# Selling Pseudo-Science: Is More Regulation Needed? *

Current version: November 2026

Rui A[†] Elisabeth Honka[‡]

**Abstract**

We study the effects of pseudo-scientific claims on consumer shopping behavior. Pseudo-scientific claims are product claims that adopt the appearance of scientific evidence (e.g., "clinical study") but lack the rigor of valid scientific methodology. Such claims may mislead consumers if they are perceived as more credible than they actually are. We conduct an incentive-aligned online experiment at a leading beauty retailer to measure how pseudo-scientific claims affect consumer search and purchases. Furthermore, we collect data on consumer beliefs and on the perceived credibility of these claims to investigate the underlying mechanism. We find that pseudo-scientific claims cause consumers to terminate their search earlier and to be more likely to purchase a product carrying such a claim. The effect is driven by consumers misinterpreting numerical information, rather than by costly signaling or credibility transfer from medical research. We discuss implications for regulation.

**Keywords:** misinformation, regulation, field experiment

**JEL Classification:** D83, L50, M37

*We thank Brett Hollenbeck, Tai Lam, Franklin Shaddy, and Stephen Spiller for their suggestions. We thank the participants of the 2025 Marketing Science conference and seminar/brown bag participants at UCL, UCLA, London Business School, and ESMT for their comments. All errors are our own.

[†]University of California Los Angeles, rui.a.phd@anderson.ucla.edu.

[‡]University of California Los Angeles, elisabeth.honka@anderson.ucla.edu.

# 1 Introduction

Firms often use scientific evidence to promote their products. However, such claims are not always reliable, even though the Federal Trade Commission (FTC) mandates that claims be truthful, non-misleading, and adequately substantiated (Federal Trade Commission 2024). Some claims, while factually accurate, may still mislead consumers. A claim is misleading if (i) it influences people's decision-making and (ii) it implies a higher level of credibility than is warranted.[1] Credibility, in this context, encompasses both internal validity – the soundness of the study design – and external validity – the generalizability of the results. We refer to such claims as "pseudo-scientific claims."

Typically, these pseudo-scientific claims are efficacy claims that are based on a collection of anecdotal evidence presented in scientific-sounding language. For example, an ad for a supplement claiming that the product improves memory and showing the results of a consumer survey reporting that more than 75% of customers noticed memory improvement; a skincare product claiming to reduce wrinkles and showing the results of a survey reporting that 90% of participants saw firmer skin after using the product. Although these claims may summarize genuine consumer experiences, they do not establish that the product *caused* the reported improvement. As the FTC notes, consumer experiences may reflect placebo effects or other factors unrelated to the product.[2] A claim is deceptive if it is likely to mislead reasonable consumers and the misleading interpretation is material to their decisions.[3] A pseudo-scientific claim is therefore deceptive if consumers read anecdotal evidence as establishing that the product works and that reading affects what they buy.

Whether pseudo-scientific claims affect consumer behavior is an open question. If consumers are sophisticated enough to discern firms' intentions and/or to access information through alternative channels, pseudo-scientific claims may have no effect on their purchase

[1] https://www.ftc.gov/legal-library/browse/ftc-policy-statement-deception

[2] Federal Trade Commission, *Health Products Compliance Guidance* (2022), p. 14; https://www.ftc.gov/system/files/ftc_gov/pdf/Health-Products-Compliance-Guidance.pdf.

[3] https://www.ftc.gov/legal-library/browse/ftc-policy-statement-deception

decisions or even a negative effect through consumer backlash. Conversely, if consumers take the claims at face value or interpret them as quality signals, the effect may be positive. Existing evidence does not settle the matter. Prior work documents declines in monthly revenue and product page visits after the FTC identifies a false claim and requires its removal (Rao and Wang 2017*a*, Rao 2022*a*). These results, however, conflate the effect of claim removal with the effects of negative publicity and other enforcement-related factors, and the effect of pseudo-scientific claims in the absence of intervention may differ in magnitude – or even in sign – from the effect of their removal.

We conduct an incentive-aligned online experiment at a leading online beauty retailer, using a browser extension that creates exogenous variation in whether participants encounter pseudo-scientific claims on a product's detail page. Across a sample of 2,366 participants shopping for anti-aging moisturizers, we find that pseudo-scientific claims have two robust effects on consumer behavior. First, they cause consumers to terminate their search earlier, reducing the total number of products examined. Second, they shift choice toward products that carry the claim: participants in the treatment condition are significantly more likely to purchase a product with a clinical results claim than participants in the control condition. Instrumental-variables estimates of the effect of *actually viewing* a claim, rather than merely being potentially exposed to one, are larger still. Pseudo-scientific claims thus systematically change both how much information consumers gather and which products they choose.

To identify the mechanism behind these effects, we run a second experiment that systematically varies the components of a clinical results claim—the presence of the phrase "clinical study," the inclusion of numerical efficacy percentages, and a disclaimer clarifying what "clinical study" means in skincare. The mere reference to a "clinical study," without accompanying numbers, has no detectable effect on consumer choice, ruling out both costly signaling and credibility transfer from medical research as drivers of our main results. Adding numerical percentages, however, produces a large and statistically significant increase in the probability of choosing the treated product, and this increase is not atten-

uated by the disclaimer. A further credibility-rating experiment reinforces the conclusion: respondents rate claims as more credible when reported percentages are higher and when the cited sample size is larger, but they do not distinguish between a "clinical study" and a "consumer-perception study." The mechanism, we conclude, is numerical misinterpretation: consumers treat reported percentages and sample sizes as evidence of effect size rather than as what they actually are – the prevalence of a subjective judgment within a non-randomized study.

This paper makes three contributions. First, we provide the first causal evidence on consumer response to pseudo-scientific claims in the absence of regulatory intervention, isolating the effect of the claims themselves from the effects of claim removal and negative publicity that have characterized prior work. Second, we contribute to the literature on firms' strategic provision of information by identifying a route to consumer deception that does not require firms to lie or to attach false content to credible-sounding labels: selective reporting of numerical information from weakly designed studies is sufficient to shift consumer behavior. Third, we inform the ongoing regulatory debate. Because the locus of the problem is the numerical content of claims rather than the language surrounding them, regulatory leverage lies in the numbers themselves and in consumers' ability to interpret them, not in the terminology of clinical testing.

The remainder of this paper is organized as follows: In the next section, we review the relevant literature. In Section 3, we show evidence on the prevalence of pseudo-scientific claims. We discuss our field experiment and its results in Section 4. We present evidence for the underlying mechanism in Section 5 and discuss implications for regulation in Section 6. Finally, we conclude in Section 7.

# 2 Relevant Literature

This paper is related to four streams of literature: on strategic information provision, on firms' deceptive behavior, on the effects of misleading claims on consumer behavior, and on the debunking of misinformation.

The first stream of literature studies firms' strategic provision of information. Grossman (1981) and Milgrom (1981) show that when disclosure is costless and verifiable, higher-quality firms disclose in equilibrium, forcing all but the lowest-quality firms to follow. Jovanovic (1982) shows that when disclosure is costly, only firms above a quality threshold disclose. A large subsequent literature has relaxed the assumptions of costless and verifiable disclosure and of fully rational receivers, identifying a range of conditions under which firms strategically withhold information (see Dranove and Jin 2010 for a review).

Empirical work finds partial unraveling and shows that disclosure changes what firms sell. For example, Mathios (2000) finds that before nutrition labels became mandatory, every low-fat salad dressing carried a label while most high-fat dressings did not. Ippolito and Mathios (1990) find that after the ban on health claims for foods was lifted in the mid-1980s, consumption of high-fiber cereals rose and new cereals contained more fiber. Jin and Leslie (2003) find that the requirement to post restaurant hygiene grades raised inspection scores and reduced foodborne-illness hospitalizations. Pseudo-scientific claims fit into this framework as a form of costly disclosure: a firm pays to conduct a study and reports its result.

The second stream of literature documents false claims for a variety of products and services. Zinman and Zitzewitz (2016) document the extent of deceptive advertising by ski resorts. They find that ski resorts report 12%–23% more snowfall than government sources indicate. This tendency to exaggerate snow conditions is particularly pronounced on weekends, when the incentives to over-report are strongest. Heneghan et al. (2012), Fowler et al. (2015), and Peake et al. (2018) review advertising claims for cosmetics, sports-performance products, and consumer health-monitoring devices using expert judgments, and

find that the majority of claims lack reliability.

The third stream studies the effects of misleading claims on consumer behavior. Rao and Wang (2017*b*) investigate the impact of exposing deceptive advertising claims on demand for four products. The findings indicate a monthly revenue loss ranging from 12% to 67% across these products. This decline is primarily attributed to newcomers—households that began buying only after the false claim appeared. Rao (2022*b*) examines another form of deceptive advertising, in which advertisements mimic news articles. Using a natural experiment in which the FTC shut down over 150 fake news websites, the author observes a 22% to 41% increase in the share of merchants that received no visits in the month following the shutdown. This suggests that fake news websites had been effective in boosting consumer interest in the advertised products. These two studies capture the combined effect of the removal of misleading claims and the accompanying negative publicity. In contrast, we aim to recover the effect of misleading claims only.

The last stream of literature examines the debunking of misinformation. Given its prevalence, Fong et al. (2024) evaluate the efficacy of debunking misinformation about consumer products through a randomized experiment. Their findings reveal that the disbelief induced by misinformation can be reversed through debunking efforts. Both regulators and competitors are found to be effective in debunking; however, the study suggests that, in equilibrium, competitors' best response is to join the misinformation campaign rather than engage in debunking. Similarly, Shi et al. (2022) examine whether various information channels—such as research articles, news, health-related web articles, and online customer reviews—can correct misleading health information. They find that news articles and online customer reviews are largely ineffective in correcting biases, tending instead to reinforce rather than rectify misleading information. Research articles, meanwhile, are found to be too slow to effectively address the misinformation problem.

