

Generative search therefore changes the mapping from organic visibility to external visits. Rank still governs visibility, but a fixed rank no longer produces the same traffic once the search engine itself supplies part of the information that used to require a visit. For merchants, holding a ranking no longer guarantees the traffic that the ranking used to deliver, and search engine optimization aimed only at position captures a shrinking share of the value of organic search.

## 9.2 Traffic Loss and Customer Acquisition

The central distinction in our results is between lost visits and lost customers. Sessions fall by 2.77% and transactions by 0.19%, so transactions per session rise by 2.57%, and every estimator preserves this ordering. We call this pattern selective referral compression. AIOV removes visits faster than it removes purchases, and the visits it removes come disproportionately from results below the top three and from informational queries, the searches furthest from a transaction. Revenue reinforces the distinction: revenue per session rises by 2.05% while average order value does not change, so revenue holds up relative to traffic because more of the remaining visits end in an order, not because

orders become larger. Category heterogeneity points the same way. Session effects vary widely across merchant categories while transaction effects do not, so a ranking of AIOV exposure based on traffic differs from one based on realized customer acquisition.

These findings change how search performance should be measured. Organic traffic remains valuable, because a visit brings the shopper onto the merchant's own site, where the merchant controls the presentation and the relationship. But once the search engine answers part of the query before the click, visits no longer measure the commercial return to search visibility. Merchants should track organic reach and organic sales as separate metrics, and studies of platform changes should follow the funnel past clicks to transactions. More broadly, generative search moves part of the information search onto the results page, and the commercial cost to merchants depends on which stages of the journey move. In our data, the visits that disappear carry few purchases.

## 9.3 Scope and Limitations

The estimates measure the early effect of market-level AIOV availability. Treatment is documented availability in a market, not the display of an AI Overview on a particular query, so the estimand is the intention-to-treat effect of entering an AIOV-enabled search environment. Because the estimates average weeks 2–8 after rollout, they describe the short-run response, not longer-run adaptation by merchants or users. The design is observational and rests on staggered rollout timing and conditional parallel trends. Coefficients more than four weeks before rollout are less uniform for transactions and revenue than the immediate pre-rollout coefficients, and effects differ across rollout cohorts, with the most recent cohort contributing substantially to the pooled estimate (Appendices C and D). The alternative estimators, merchant-by-week specification, sample restrictions, placebo dates, and spillover checks support the design without converting the rollout into randomized exposure.

Measurement sets further boundaries. Transactions and revenue are those GA attributes to Google Organic, so purchases completed after a user returns through another channel fall outside this measure. The outcomes are log-plus-one transforms whose mapped effects approximate percentage changes when counts are large. The sample consists of e-commerce merchants whose

commercial value is observable through sessions, transactions, and revenue; publishers that earn from page views, advertising, or subscriptions may respond differently, since a lost visit costs them revenue even when no purchase was at stake. Finally, the position and intent results show where referral losses occur, not why an individual user stops. Predetermined position is not randomly assigned, intent is inferred from query text, and aggregate data do not separate selection of who visits from changes in what visitors know when they arrive. Query-level AIOV display data and linked search-and-purchase paths would separate the two, and longer post-rollout panels would show whether merchants adapt their content, search strategy, or conversion design.

# 10 Conclusion

Google AI Overviews reduce organic traffic to e-commerce merchants much more than they reduce purchases. Across 4,658 domains and 131 markets, sessions fall by about 2.8% in the weeks after rollout, while transactions and revenue barely move and transactions per session rise. The lost clicks come from results below the top three and most strongly from informational searches, and the same ordering holds across estimators and robustness checks.

The broader implication concerns measurement. Organic traffic remains important, because a visit brings the shopper onto the merchant's own site. But as AI search moves part of the information task onto the results page, clicks and sessions diverge from purchases, and fewer visits need not mean proportionally fewer customers. Evaluating search performance in this environment requires following the acquisition funnel beyond impressions, clicks, and sessions to realized transactions and revenue. The same applies to research on platform changes: the commercial incidence of generative search depends on which stages of the customer journey move onto the platform, not on how many external clicks disappear.