The psychology literature on correcting false beliefs is similarly mixed: misinformation often continues to influence judgment even after it has been retracted (Lewandowsky et al.

2012), corrective information fails to reduce, and can even reinforce, belief in the original misinformation (Nyhan and Reifler 2010), and corrective messages' effectiveness depends on how the correction is delivered and on audiences' prior beliefs (Chan et al. 2017). Consistent with this mixed evidence, we find that a corrective disclaimer does not attenuate the effect of pseudo-scientific claims on consumer choice.

# 3 Prevalence and Characteristics of Pseudo-Scientific Claims

A clinical results claim typically reports results related to the product's efficacy from a study, followed by a brief note on the study. Figure 1 shows an example: four statements such as "93% agreed eye area looks more radiant," followed by a note that they are "based on the self-assessment survey portion of an 8-week clinical study on 30 women ages 20-49." These statements and the note are all that consumers learn about the study.




**Figure 1:** Example of a Clinical Results Claim

While Figure 1 shows a typical example of a pseudo-scientific claim, we scraped information from sephora.com for all products in four categories (skincare, haircare, make-up,

and bath & body) to gather representative data on the prevalence and characteristics of pseudo-scientific claims.[4] Clinical results claims are common across all four categories: 52% of skincare, 40% of haircare, 36% of make-up, and 29% of bath & body products contain such claims.

Table 1 contains descriptives of the clinical results claims. We define a study as a test, survey, or panel cited in a product's clinical results claim; a claim may cite more than one. The products in our data which make clinical results claims cite 1.4 studies per product on average. Three patterns stand out. First, the studies are small and short. The median sample size of a study is 33 participants and the median study length is 28 days. Second, the wording of the study type varies widely. "Clinical study," the most common label, accounts for only 12% of studies, the three most common phrases together for 27%. And finally, the studies report high numbers. Agreement statements, such as "93% agreed skin looked firmer," report an average of 94% of participants agreeing. Statements of measured changes, such as "+36% hydration," report an average change of 72%.

[4]Sephora was the largest U.S. online beauty store by net sales in 2025 (Statista 2025).

**Table 1:** Characteristics of Clinical Results Claims

| | Skincare | Haircare | Make-up | Bath & Body | All |
|---|---|---|---|---|---|
| *Panel A: Descriptives by Product* | | | | | |
| Number of Products | 2,808 | 1,616 | 2,213 | 780 | 7,417 |
| Number of Products with Clinical Results Claims | 1,464 | 652 | 806 | 227 | 3,149 |
| Percentage of Products with Clinical Results Claims | 52% | 40% | 36% | 29% | 43% |
| Average Number of Studies Cited per Product with Claims | 1.46 | 1.32 | 1.29 | 1.31 | 1.38 |
| *Panel B: Descriptives by Study* | | | | | |
| Number of Studies Cited | 2,135 | 860 | 1,043 | 297 | 4,335 |
| Three Most Common Study Types: | | | | | |
| "Clinical Study" | 13% | 10% | 11% | 15% | 12% |
| "Consumer Perception Study" | 7% | 11% | 7% | 13% | 8.0% |
| "Consumer Study" | 6% | 7% | 8% | 10% | 7% |
| Median Number of Participants per Study | 33 | 40 | 33 | 37 | 33 |
| Median Study Length in Days | 28 | 14 | 7 | 21 | 28 |
| *Panel C: Descriptives by Statement* | | | | | |
| Average Percentage of Agreement Statements | 93% | 93% | 95% | 95% | 94% |
| Average Percentage Change in Measured-Change Statements | 60% | 79% | 95% | 50% | 72% |

A product has a *clinical results claim* if its product detail page contains a "Clinical Results" block. A *study* is one test, survey, or panel cited in the block; a block may cite several. *Study type* is the phrase the claim uses for the study; the three most common phrases are shown. *Participants* is the number of people in the study, for the 73% of studies that report it. *Study length* is the longest follow-up at which a result is reported, for the 60% of studies that report one. *Agreement statements* (7,252) are percentages of participants who agreed, said, or saw something ("93% agreed skin looked firmer"); *measured changes* (1,195) are percentage changes in a measured quantity ("+36% hydration"). Each row averages the reported percentages over statements.

To understand which products are most likely to carry clinical results claims, we estimate linear probability models regressing an indicator for the presence of clinical results claims on product attributes. The results are shown in Table 2. In all four categories, products with more reviews are more likely to display clinical results claims. The estimates for other product attributes are mixed, e.g., a higher price is associated with a higher probability of a pseudo-scientific claim is three of the four categories and a lower probability of a pseudo-scientific claim in the fourth category. Higher-rated products are more likely to display clinical results claims in only two of the four categories. We conclude that pseudo-scientific claims are used without a clear overall pattern in terms of which types of products are most likely to carry them.

**Table 2:** Determinants of Whether a Product Displays Clinical Results Claims

| | Skincare (i) | Haircare (ii) | Make-up (iii) | Bath & Body (iv) |
|---|---|---|---|---|
| log(Price per oz) | 0.0337** | 0.0740*** | -0.0459*** | 0.0801*** |
| | (0.0106) | (0.0161) | (0.0091) | (0.0203) |
| Rating Score | 0.0825*** | 0.0642* | -0.0480* | -0.0049 |
| | (0.0218) | (0.0291) | (0.0221) | (0.0387) |
| log(1+Number of Reviews) | 0.0320*** | 0.0792*** | 0.0496*** | 0.0642*** |
| | (0.0069) | (0.0103) | (0.0090) | (0.0138) |
| log(1+Loves Count) | -0.0068 | -0.0447*** | -0.0449*** | -0.0199 |
| | (0.0058) | (0.0122) | (0.0099) | (0.0141) |
| Dummy for Middle-Aged Brand | 0.1236*** | -0.0320 | 0.0944*** | 0.0157 |
| | (0.0222) | (0.0287) | (0.0241) | (0.0403) |
| Dummy for Old Brand | 0.0548* | -0.0960** | 0.1246*** | 0.0224 |
| | (0.0250) | (0.0301) | (0.0235) | (0.0482) |
| Dummy for Missing Price | -0.0061 | 0.1803** | -0.3054*** | -0.0565 |
| | (0.0307) | (0.0600) | (0.0223) | (0.0481) |
| Dummy for Rating Score Unavailable | 0.5213*** | 0.4875*** | -0.1685 | 0.0918 |
| | (0.1196) | (0.1477) | (0.1387) | (0.1903) |
| Subcategory Fixed Effects | Yes | Yes | Yes | Yes |
| Number of Observations | 2,808 | 1,616 | 2,213 | 780 |
| Adjusted $R^2$ | 0.072 | 0.093 | 0.114 | 0.079 |

Robust standard errors in parentheses.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.
*Dummy for Missing Price* equals one for products with no per-ounce price: multi-item sets and kits, tools and devices, and products whose listed size is missing or is a count of items; their price term is held at the category mean. *Dummy for Rating Score Unavailable* equals one if the product has no reviews. Brand-age categories: *young* (founded 2015 or later), *middle-aged* (1995–2014), and *old* (before 1995). Product subcategories for skincare: Cleanser, Toner & Essence, Serum & Treatment, Face Oil, Moisturizer, Eye Care, Mask & Exfoliant, Sunscreen, and "Other." For haircare: Shampoo, Conditioner, Styling, and "Other." For makeup: Face Makeup, Eye Makeup, Lip Makeup, Tools & Brushes, and "Other." For bath & body: Lotions & Moisturizers, Body Wash & Cleansers, Deodorant & Care, Scrubs & Exfoliants, and "Other."

# 4 Do Pseudo-Scientific Claims Affect Consumers?

## 4.1 Experimental Design

We conduct a pre-registered, incentive-compatible field experiment at a leading online beauty retailer, using a browser extension that modifies the content displayed on the retailer's product detail pages and tracks participants' browsing and purchase behavior.[5] The design creates exogenous variation in the clinical results claims participants encounter while hold-

[5] Pre-registered on AsPredicted.org.

ing the underlying products and the retailer's live interface constant, allowing us to identify the causal effect of these claims on consumer decisions. Compared to a conjoint study, this approach has two advantages: it embeds the experiment in a real shopping environment, yielding higher external validity; and it generates rich data on browsing behavior – what participants view, how long they linger, and in what order – rather than on final choices alone.

We recruited 4,436 participants between April 28 and June 30, 2025, through CloudResearch Connect, an online survey platform. The sample was restricted to women living in the United States who reported using a facial cream more than twice a week. After providing information on their socio-demographic characteristics and online shopping habits, participants installed the browser extension and completed a shopping task on the modified version of the retailer's website. Upon completing the shopping task, participants made a second product choice and evaluated the credibility of clinical results claims shown alongside real products from our sample.[6]

Participants were told that the study "explores how people make shopping decisions in an online store" and that the store "has been modified for research purposes," without specifying the modifications. Instructions were identical across conditions and did not mention anything about "clinical study claims." Participants were instructed to shop as they normally would, browse products, and select a product they liked. We relabeled the "Add to basket" button on product detail pages as "Add to wishlist," so participants had to click through to at least one product detail page to complete the task. Clicking on "Add to wishlist" ended the shopping task without proceeding to a shopping basket and recorded the product choice.

The experiment was incentive-compatible: participants received $3 for completing the study and had a 1-in-30 chance of winning a $90 prize: the product they chose during the shopping task, plus a gift card for the difference between $90 and the product's price including tax and shipping.

[6]The complete experimental survey is available at `experimental_survey.pdf`.

The experimental design involves three randomization steps—product, product ranking, and participant randomization—which we describe in turn.

### 4.1.1 Product Ranking Randomization

Participants encounter products on a category list page; Figure 2 displays an example. Because prior research shows that the position of a product in a list substantially affects the probability that it is viewed and purchased (Ursu 2018, Donnelly et al. 2024, Compiani et al. 2024, Greminger 2023), we randomize product rank at the participant level. We verify that the randomization was successful by regressing each product's rank on a full set of product fixed effects; no coefficient is statistically significantly different from zero.

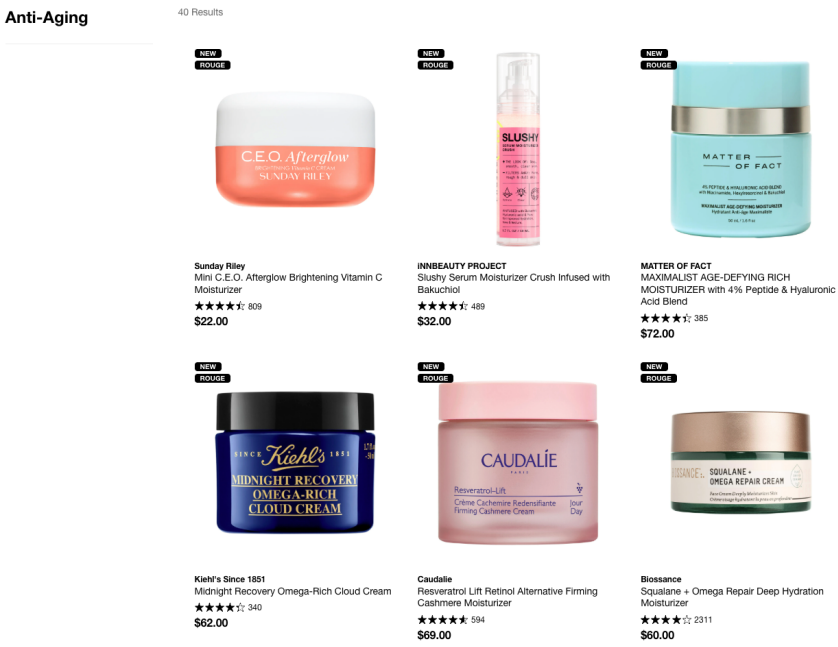



**Figure 2:** Example of a Product List Page

### 4.1.2 Product Randomization

Our study focuses on anti-aging moisturizers. From the full set of products sold by the retailer in this category, we exclude items priced above $90, items targeted at male consumers, and

items not classified by the retailer as moisturizers.[7] For products available in multiple sizes, we display only the price and size of the standard option. These criteria leave us with 40 anti-aging moisturizers.[8]

The retailer displays pseudo-scientific claims in two places on a product detail page: the product image gallery and a part of the product description labeled "Clinical Results." Of the 40 products, 33 carried such a claim on the retailer's website. The browser extension controls whether a product's claim is displayed. When the claim is off, the extension removes the claim image from the gallery and the "Clinical Results" part of the description. When the claim is on, both appear. All other page content is the retailer's live content and is identical across participants.

We randomly assigned 20 of the 40 products to be eligible to display a claim.[9] To choose the 20 moisturizers, we pairwise-matched products on price, consumer ratings, number of reviews, and three measures of brand equity – brand awareness, brand loyalty, and perceived quality – using an adapted version of the scale developed by Yoo and Donthu (2001). The brand equity measures were collected through a separate survey of 413 individuals conducted prior to the main experiment.[10]

Within each matched pair, we randomly assigned one product to Group I and the other to Group II, and we hold this assignment fixed for all participants. Group I products can display pseudo-scientific claims; Group II products never do. 18 of the 20 Group I products display their own claim and two a constructed one.[11] Table 3 shows that Group I and Group II products are balanced across observable attributes: none of the estimated differences between the two groups is statistically significant.

[7] We follow the retailer's own website categorization when defining product types.

[8] One product was removed from the retailer's website immediately before the study launched and was therefore excluded from the experiment, leaving 40 products.

[9] We chose half of the products to display a claim for two reasons. First, the power to detect any effect is largest when half of the products carry claims. Second, this allocation makes claim prevalence similar to that in the retailer's skincare category, where 52% of products show clinical results claims (Section 3).

[10] Web Appendix A provides details on the brand equity survey.

[11] Each constructed claim is based on a real claim from one of the 33 products in our sample that display such claims, with each reported percentage shifted by a draw from $U(-3, 3)$ percentage points and capped at 100%; its image matches the layout and style of the real claim images.

Randomizing within matched pairs ensures that the products displaying a claim are not selected on price, popularity, or brand, which predict claim use in the market (Table 2).

**Table 3:** Covariate Balance of Attributes for Group I and Group II Products

| | **Price** | **Avg. Rating** | **Number of Reviews** | **Brand Awareness** | **Brand Loyalty** | **Perceived Quality** |
|---|---|---|---|---|---|---|
| Dummy for Group I | -1.300 | -0.018 | -293.850 | 0.245 | 0.085 | 0.042 |
| | (5.062) | (0.070) | (415.608) | (0.841) | (0.209) | (0.140) |
| Constant | 60.600*** | 4.412*** | 1081.350*** | 8.882*** | 5.650*** | 6.912*** |
| | (3.579) | (0.050) | (293.879) | (0.595) | (0.148) | (0.099) |
| Number of Observations | 40 | 40 | 40 | 40 | 40 | 40 |
| $R^2$ | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |

Standard errors in parentheses.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Throughout the online experiment, we hold constant which products belong to Group I and which products belong to Group II. In the control condition, all products (belonging to both Group I and Group II) were shown without pseudo-scientific claims. In the treatment condition, Group I products displayed clinical results claims, while Group II products did not.[12]

### 4.1.3 Participant Randomization

Participants were stratified by age and education level and, within each stratum, randomly assigned to either the control or the treatment condition. In the control condition, participants visited a version of the retailer's website in which no product displayed clinical results claims. In the treatment condition, participants visited a version in which Group I products displayed clinical results claims and Group II products did not.

All participants saw the same set of products. The treatment manipulation was applied only on product detail pages: list-page content was identical across conditions, and differences

[12] An alternative experimental design would be to re-draw which product in a pair displays a claim for each participant, so that every product appears both with and without a claim across participants. Both experimental designs yield unbiased estimates of the same average treatment effect over possible product assignments. However, re-drawing requires a separate version of the retailer's website for every participant – making the technical implementation significantly more challenging – whereas holding Group I fixed requires only two versions, keeping the implementation magnitudes simpler.

appeared exclusively in the detail-page content, specifically in whether clinical results claims were present. Figure 3 shows examples of product detail pages for Group I and Group II products in each condition.

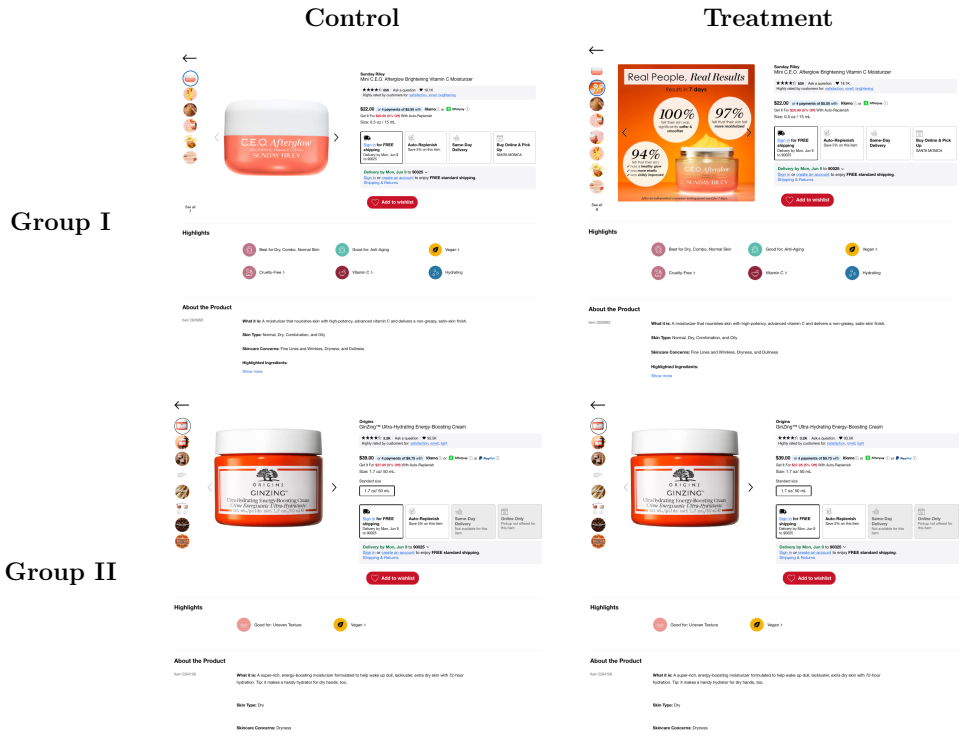



**Figure 3:** Examples of Product Detail Pages in the Control and Treatment Conditions for Group I and II Products

Table 4 reports balance checks on demographic and behavioral characteristics across the control and treatment conditions; no characteristic differs significantly between groups.