# References


Agarwal A, Hosanagar K, Smith MD (2015) Do organic results help or hurt sponsored search performance? *Information Systems Research* 26(4):695–713, URL http://dx.doi.org/10.1287/isre.2015.0593.

Agarwal S, Sen A (2026) The impact of google ai overviews on publisher traffic and user experience: Evidence from a field experiment. Working paper, URL http://dx.doi.org/10.2139/ssrn.6513059, sSRN 6513059.

Animesh A, Viswanathan S, Agarwal R (2011) Competing "creatively" in sponsored search markets: The effect of rank, differentiation strategy, and competition on performance. *Information Systems Research* 22(1):153–169, URL http://dx.doi.org/10.1287/isre.1090.0254.

Baye MR, De los Santos B, Wildenbeest MR (2016) Search engine optimization: What drives organic traffic to retail sites? *Journal of Economics & Management Strategy* 25(1):6–31, URL http://dx.doi.org/10.1111/jems.12141.

Berman R, Katona Z (2013) The role of search engine optimization in search marketing. *Marketing Science* 32(4):644–651, URL http://dx.doi.org/10.1287/mksc.2013.0783.

Burtch G, Lee D, Chen Z (2024) The consequences of generative ai for online knowledge communities. *Scientific Reports* 14:10413, URL http://dx.doi.org/10.1038/s41598-024-61221-0.

Callaway B, Sant'Anna PHC (2021) Difference-in-differences with multiple time periods. *Journal of Econometrics* 225(2):200–230, URL http://dx.doi.org/10.1016/j.jeconom.2020.12.001.

Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave E, Ott M, Zettlemoyer L, Stoyanov V (2020) Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451, URL http://dx.doi.org/10.18653/v1/2020.acl-main.747.

de Chaisemartin C, D'Haultfoeuille X (2020) Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9):2964–2996, URL http://dx.doi.org/10.1257/aer.20181169.

Dukes A, Liu L (2016) Online shopping intermediaries: The strategic design of search environments. *Management Science* 62(4):1064–1077, URL http://dx.doi.org/10.1287/mnsc.2015.2176.

Ghose A, Yang S (2009) An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science* 55(10):1605–1622, URL http://dx.doi.org/10.1287/mnsc.1090.1054.

Goldstein A, Oestreicher-Singer G, Barzilay O, Yahav I (2022) Are we there yet? analyzing progress in the conversion funnel using the diversity of searched products. *MIS Quarterly* 46(4):2015–2054, URL http://dx.doi.org/10.25300/MISQ/2022/15524.

Goodman-Bacon A (2021) Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225(2):254–277, URL http://dx.doi.org/10.1016/j.jeconom.2021.03.014.

Google (2024a) Ai overviews in search are coming to more places around the world. The Keyword, URL https://blog.google/products-and-platforms/products/search/ai-overviews-search-october-2024/, october 28, 2024.

Google (2024b) Generative ai in search: Let google do the searching for you. The Keyword, URL https://blog.google/products-and-platforms/products/search/generative-ai-google-search-may-2024/, may 14, 2024.

Google (2024c) New ways to connect to the web with ai overviews. The Keyword, URL https://blog.google/products-and-platforms/products/search/new-ways-to-connect-to-the-web-with-ai-overviews/.

Google (2025a) Ai overviews expand to over 200 countries and territories, more than 40 languages. The Keyword, URL https://blog.google/products-and-platforms/products/search/ai-overview-expansion-may-2025-update/, may 20, 2025.

Google (2025b) We're bringing the helpfulness of ai overviews to more countries in europe. The Keyword, URL https://blog.google/feed/were-bringing-the-helpfulness-of-ai-overviews-to-more-countries-in-europe/, march 25, 2025.

Gu C, Wang Y (2022) Consumer online search with partially revealed information. *Management Science* 68(6):4215–4235, URL http://dx.doi.org/10.1287/mnsc.2021.4104.

Jiang DL, Ye S, Zhao L, Gu B (2025) Do reductions in search costs for partial information on online platforms lead to better consumer decisions? evidence of cognitive miser behavior from a natural experiment. *Information Systems Research* 36(3):1780–1798, URL http://dx.doi.org/10.1287/isre.2022.0432.

Kaiser M, Schulze C (2026) Frontiers: Chatgpt referrals to e-commerce websites: How do llms compare against traditional channels? *Marketing Science* URL http://dx.doi.org/10.1287/mksc.2025.0489, articles in Advance.

Khosravi M, Yoganarasimhan H (2026) Impact of ai search summaries on website traffic: Evidence from google ai overviews and wikipedia. Working paper, URL http://dx.doi.org/10.2139/ssrn.6164926, sSRN 6164926.

Padilla N, Lam HT, Lambrecht A, Hollenbeck B (2026) The impact of llm adoption on online user behavior. Working paper, URL https://ssrn.com/abstract=5393256, sSRN 5393256, revised September 2026.

Shan G, Qiu L (2026) Examining the impact of generative ai on users' voluntary knowledge contribution: Evidence from a natural experiment on stack overflow. *Information Systems Research* 37(2):1021–1041, URL http://dx.doi.org/10.1287/isre.2023.0332.

Shi SW, Trusov M (2021) The path to click: Are you on it? *Marketing Science* 40(2):344–365,

URL http://dx.doi.org/10.1287/mksc.2020.1253.

Sun L, Abraham S (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2):175–199, URL http://dx.doi.org/10.1016/j.jeconom.2020.09.006.

Weber TA, Zheng ZE (2007) A model of search intermediaries and paid referrals. *Information Systems Research* 18(4):414–436, URL http://dx.doi.org/10.1287/isre.1070.0139.

Yang S, Ghose A (2010) Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science* 29(4):602–623, URL http://dx.doi.org/10.1287/mksc.1090.0552.

Zhuang M, Fang E, Lee J, Li X (2021) The effects of price rank on clicks and conversions in product list advertising on online retail platforms. *Information Systems Research* 32(4):1412–1430, URL http://dx.doi.org/10.1287/isre.2021.1039.

# Appendix

## A Sample Construction and Rollout Cohorts

The unit of analysis is an e-commerce domain in a market, observed weekly. The primary panel contains 4,658 domains, 97,236 domain–country units, 131 markets, and 62 calendar weeks, for a total of 5,180,452 unit-week observations. These units account for 99% of pre-treatment Google Organic sessions among the 7,083 eligible domains with linked GSC and GA data, and the selection rule uses only pre-treatment information. A fully balanced grid of all units and weeks would contain 6,028,632 unit-weeks; 848,180 cells fall outside a unit's valid measurement window, and 41,093 units are absent in at least one week. The estimator uses within-unit changes for which both the cohort-specific base week and the comparison week are observed, so each unit contributes to the group-time comparisons its data support. Table A.1 summarizes the sample and estimand.

**Table A.1:** Sample and Estimand

| Item | Value |
|---|---|
| Eligible linked e-commerce domains | 7,083 |
| Domains in primary sample | 4,658 |
| Domain–country units | 97,236 |
| Markets | 131 |
| Unit-week observations | 5,180,452 |
| Calendar weeks | 62 |
| Treated rollout cohorts | 5 |
| Post-treatment window | $e = 2, \ldots, 8$ |
| Comparison group | Not-yet-treated markets |
| Inference | Clustered by country |

*Notes.* The 99% coverage refers to pre-treatment Google Organic sessions within the eligible universe of linked domains.

Table A.2 describes the five rollout cohorts. The October 2024 cohort, which contains 112 markets, carries the largest share of the headline estimate (46.0%). The March 2025 cohort carries 29.4%. Because most markets had received AIOV by early 2025, this cohort is compared with two markets that were not yet treated over its event window. Appendix D reports how the results depend on this cohort.