**Table 4:** Randomization Check: Demographic and Behavioral Balance

| | Control | Treatment | p-value |
|---|---|---|---|
| **Demographics** | | | |
| *Age* | | | |
| 18–24 | 20% | 17% | 0.168 |
| 25–34 | 37% | 38% | 0.679 |
| 35–44 | 25% | 24% | 0.484 |
| 45+ | 19% | 21% | 0.112 |
| *Education* | | | |
| Less than BA | 40% | 42% | 0.221 |
| BA/Grad | 60% | 57% | 0.193 |
| Unknown | 0% | 0% | 0.336 |
| *Race* | | | |
| Non-White | 35% | 36% | 0.740 |
| White | 64% | 64% | 0.688 |
| Unknown | 1% | 1% | 0.658 |
| *Annual income* | | | |
| Below $50,000 | 45% | 45% | 0.970 |
| $50,000 - $100,000 | 37% | 35% | 0.271 |
| Above $100,000 | 15% | 17% | 0.236 |
| Unknown | 2% | 2% | 0.622 |
| *Employment* | | | |
| Full-time | 46% | 46% | 0.851 |
| Other | 54% | 53% | 0.671 |
| Unknown | 0% | 1% | 0.065 |
| **Behavioral** | | | |
| *Use frequency* | | | |
| >1×/day | 43% | 41% | 0.241 |
| *Skincare importance* | | | |
| 4–5 | 74% | 75% | 0.727 |
| *Purchase frequency* | | | |
| ≤ monthly | 37% | 37% | 0.863 |
| monthly | 44% | 44% | 0.826 |
| ≥ twice/month | 19% | 20% | 0.626 |
| *Purchase amount* | | | |
| < $10 | 12% | 13% | 0.594 |
| $10–$40 | 49% | 48% | 0.567 |
| $40–$70 | 26% | 27% | 0.852 |
| ≥ $70 | 13% | 13% | 0.933 |
| Number of Observations | 1153 | 1213 | |

**Notes.** Survey questions: *How frequently do you use a facial cream or moisturizer?* (Use frequency); *How important is skincare in your daily life?* (Skincare importance); *How frequently do you purchase skincare products?* (Purchase frequency); *How much do you spend on skincare products per month?* (Purchase amount). *Skincare importance* (4 - 5) means the consumer thinks skincare is important (selected 4–5 on a 1–5 scale).

### 4.1.4 What the Experimental Design Identifies

In this section, we establish which causal effect the experiment measures using the potential outcomes framework. Consider a store with six products in three matched pairs, $\{1, 2\}$, $\{3, 4\}$, and $\{5, 6\}$, and suppose product randomization has placed products 1, 3, and 5 in Group I, i.e., $G = \{1, 3, 5\}$. In the treatment condition, products 1, 3, and 5 display their claims; in the control condition, no product does (see Figure 4 for illustration). Each participant is randomly assigned to one of the two conditions and chooses one product.

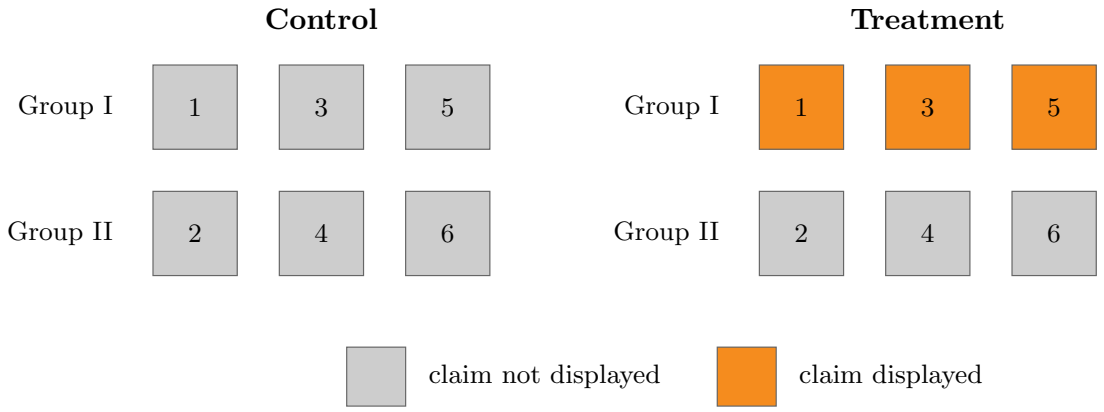



**Figure 4:** Six-Product Illustration

Let $D_i = 1$ if participant $i$ is assigned to the treatment condition and $D_i = 0$ if she is assigned to the control condition. For each product $j$, participant $i$ has two potential outcomes: $Y_{ij}(1) = 1$ if she would choose product $j$ in the treatment condition and 0 otherwise, and $Y_{ij}(0) = 1$ if she would choose product $j$ in the control condition and 0 otherwise. Since she only chooses exactly one product, $\sum_j Y_{ij}(1) = \sum_j Y_{ij}(0) = 1$. We only observe the outcomes for the condition she is assigned to, i.e., $Y_{ij} = D_i\, Y_{ij}(1) + (1 - D_i)\, Y_{ij}(0)$.

Recall that $G = \{1, 3, 5\}$ is the set of Group I products and only Group I products show claims in the treatment condition. Then the average treatment effect (ATE) of displaying these claims, denoted by $\tau$, is the expected difference in choosing Group I products between

treatment and control, i.e.,

$$\tau \;=\; E\left[\sum_{j\in G} Y_{ij}(1) - \sum_{j\in G} Y_{ij}(0)\right] \;. \tag{1}$$

Since participants are randomly assigned to the control and treatment conditions, $D_i$ is independent of potential outcomes. Therefore, $E[Y_{ij} \mid D_i = 1] = E[Y_{ij}(1)]$ and $E[Y_{ij} \mid D_i = 0] = E[Y_{ij}(0)]$ for every product $j$, so

$$\tau \;=\; E\left[\sum_{j\in G} Y_{ij} \mid D_i = 1\right] - E\left[\sum_{j\in G} Y_{ij} \mid D_i = 0\right] \;, \tag{2}$$

i.e., $\tau$ equals the difference in the probability of choosing a Group I product between the treatment and control conditions. Its sample analogue is the difference between the two conditions in the share of participants who chose a Group I product.

#### 4.1.5 Data Collection and Data Cleaning

The browser extension records each action a participant takes while navigating the retailer's website. For each shopping session we observe: (i) the product the participant chose;[13] (ii) the participant-specific rank of each product on the list page; (iii) whether a product actually appeared on the participant's screen as they scrolled; and (iv) granular engagement statistics for each information block on the product detail page, including the clinical results claim image, the claim text, the ratings-and-reviews section, and the ingredients section. For each block, the extension records both a binary exposure indicator and the exact dwell time.

Of the 4,436 recruited participants, 36 failed attention checks and 1,076 did not complete either the shopping task or the survey; we exclude another 252 participants whose browsing or purchase data could not be recorded due to technical issues. These filters leave 3,049 participants. We further drop 22 participants whose number of product detail page visits

[13] Participants could choose a back-up product in case their first choice was unavailable at checkout. Participants were informed of this back-up option only after they had selected their first product.

exceeds 15—the 99th percentile of the distribution—and 683 participants who spent less than 30 seconds on the shopping task.[14] The final analysis sample contains 2,366 participants.[15]

## 4.2 Data Description

As expected and documented by previous literature (e.g., Ursu 2018), the probability that a consumer searches and/or chooses a product declines with the product's ranking, i.e., the further down a product is on the product list page, the less likely it is that a consumer searches or purchases the product. For example, the probability that an individual searches (purchases) the product ranked first is 4.65% (4.14%), while the probability that an individual searches (purchases) the product ranked 10th is 2.55% (2.54%).

Figure 5 displays the distributions of the number of searches individuals in the control and treatment conditions make. On average, consumers in the control condition search significantly more products (3.55 vs. 3.36, difference significant at the 5% level).

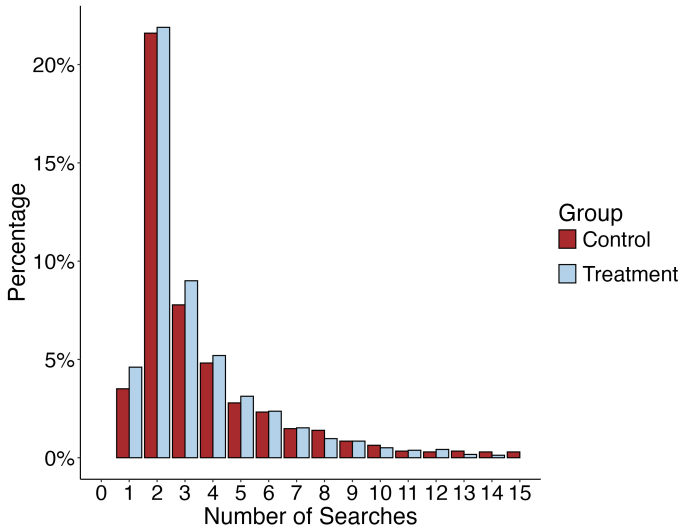



**Figure 5:** Histogram of the Number of Searches

Next, we show evidence that a significant proportion of participants looked at clinical results claims and that individuals were more likely to buy products with clinical results

[14] The mean and median shopping-task durations are 240 and 96 seconds, respectively.

[15] Web Appendix B provides a detailed attrition table.

claims. Our data shows that 266 participants (22% of participants in the treatment group) looked at clinical results claims.[16] Participants in the treatment group—who may see a clinical results claim—were also more likely to choose a Group I product (with clinical results claims) compared to those in the control group. Comparing product choices across control and treatment, 50.4% of participants chose a Group I product in the control group, while 53.6% of participants chose a Group I product in the treatment group (see Figure 6).

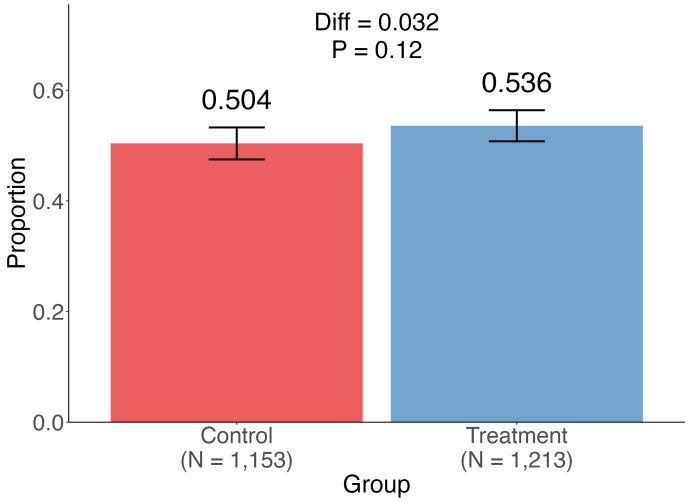



**Figure 6:** Proportion of Participants Buying a Group I Product

## 4.3 Causal Effects

We estimate linear probability models to measure the intention-to-treat (ITT) effects of the treatment assignment on three outcome measures: whether participants made a second search,[17] the number of searches they made, and whether they chose a Group I product. Recall that a Group I product is a product that shows clinical results claims in the treatment but not in the control condition. Table 5 displays the results.