**Table A.2:** Rollout Cohorts and Comparison Markets

| Rollout cohort | Treated markets | Treated units | Weight (%) | Min. comparison markets |
|---|---|---|---|---|
| May 2024 | 1 | 4,192 | 4.61 | 130 |
| August 2024 | 6 | 16,708 | 18.38 | 124 |
| October 2024 | 112 | 41,819 | 46.01 | 11 |
| Late 2024 | 1 | 1,426 | 1.57 | 11 |
| March 2025 | 9 | 26,751 | 29.43 | 2 |
| Not treated by panel end | 2 | 6,340 | – | – |

*Notes.* Weight is each cohort's contribution to the headline estimate over $e = 2, \ldots, 8$. Min. comparison markets is the smallest number of not-yet-treated markets available to the cohort over that window. All five cohorts are observed over the full window.

# B Outcomes and Estimation

The funnel margins in Section 3.4 are constructed at the unit-week level before estimation,

$$CVR_{it} = T_{it} - S_{it}, \qquad AOV_{it} = R_{it} - T_{it}, \qquad RPS_{it} = R_{it} - S_{it}, \tag{B.1}$$

and each is then estimated as its own outcome. GA transactions count completed purchases attributed to Google Organic; they do not identify distinct first-time customers. Because the pre-treatment adjustment profile is specific to each outcome, the directly estimated effect on a margin need not equal the difference between the separately estimated effects on its two components. For example, the separately estimated transaction and session effects differ by $-0.0019 - (-0.0281) = 0.0262$, while the directly estimated transaction-minus-session effect is 0.0254. The difference reflects the outcome-specific nuisance models, not the data, in which the identity $CVR_{it} = T_{it} - S_{it}$ holds exactly. Inference for each margin uses its direct estimate and therefore its own covariance structure.

For log-plus-one outcomes, the mapped effect $100[\exp(\widehat{ATT}) - 1]$ describes the change in the transformed outcome averaged across units. It differs from the percentage change in summed raw activity, which gives large merchant-markets more influence. Appendix E reports estimates on the count scale.

# C Pre-Treatment Coefficients and Placebo Dates

Table C.1 reports joint tests of the pre-treatment coefficients over two windows: the four weeks before rollout, which the main text reports, and the eight weeks before rollout. The transaction-minus-session margin, the outcome that carries the paper's central comparison, does not reject in either window ($p = .468$ and $p = .096$), and neither do sessions ($p = .391$ and $p = .077$). Transactions and revenue do not reject over the four-week window but reject over the eight-week window, where coefficients far from rollout are less uniform. The headline comparison between transactions and sessions is estimated directly on the margin, whose pre-treatment coefficients are flat over both windows.

**Table C.1:** Joint Tests of Pre-Treatment Coefficients

| Outcome | Weeks −4 to −2 | Weeks −8 to −2 |
|---|---|---|
| Sessions | 0.391 | 0.077 |
| Transactions | 0.511 | 0.001 |
| Revenue (signed log) | 0.425 | $< 0.001$ |
| Transaction-minus-session margin | 0.468 | 0.096 |

*Notes.* Entries are $p$-values from country-clustered joint tests that all pre-treatment coefficients in the window equal zero, under the preferred Callaway–Sant'Anna specification.

A separate placebo exercise assigns rollout dates four and eight weeks before the documented dates and re-estimates $CVR$, $AOV$, and $RPS$. None of the six placebo estimates is individually significant, and a joint maximum-$t$ test gives $p = .403$.

# D Cohort Heterogeneity and Comparison-Market Support

The headline estimate averages five rollout cohorts. Table D.1 reports each cohort's effect over weeks 2–8. The two largest cohorts by weight both show the pattern of the pooled estimate. In the October 2024 cohort, which carries 46% of the weight and has at least 11 comparison markets throughout, sessions fall ($-0.0241$, $p = .018$) and the transaction-minus-session margin rises ($0.0201$, $p = .046$). In the March 2025 cohort, sessions fall by more ($-0.0593$, $p = .003$) and the

margin rises by more (0.0604, $p < .001$), while transactions do not change (0.0015, $p = .790$). The two earliest cohorts, the United States in May 2024 and six markets in August 2024, show no session decline over this window.