[16] For comparison, 48% of participants looked at reviews, 12% of participants looked at the ingredients, 11% of participants looked at the "show more" section, 4% of participants looked at the "how to use" section, and 0.3% of participants looked at the brand description. There are no significant differences in the proportions of participants in treatment and control looking at the before mentioned five parts of the product detail pages.

[17] All participants had to click on at least one product detail page, i.e., make at least one search.

**Table 5:** Intention-to-Treat Effects on Search and Purchase Outcomes

| | (i) Second Search | (ii) Number of Searches | (iii) Purchase |
|---|---|---|---|
| Treatment | -0.017 | -0.205* | 0.040* |
| | (0.011) | (0.101) | (0.019) |
| Constant | 0.297 | -5.885* | 0.658† |
| | (0.194) | (2.593) | (0.337) |
| Demographic Fixed Effects | Yes | Yes | Yes |
| Product Attribute Controls | Yes | Yes | Yes |
| Product Ranking Controls | Yes | Yes | Yes |
| Number of Observations | 2,366 | 2,366 | 2,366 |
| $R^2$ | 0.045 | 0.052 | 0.157 |

Robust standard errors in parentheses.
We control for average product attributes and average product ranking of searched products in column (ii).
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

Column (i) presents the effect of treatment assignment on whether a participant decided to make a second search. Participants in the treatment condition are less likely to make a second search but the effect is not significant. Column (ii) displays the effect of treatment assignment on the number of searches a person made. The coefficient is negative and statistically significant, i.e., consistent with the descriptive results, individuals assigned to treatment make fewer searches than individuals in the control condition. Column (iii) presents the effect of being assigned to treatment on the likelihood of choosing a Group I product: the effect of treatment assignment is positive and statistically significant. Overall, the ITT results show that participants assigned to the treatment condition search less and that they are more likely to purchase a product with clinical results claims.

Table 6 reports instrumental variables estimates of the local average treatment effect (LATE) of looking at the clinical result claims on the same three outcome measures as before.[18] The LATE estimates are directionally consistent with the ITT estimates, but (as expected) larger in magnitude. They provide causal evidence that viewing a pseudo-scientific claim makes consumers search less (decreases the number of searches by about

[18]The $F$-statistics in all three 1st stage regressions are larger than 138.

0.93) and increases the likelihood that consumers buy a product with clinical results claims by about 36 percentage points.

**Table 6:** Local Average Treatment Effect on Search and Purchase Outcomes

| | (i) Second Search | (ii) Number of Searches | (iii) Purchase |
|---|---|---|---|
| Dummy for Seeing Clinical Results | -0.149 | -0.934* | 0.364* |
| | (0.103) | (0.473) | (0.171) |
| Constant | 0.340† | -5.772* | 0.630† |
| | (0.198) | (2.672) | (0.330) |
| Demographic Fixed Effects | Yes | Yes | Yes |
| Product Attribute Controls | Yes | Yes | Yes |
| Product Ranking Controls | Yes | Yes | Yes |
| Number of Observations | 2,366 | 2,366 | 2,366 |

Robust standard errors in parentheses.
We control for average product attributes and average product ranking of searched products in column (ii).
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

# 5 Mechanisms

In this section, we investigate the mechanisms through which clinical results claims affect consumer decision-making. More specifically, we aim to disentangle whether these claims serve as genuine quality signals or mislead consumers through their phrasing and framing, even when no credible quality message is present. Drawing on prior work in economics, marketing, and psychology, we focus on three mechanisms that have been identified as leading candidates for how scientific-sounding claims persuade consumers: costly signaling, credibility transfer from medical research, and numerical misinterpretation.

The first potential mechanism is quality signaling through costly firm actions (Nelson 1974). A large literature in marketing and economics has argued that costly actions such as warranties, advertising, or visible product testing can serve as credible signals of quality, because low-quality firms cannot profitably imitate them (Grossman 1981, Milgrom and Roberts 1986, Kirmani and Wright 1997). Consumers may recognize that conducting a

"clinical study" is costly, and thus infer higher product quality from the mere existence of such a study, regardless of its content or rigor.

The second potential mechanism is credibility transfer from medical research. Consumers encounter the phrases "clinical study" and "clinically tested" primarily in the context of medical trials, where these terms denote rigorous and tightly regulated scientific procedures. When the same vocabulary appears on a beauty product, consumers may import the credibility of medical research onto the beauty claim—what cognitive psychology refers to as schema-based processing (Alba and Hasher 1983, Brewer and Treyens 1981). Under this mechanism, the vocabulary of clinical testing can make efficacy claims appear more credible even when the advertisement makes no explicit reference to a medical trial and when the underlying study would not meet medical-trial standards.[19]

The third potential mechanism is that consumers treat the numerical component of a clinical results claim as evidence of effect size rather than of prevalence. A claim stating "93% see improvement," for example, reports that 93% of study participants observed some improvement in a particular attribute. Consumers, however, may interpret this either as a literal statement that participants experienced a 93% improvement, or as an inference that a higher level of consensus among participants implies a larger underlying effect. Under either interpretation, consumers conclude that the product is more effective than the evidence warrants. A substantial literature documents systematic biases in consumers' interpretation of probabilistic and quantitative information (Kahneman et al. 1982, Peters et al. 2006, Overton et al. 2025).

These three mechanisms generate different predictions about which components of a clinical results claim matter, and they carry different implications for regulation. If quality signaling is the operative mechanism, pseudo-scientific claims convey meaningful information about firm commitment and raise limited regulatory concern. If credibility transfer drives the effect, disclaimers clarifying what "clinical study" means in the beauty context may mitigate

[19] https://www.nytimes.com/2023/07/26/well/live/wellness-products-false-claims.html; https://www.beautyindependent.com/what-science-washing-can-the-beauty-industry-do-anything/

consumer confusion. If numerical misinterpretation is the primary driver, disclosure alone is unlikely to suffice, and broader educational interventions may be required. In the following subsection, we design an experiment that systematically varies the components of clinical results claims—the study-type label, the accompanying efficacy percentages, and a clarifying disclaimer—to isolate which of these mechanisms is at work.

## 5.1 Experimental Design

To empirically test for the presence of these mechanisms, participants were randomly assigned to one of the following experimental conditions and asked to choose one of two fictitious skincare products.

- **Benchmark Condition:** Both products include verbal efficacy claims without any reference to a clinical study (see Figure 7(a)).
- **Condition I (Clinical Study):** One product states that its efficacy claim is based on a "clinical study," while the other remains identical to the benchmark condition (see Figure 7(b)).
- **Condition II (Clinical Study with Percentage):** One product mentions a clinical study and provides claim percentages (e.g., "82% agreed it reduced wrinkles"), while the other remains identical to the benchmark condition (see Figure 7(c)).
- **Condition III (Clinical Study with Disclaimer):** Participants first read a clarification of what "clinical study" means in skincare, and then see the same product choice as in Condition II (see Figure 7(d)).




(a) Benchmark

(b) Condition I (Clinical Study)

**What is a "clinical study" in skincare?**
In skincare, "clinical study" has no formal legal definition and does not follow strict medical trial standards. It often means the product was tested on people, but the methods may not be scientifically rigorous.




(c) Condition II (Clinical Study with Percentage)

(d) Condition III (Clinical Study with Disclaimer)

**Figure 7:** Conditions for Mechanism Testing

Comparing the outcomes in the Benchmark and Condition I allows us to determine whether the mere mention of a clinical study affects consumer decision-making. If significantly more consumers choose the treated product in Condition I than in the Benchmark, this provides supporting evidence for the first and second mechanisms (quality signaling and

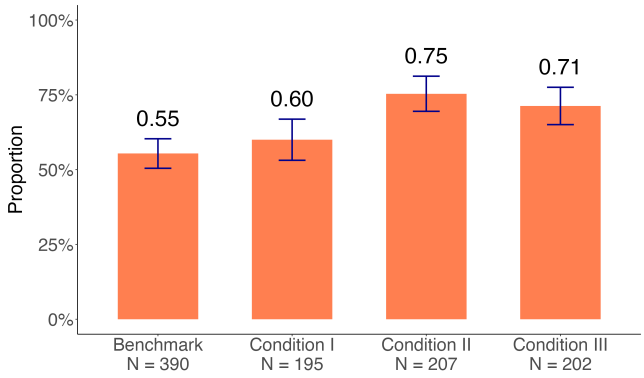
cognitive associations). Subsequent conditions allow us to differentiate between these two mechanisms. Conversely, if there is no significant increase in the proportion of consumers choosing the treated product between Condition I and the Benchmark, we can rule out the first two mechanisms as the driving forces behind the causal effect of clinical results claims.

Comparing Conditions I and II provides evidence on the presence or absence of the third mechanism (numerical information). If significantly more consumers choose the treated product in Condition II than in Condition I, that suggests the presence of the third mechanism; if no such increase is observed, we can rule it out.

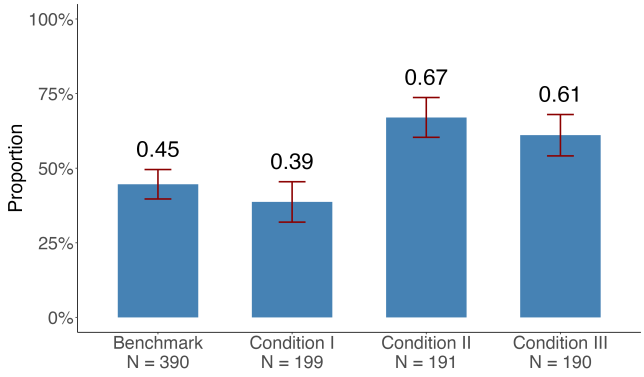
Finally, comparing Conditions II and III allows us to investigate further whether the second mechanism (cognitive associations) is present. If reading a disclaimer in Condition III significantly reduces the choice probability of the treated product relative to Condition II, this is evidence for the presence of the second mechanism.

## 5.2 Results

Figure 8 displays the proportion of participants who chose the treated product across the four experimental conditions, with 95% confidence intervals. Because participants in the Benchmark condition showed a slight baseline preference for the product on the left (55%) over the product on the right (45%), we report results separately for cases in which the treated product was positioned on the left (Figure 8a) and cases in which it was positioned on the right (Figure 8b). The patterns are consistent across the two panels, and we therefore discuss them jointly.




(a) Treated Product on the Left




(b) Treated Product on the Right

**Figure 8:** Proportion of Participants Choosing the Treated Product Across Experimental Conditions

Comparing the Benchmark and Condition I, we find no statistically significant difference in the share of participants choosing the treated product (55% vs. 60% when the treated product is on the left; 45% vs. 39% when it is on the right). Adding the phrase "clinical study" to the efficacy claim, without any accompanying numerical information, does not shift consumer choice. This null result is inconsistent with mechanisms 1 and 2: if consumers inferred higher quality from the mere existence of a study, or imported the credibility of medical research from the word "clinical," we would expect choice probabilities in Condition I to rise relative to the Benchmark.

Adding numerical efficacy percentages, however, produces a large and statistically significant shift in choice. In Condition II, the treated product is chosen 75% of the time when on the left (up from 55% in the Benchmark) and 67% of the time when on the right (up from

45%). This roughly 20-percentage-point increase relative to the Benchmark – and the 15- to 28-percentage-point increase relative to Condition I – provides direct evidence for mechanism 3 (numerical misinterpretation). What changes consumer choice is not the reference to a clinical study but the inclusion of percentages such as "82% agreed it reduced wrinkles." Condition III, which retains the numerical percentages but adds a disclaimer clarifying that "clinical study" in skincare does not follow strict medical-trial standards, produces choice shares that are statistically indistinguishable from Condition II (71% on the left and 61% on the right, compared with 75% and 67% in Condition II). The disclaimer about what "clinical study" means in skincare therefore does not attenuate the effect of the claim on consumer choice. This reinforces the conclusion that credibility transfer (mechanism 2) is not what drives the effect: if consumers were responding to the word "clinical" because of its medical-trial connotations, clarifying that the word does not carry those connotations in skincare should reduce the treatment effect, yet it does not.

Taken together, the results support only one of the three mechanisms we considered. Consumers respond to the presence of numerical efficacy results, not to the language of clinical testing or to its implied costliness. Stating that a claim is "based on a clinical study" alone does not affect consumer choice, and clarifying that such studies fail to meet medical-trial standards does not attenuate the effect either. The driver is the inclusion of numeric percentages, e.g., the share of study participants reporting an improvement.

## 5.3 Follow-up Experiment

The results in Section 5.2 identify numerical information as the component of clinical results claims that shifts consumer choice. To understand which features of that numerical information matter, we conducted a follow-up experiment in which respondents rated the perceived efficacy implied by different claims. Respondents were shown a set of efficacy statements (e.g., "93% showed diminished appearance of fine lines") accompanied by a footnote specifying study details (e.g., "Based on a clinical study with 202 participants after 1 month of

regular use") and were asked, on a seven-point Likert scale, how certain they were that the product effectively achieved the stated functions.

Each participant evaluated three claims: a benchmark claim, identical across participants, and two randomly selected variants. Each variant holds all elements of the benchmark fixed except one – the study-type label, the sample size, or the reported efficacy percentages – allowing us to isolate the marginal contribution of each element. Table 7 summarizes the five claims.

| **Condition** | **Study Type** | **Sample Size** | **Duration** | **Percentage** |
|---|---|---|---|---|
| Benchmark | Clinical study | 31 participants | 8 weeks | 100%, 96%, 93% |
| Clinical study | Clinical study | 202 participants | 1 month | 93%, 86%, 82% |
| Consumer perception | Consumer-perception study | 202 participants | 1 month | 93%, 86%, 82% |
| Small sample | Clinical study | 35 participants | 1 month | 93%, 86%, 82% |
| High agreement | Clinical study | 202 participants | 1 month | 100%, 98%, 98% |

**Table 7:** Experimental Conditions for Clinical Results Claims

Figure 9 plots the average *standardized* rating for each claim, defined as the respondent's rating for that claim minus their own rating for the benchmark. This within-respondent differencing removes individual variation in Likert-scale usage and isolates the effect of each manipulated feature.

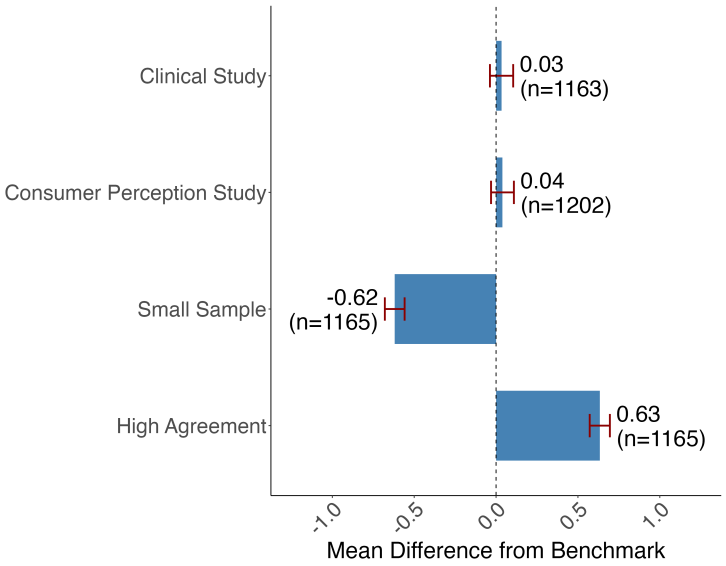



**Figure 9:** Distribution of Perceived Credibility of Different Product Claims

The pattern of ratings mirrors the choice-behavior results in Section 5.2 and sharpens the mechanism interpretation. First, ratings of the clinical-study and consumer-perception-study claims are statistically indistinguishable (0.03 vs. 0.04). Relabeling the underlying study from a "clinical study," which consumers may associate with medical-trial rigor, to a "consumer-perception study," which in industry usage indicates a less rigorous self-assessment, does not move perceived efficacy. This null result reinforces the conclusion from Section 5.2 that credibility transfer (mechanism 2) is not the operative channel. Second, reducing the reported sample size from 202 to 35 participants lowers the rating substantially (from 0.03 to -0.62). Third, raising the reported efficacy percentages from 93%/86%/82% to 100%/98%/98% raises the rating by a comparable magnitude (from 0.03 to 0.63). Respondents thus treat larger sample sizes and higher agreement percentages as stronger evidence of product effectiveness, but do not treat the label "clinical" as more informative than "consumer perception."

These ratings refine the interpretation of mechanism 3. Consumers do not simply misread "93% showed improvement" as a "93% improvement"; under such a literal misreading, varying the study-type label or the sample size should not affect perceived efficacy, since both

manipulations leave the percentages unchanged. Yet the sample-size manipulation produces a large shift in ratings, indicating that consumers are interpreting the percentages in light of the underlying study's informativeness. The pattern is most consistent with consumers inferring a positive relationship between the reported agreement rate and the underlying probability that the product works, with larger sample sizes strengthening that inference. This interpretation aligns with recent work showing that consumers conflate consensus-based metrics with causal evidence (Overton et al., 2025), and with a broader literature documenting that people overweight numerosity and fail to integrate denominators or units properly (Denes-Raj et al., 1995, Pandelaere et al., 2011, Peters et al., 2006).

Taken together, the choice and belief measures converge on a single mechanism: consumers respond to the numeric content of clinical results claims rather than to the existence or type of the underlying study. The word "clinical" neither costly-signals nor schema-transfers credibility in this setting; instead, consumers read the reported percentages as evidence of efficacy, interpret higher percentages as evidence of a larger effect, and treat larger study samples as sharpening that evidence.

# 6 Implications for Regulation

Our findings have direct implications for how pseudo-scientific claims should be regulated. Because consumer response is driven by the numerical content of the claim – the reported percentages and sample size – rather than by the presence or type of the underlying study, regulation targeting the language of clinical testing is unlikely to reduce consumer susceptibility. In our experiments, a disclaimer explicitly stating that a "clinical study" in skincare need not meet medical-trial standards had no attenuating effect on consumer choice. The locus of the problem lies in how consumers interpret percentages and sample sizes, not in whether they understand the term "clinical."

The current FTC framework already contains several relevant elements. The FTC re-

quires that advertisers possess at least the level of clinical evidence the wording of their claims implies, and for health benefits this generally means well-designed, randomized, controlled human clinical testing.[20] Testimonials and consumer-perception statements do not convert anecdotes into evidence; advertisers are liable for the expressed and implied efficacy conveyed through endorsements and must hold adequate substantiation as if they had made the claim directly.[21] In practice, however, these rules are primarily enforced *ex post* through challenges to specific claims, and the beauty category, where most claims concern cosmetic rather than health benefits, receives limited scrutiny. The pattern we document, in which studies are routinely described with minimal methodological detail and accompanied by percentages that appear precise but lack substantiation, operates largely below the threshold of active enforcement.