**Table D.1:** Effects by Rollout Cohort, Weeks 2–8

| | May 2024 | Aug. 2024 | Oct. 2024 | Late 2024 | Mar. 2025 |
|---|---|---|---|---|---|
| Weight (%) | 4.61 | 18.38 | 46.01 | 1.57 | 29.43 |
| Sessions | 0.0085 | 0.0061 | −0.0241 | −0.0669 | −0.0593 |
| $p$-value | 0.120 | 0.585 | 0.018 | < 0.001 | 0.003 |
| Transactions | −0.0080 | −0.0033 | −0.0034 | 0.0102 | 0.0015 |
| $p$-value | 0.058 | 0.068 | 0.004 | < 0.001 | 0.790 |
| Transaction-minus-session | −0.0221 | −0.0099 | 0.0201 | 0.0766 | 0.0604 |
| $p$-value | < 0.001 | 0.374 | 0.046 | < 0.001 | < 0.001 |

*Notes.* Preferred Callaway–Sant'Anna estimates on the transformed scale, averaged over $e = 2, \ldots, 8$ within each cohort. The May 2024 and Late 2024 cohorts each contain one treated market.

Table D.2 reports the pooled estimates without the March 2025 cohort, computed in two ways. Reaggregation drops the cohort's contribution to the average and keeps the remaining cohort estimates unchanged. Refitting removes the cohort's units from the sample and re-estimates the model, which also removes them from the comparison group of earlier cohorts and leaves the October 2024 cohort with fewer comparison markets over its window. Under reaggregation, sessions fall by 1.49% and the transaction-minus-session margin rises by 1.08%. Under refitting, the corresponding estimates are −0.81% and +0.40%, and neither differs from zero ($p = .517$ and $p = .699$); the refitted model also rejects the four-week pre-treatment test for sessions ($p < .001$) and the margin ($p = .002$). Transaction effects stay small in every version. The March 2025 cohort therefore contributes a substantial part of the pooled gap between sessions and transactions.

**Table D.2:** Pooled Estimates Without the March 2025 Cohort

| | Sessions | Transactions | Revenue | Transaction-minus-session |
|---|---|---|---|---|
| Full sample | −2.77 | −0.19 | −0.54 | +2.57 |
| Excluding March 2025: reaggregation | −1.49 | −0.34 | −0.92 | +1.08 |
| Excluding March 2025: refit | −0.81 | −0.24 | −0.39 | +0.40 |

*Notes.* Mapped effects in percent, $100[\exp(\widehat{ATT}) - 1]$, averaged over $e = 2, \ldots, 8$.

Because the March 2025 cohort is compared with two markets, Table D.3 varies the comparison

markets used for this cohort. Across these checks, the cohort's session effect ranges from −11.03% to −2.45% and its transaction-minus-session margin from +3.19% to +10.52%. The direction of both is the same under every choice of comparison markets. The cohort's transaction effect ranges from −3.44% to +0.81%.

**Table D.3:** March 2025 Cohort: Sensitivity to Comparison Markets

| Outcome | Both comparison markets | Minimum | Maximum |
|---|---|---|---|
| Sessions | −5.76 | −11.03 | −2.45 |
| Transactions | +0.15 | −3.44 | +0.81 |
| Transaction-minus-session | +6.22 | +3.19 | +10.52 |

*Notes.* Mapped effects in percent for the March 2025 cohort, averaged over $e = 2, \ldots, 8$. Minimum and maximum are taken across alternative comparison-market definitions for this cohort.

# E Count-Scale Estimates

The headline estimand averages log-plus-one outcomes across units, so a small merchant-market and a large one carry similar influence. A count-scale estimand instead measures the percentage change in summed sessions and transactions and gives large merchant-markets more weight. We estimate count-scale group-time models for sessions and transactions on the same panel. Let $\mu_{1,k}$ and $\mu_{0,k}$ be the observed and counterfactual mean counts of outcome $k$ over treated unit-weeks with paired support. The count-scale effect is $100(\mu_{1,k} - \mu_{0,k})/\mu_{0,k}$, and the funnel gap is the transaction effect minus the session effect, in percentage points.