Our results argue for a regulatory shift toward the numerical content of claims rather than toward the language surrounding them. Two specific directions are suggested by our findings. First, because consumers interpret higher agreement rates as evidence of larger underlying effects, regulators could require that reported percentages be accompanied by a disclosure of the typical *magnitude* of the effect observed among those who reported improvement. A claim that "93% of participants reported an improvement" is consistent with improvements ranging from barely perceptible to dramatic, and disclosing the magnitude would allow consumers to distinguish these cases. Second, because consumers treat larger sample sizes as evidence of stronger effects, regulators could require that numerical efficacy claims be based on studies meeting specified design standards – adequate sample size, controls, and randomization – preventing firms from generating credible-looking numerical evidence from small, uncontrolled self-assessments.

Our disclaimer results also indicate that informational interventions alone are unlikely to suffice. A disclaimer clarifying what "clinical study" means in skincare did not attenuate consumer response in our setting, consistent with the broader literature showing that

[20] `https://www.ftc.gov/business-guidance/resources/health-products-compliance-guidance`

[21] FTC Endorsement Guides, 16 C.F.R. §255.2(a). See also FTC Health Products Compliance Guidance (Dec. 2022).

the effectiveness of corrective messages depends heavily on how they are delivered and on audiences' prior beliefs (Chan et al. 2017, Lewandowsky et al. 2012). Improving consumer numeracy – particularly the interpretation of agreement rates as evidence of efficacy – may therefore be a necessary complement to any disclosure-based remedy. Absent such complementary measures, our results imply that simple disclaimers will fall short of protecting consumers from the specific form of misinterpretation we identify. More regulation is warranted, but its leverage lies in the numerical content of claims and in consumers' ability to interpret that content, not in the language of clinical testing that surrounds it.

# 7 Conclusion

Pseudo-scientific claims are widespread on the shelves of major beauty retailers, yet remarkably little is known about how consumers actually respond to them. This paper closes that gap. Using an incentive-aligned field experiment on a leading online beauty retailer's website, we show that clinical results claims cause consumers to terminate their product search earlier and to choose products carrying such claims at higher rates. A follow-up mechanism experiment isolates the driver of this effect: consumers do not respond to the mere existence of a "clinical study," nor to the scientific-sounding language that surrounds it, nor to schema-based associations with medical research. What moves choice is the numerical content of the claim – the percentage of study participants who reported an improvement and the size of the study – which consumers read as direct evidence of product effectiveness.

This pattern is inconsistent with interpretations under which pseudo-scientific claims are benign. If consumers inferred quality from the costly action of conducting a study, disclosure of that fact alone should suffice to shift choice; it does not. If consumers imported the credibility of medical trials through the word "clinical," clarifying that the word does not carry its medical meaning in skincare should attenuate the effect; again, it does not. Instead, consumers appear to treat numerical agreement rates as informative about effect

size, and treat larger study samples as strengthening the inference – behavior consistent with a substantial literature on biases in the interpretation of probabilistic and quantitative information.

The welfare consequences depend on which firms use such claims. If only higher-quality firms conduct and disclose studies, consumer reliance on them need not generate significant welfare loss, since the heuristic "this product is backed by a study" may correlate with genuine quality differences in equilibrium. If lower-quality firms also make such claims, which our descriptive evidence in Section 3 suggests is the case, the market becomes susceptible to systematic deception: consumers misallocate demand toward inferior products, and firms' incentives to invest in genuine quality improvements erode. Even when the underlying study is internally valid, the reported percentage may reflect non-causal factors such as placebo responses, self-selection, or seasonal and environmental confounders. In the absence of transparency about study design and causal identification, the reported numbers carry little evidentiary weight; yet consumers treat them as if they did.

Several broader contributions follow. Empirically, we provide the first causal evidence on how consumers respond to pseudo-scientific claims in the absence of regulatory intervention, separating the effect of the claim itself from the combined effect of claim removal and negative publicity that dominates the existing literature (Rao and Wang 2017*b*, Rao 2022*b*). Methodologically, we demonstrate a design combining a browser-extension field experiment with follow-up mechanism experiments that may be useful in other settings in which claims, labels, or certifications are difficult to vary exogenously at scale. Substantively, the findings contribute to the literature on firms' strategic provision of information: firms need not lie, and need not attach false content to credible-sounding labels, to mislead consumers; reporting *selective* numerical information from weakly designed studies is sufficient.

The findings also inform regulatory debate. As we discuss in Section 6, our results imply that interventions targeting the language of clinical testing are unlikely to protect consumers. The locus of the problem is the numerical content of the claim, and effective intervention will

require either (i) design standards for the studies whose results may be cited in advertising, (ii) disclosure of the magnitude, not just the prevalence, of reported effects, or (iii) broader efforts to improve consumer numeracy. The relative costs of these interventions, and their distributional consequences for firms and consumers, are an important open question.

Several limitations warrant acknowledgment. Our field experiment focuses on a single product category, anti-aging moisturizers, at a single retailer, and our sample is restricted to women in the United States who use facial creams more than twice a week. Whether the mechanism we document extends to other categories, other consumer segments, and other retail environments is an empirical question that we leave to future work. Our mechanism experiments use fictitious products in a choice setting that, while incentive-aligned, is not identical to a real shopping environment; the follow-up credibility ratings rely on self-reports rather than revealed behavior. Finally, we do not estimate long-run consequences: whether consumers learn to discount pseudo-scientific claims after repeated exposure, how firms' use of such claims responds to category-level disclosure or enforcement, and how product quality itself responds to the regulatory interventions we propose are all beyond the scope of the present paper.

These limitations, in turn, suggest natural directions for future research. On the consumer side, characterizing heterogeneity in susceptibility – by age, education, numeracy, or prior experience with scientific claims – would sharpen the targeting of educational interventions. On the firm side, observing how firms adjust the content and framing of their claims in response to shifts in consumer behavior or regulatory policy would shed light on the equilibrium implications of the patterns we document. On the regulatory side, a direct comparison of the alternative interventions we propose – design-standard requirements, magnitude disclosures, and numeracy-oriented consumer education – would allow the field to move from our qualitative recommendations to quantitative guidance on which instruments work best. We hope that the findings reported here provide a foundation for that subsequent work.

# References


Alba, Joseph W. and Lynn Hasher (1983), 'Is memory schematic?', *Psychological Bulletin* **93**(2), 203–231.

Brewer, William F. and James C. Treyens (1981), 'Role of schemata in memory for places', *Cognitive Psychology* **13**(2), 207–230.

Chan, Man-pui Sally, Christopher R. Jones, Kathleen Hall Jamieson and Dolores Albarracín (2017), 'Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation', *Psychological Science* **28**(11), 1531–1546.

Compiani, Giovanni, Greg Lewis, Sida Peng and Will Wang (2024), 'Online search and optimal product rankings: An empirical framework', *Marketing Science* **43**(3), 615–636.

Denes-Raj, Veronika, Seymour Epstein and Joan Cole (1995), 'The generality of the ratio-bias phenomenon', *Personality and Social Psychology Bulletin* **21**(10), 1083–1092.

Donnelly, Robert, Ayush Kanodia and Ilya Morozov (2024), 'Welfare effects of personalized rankings', *Marketing Science* **43**(1), 92–113.

Dranove, David and Ginger Zhe Jin (2010), 'Quality disclosure and certification: Theory and practice', *Journal of Economic Literature* **48**(4), 935–963.

Federal Trade Commission (2024), 'Ftc policy statement regarding advertising substantiation'.
**URL:** *https://www.ftc.gov/legal-library/browse/ftc-policy-statement-regarding-advertising-substantiation*

Fong, Jessica, Tong Guo and Anita Rao (2024), 'Debunking misinformation about consumer products: Effects on beliefs and purchase behavior', *Journal of Marketing Research* **61**(4), 659–81.

Fowler, Jie G., Timothy H. Reisenwitz and Les Carlson (2015), 'Deception in cosmetics advertising: Examining cosmetics advertising claims in fashion magazine ads', *Journal of Global Fashion Marketing* **6**(3), 194–206.

Greminger, Rafael (2023), Heterogeneous position effects and the power of rankings. Working Paper.

Grossman, Sanford J. (1981), 'The informational role of warranties and private disclosure about product quality', *Journal of Law and Economics* **24**(3), 461–483.

Heneghan, Carl, Jeremy Howick, Braden O'Neill, Peter J. Gill, Daniel S. Lasserson, Deborah Cohen, Ruth Davis, Alison Ward, Adam Smith, Greg Jones et al. (2012), 'The evidence underpinning sports performance products: a systematic assessment', *BMJ Open* **2**(4).

Ippolito, Pauline M. and Alan D. Mathios (1990), 'Information, advertising and health choices: A study of the cereal market', *RAND Journal of Economics* **21**(3), 459–480.

Jin, Ginger Zhe and Phillip Leslie (2003), 'The effect of information on product quality: Evidence from restaurant hygiene grade cards', *Quarterly Journal of Economics* **118**(2), 409–451.

Jovanovic, Boyan (1982), 'Truthful disclosure of information', *Bell Journal of Economics* **13**(1), 36–44.

Kahneman, Daniel, Paul Slovic and Amos Tversky (1982), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Kirmani, Amna and Peter Wright (1997), 'The effect of perceived advertising costs on brand perceptions', *Journal of Consumer Research* **24**(1), 54–63.

Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz and John Cook (2012), 'Misinformation and its correction: Continued influence and successful debiasing', *Psychological Science in the Public Interest* **13**(3), 106–131.

Mathios, Alan D. (2000), 'The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market', *Journal of Law and Economics* **43**(2), 651–678.