Table E.1 reports the results. In the full sample, summed sessions fall by 18.7% while summed transactions do not change (+0.1%), a gap of 18.7 percentage points ($p < .001$). The count-scale gap is concentrated in the March 2025 cohort, whose summed sessions fall by 44.1%. Without that cohort, the gap is negative and imprecise (−4.5 percentage points, $p = .177$, under reaggregation; −4.3, $p = .337$, under refitting). Excluding the largest 1% of units by pre-treatment traffic widens the full-sample gap to 51.2 percentage points. The count-scale estimates therefore depend heavily on which cohorts and which merchant-markets receive the most weight, and the log-scale headline

estimand is the more stable summary of the average merchant-market response.

**Table E.1:** Count-Scale Effects on Sessions and Transactions

| Sample | Sessions (%) | Transactions (%) | Gap (pp) | Gap $p$-value |
|---|---|---|---|---|
| Full sample | −18.68 | +0.07 | +18.75 | < 0.001 |
| Excluding March 2025: reaggregation | −0.08 | −4.59 | −4.51 | 0.177 |
| Excluding March 2025: refit | +7.30 | +2.99 | −4.31 | 0.337 |
| Excluding top 1% units by traffic | −38.67 | +12.57 | +51.24 | < 0.001 |

*Notes.* Percentage changes in mean counts relative to the estimated counterfactual, computed over treated unit-weeks with paired support and averaged over $e = 2, \ldots, 8$. Gap is the transaction effect minus the session effect, in percentage points. Inference uses country-clustered influence functions.

# F Additional Specifications

Table F.1 collects the additional specifications reported in Section 8. Each addresses a different threat: merchant-wide shocks, panel support, anticipation, interference across a merchant's markets, and the weighting of conventional two-way fixed effects.

**Table F.1:** Additional Specifications

| Specification | Result |
|---|---|
| Domain-by-week and domain–country fixed effects | Sessions −2.32% ($p$ = .005); transactions −0.42% ($p$ = .001); transaction-minus-session +1.94% ($p$ = .010). Static post-rollout indicator. |
| Cross-market exposure | Raising the share of a merchant's other markets already treated from 0 to one third changes sessions by +0.36% (95% CI −2.7% to +3.5%) in markets not yet treated; transactions essentially unchanged. |
| Balanced sample | Sessions −2.82%; transaction-minus-session +2.64%. |
| Common-support sample | Sessions −2.99%; transaction-minus-session +2.69%. |
| Placebo rollout dates (−4, −8 weeks) | Six estimates for $CVR$, $AOV$, $RPS$, none individually significant; joint maximum-$t$ $p$ = .403. |
| Goodman–Bacon decomposition (balanced TWFE) | Exact reconstruction; all weights nonnegative; 31.2% of weight from comparisons with already-treated controls. |

# G Query-Level Position and Intent Measures

The position and intent analyses use query-level GSC records aggregated to the merchant–market–week. Each query's position group is set from its pre-treatment GSC position and held fixed after rollout, so that the outcome measures the clicks generated by an established position, not changes in the position itself. Because GSC records all links inside an AI Overview at the position of the overview block, fixing positions before rollout also keeps the groups unaffected by this recording rule.

Intent is classified into four nonexclusive labels: commercial investigation, informational, navigational, and transactional. Let $p_{qk}$ be the predicted probability of label $k$ for query $q$. For merchant–market $i$, week $t$, and position group $b$, the outcome is

$$C_{it}^{k,b} = \sum_{q \in b} p_{qk} C_{qit}, \qquad Y_{it}^{k,b} = \log\left(1 + C_{it}^{k,b}\right). \tag{G.1}$$

A query contributes to every label in proportion to its predicted probability, and the probabilities are not renormalized across labels, so the four outcomes overlap and do not partition total clicks. The cross-intent contrasts in Table 2, Panel B, are estimated directly as differences of unit-week outcomes, not computed from the rounded estimates in Panel A.

# H Data

The analysis links two Google measurement systems for each merchant. Google Search Console records impressions, outbound clicks, query text, and position on Google's results page. Google Analytics records Google Organic sessions, completed transactions, and attributed revenue on the merchant's site. The primary panel is built at the domain–country–week level; the query-level analyses use a separate GSC query history linked to the same domains and markets. The merchant data are covered by data-sharing agreements and cannot be publicly released.