Milgrom, Paul and John Roberts (1986), 'Price and advertising signals of product quality', *Journal of Political Economy* **94**(4), 796–821.

Milgrom, Paul R. (1981), 'Good news and bad news: Representation theorems and applications', *Bell Journal of Economics* **12**(2), 380–391.

Nelson, Phillip (1974), 'Advertising as information', *Journal of Political Economy* **82**(4), 729–754.

Nyhan, Brendan and Jason Reifler (2010), 'When corrections fail: The persistence of political misperceptions', *Political Behavior* **32**(2), 303–330.

Overton, Gemma, Ioannis Evangelidis and Jens Vosgerau (2025), 'People believe if 90% prefer a over b, a must be much better than b. are they wrong?', *Journal of Consumer Research* **52**(1), 135–156.

Pandelaere, Mario, Barbara Briers and Christophe Lembregts (2011), 'How to make a 29% increase look bigger: The unit effect in option comparisons', *Journal of Consumer Research* **38**(2), 308–322.

Peake, Jonathan M., Graham Kerr and John P. Sullivan (2018), 'A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations', *Frontiers in Physiology* **9**, 743.

Peters, Ellen, Daniel Västfjäll, Paul Slovic, Christopher K. Mertz, Ketti Mazzocco and Stephan Dickert (2006), 'Numeracy and decision making', *Psychological Science* **17**(5), 407–413.

Rao, Anita (2022*a*), 'Deceptive claims using fake news advertising: The impact on consumers', *Journal of Marketing Research* **59**(3), 534–554.

Rao, Anita (2022*b*), 'Industry-funded research and bias in food science', *Quantitative Marketing and Economics* **20**(1), 39–67.

Rao, Anita and Emily Wang (2017*a*), 'Demand for "healthy" products: False claims and FTC regulation', *Journal of Marketing Research* **54**(6), 968–989.

Rao, Anita and Emily Wang (2017*b*), 'Demand for "healthy" products: False claims and FTC regulation', *Journal of Marketing Research* **54**(6), 968–989.

Shi, Zijun, Xiao Liu and Kannan Srinivasan (2022), 'Hype news diffusion and risk of misinformation: The Oz effect in health care', *Journal of Marketing Research* **59**(2), 327–352.

Statista (2025), 'Top beauty online stores in the United States 2025'. Accessed: 2026-10-05.
**URL:** *https://www.statista.com/statistics/1081154/top-online-stores-beauty-united-states-ecommercedb/*

Ursu, Raluca M. (2018), 'The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions', *Marketing Science* **37**(4), 530–552.

Yoo, Boonghee and Naveen Donthu (2001), 'Developing and validating a multidimensional consumer-based brand equity scale', *Journal of Business Research* **52**(1), 1–14.

Zinman, Jonathan and Eric Zitzewitz (2016), 'Wintertime for deceptive advertising?', *American Economic Journal: Applied Economics* **8**(1), 177–192.

# Web Appendix A: Brand Equity Survey

Following Yoo and Donthu (2001), we designed a survey to measure the three aspects of brand equity (brand awareness, brand loyalty, and perceived quality) for the 24 brands in our study.[22] Table A-1 displays the items for each brand equity construct. Participants rated their agreement with each statement on a 5-point Likert scale. To ensure data quality, we did not ask each participant to evaluate all 24 brands. Instead, each participant answered the items for three randomly assigned brands.

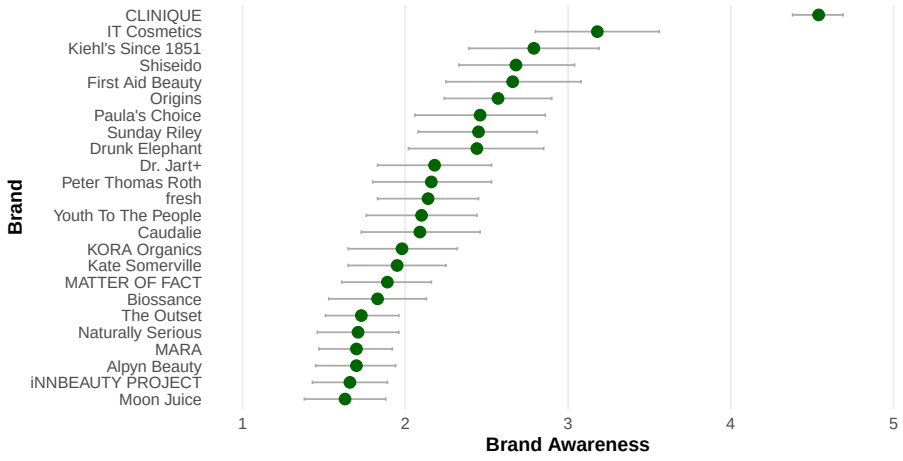
| **Construct** | **Item** |
|---|---|
| Brand Awareness | I can recognize X among other competing brands.<br>I am aware of X.<br>I can quickly recall the symbol or logo of X.<br>I have difficulty in imagining X in my mind. |
| Brand Loyalty | I consider myself to be loyal to X.<br>X would be my first choice.<br>I will not buy other brands if X is available at the store. |
| Perceived Quality | The likely quality of X is extremely high.<br>The likelihood that X would be functional is very high. |

*Note*: "X" represents the focal brand shown to each respondent.

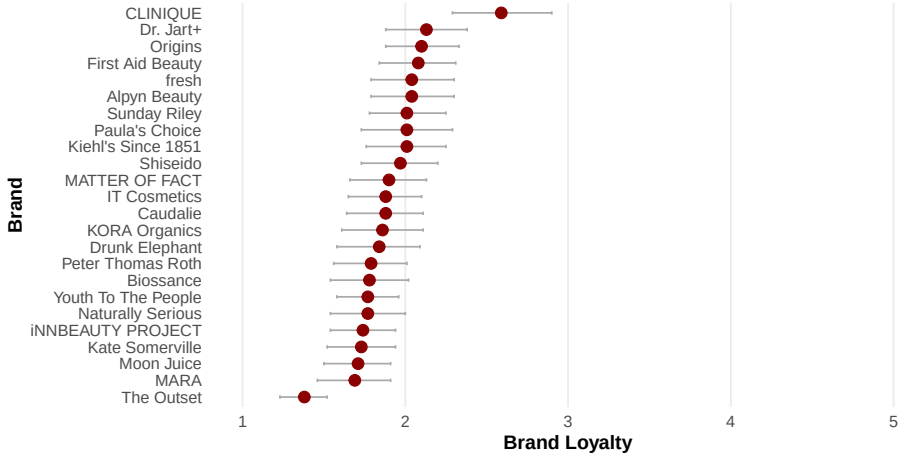
**Table A-1:** Survey Items for Brand Equity Constructs

We ran our survey on CloudResearch Connect on February 13, 2025. 413 individuals started the survey; our final sample contains 393 women living in the U.S. who passed the attention checks. For each brand equity measure, we computed the average rating across all corresponding items and all participants who evaluated that brand. Figure A-1 shows the values of the three brand equity constructs for all 24 brands.

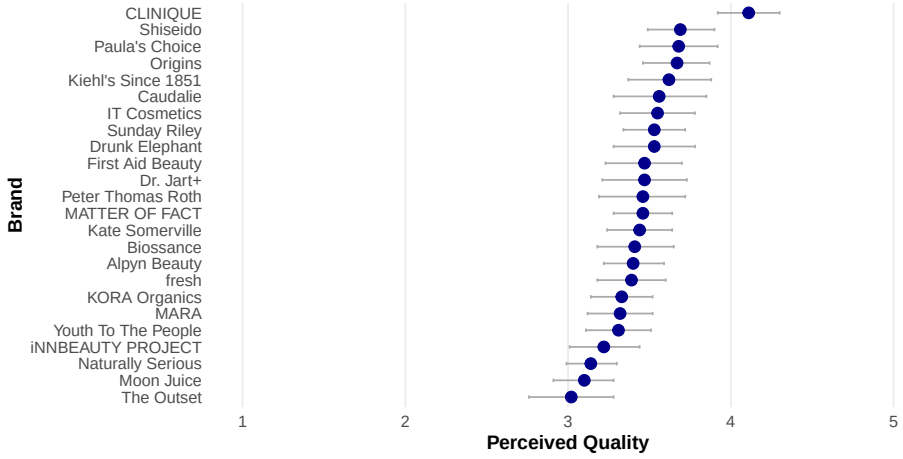
[22] The complete survey is available at `brand_equity_survey.pdf`.




(a) Brand Awareness Scores with 95% Confidence Intervals




(b) Brand Loyalty Scores with 95% Confidence Intervals




(c) Perceived Quality Scores with 95% Confidence Intervals

**Figure A-1:** Brand Equity Measures.

# Web Appendix B: Data Attrition

Table B-1 summarizes sample attrition. "Rows Remaining" reports the number of observations left after the filter described in that row was applied, while "Rows Discarded" reports how many observations the filter removed compared to the previous line. For example, 6,906 respondents started the survey. Of these, 350 did not consent, leaving 6,556 respondents who consented. Among those who consented, 681 failed the technology screener by selecting "No, I cannot install the extension or do not use Google Chrome," reducing the sample to 5,875. The subsequent rows follow the same logic.

**Table B-1:** Data Attrition

| **Filter Applied** | **Rows Remaining** | **Rows Discarded** |
|---|---|---|
| Initial sample | 6,906 | – |
| Do not consent | 6,556 | 350 |
| Tech screener failed[a] | 5,875 | 681 |
| Not located in the U.S. | 5,873 | 2 |
| Not female | 5,656 | 217 |
| Uses cream <3 times/week | 4,436 | 1,220 |
| Failed attention check | 4,400 | 36 |
| Did not finish shopping task | 3,329 | 1,071 |
| Did not finish survey | 3,323 | 6 |
| No product purchase data | 3,307 | 16 |
| Browsing record issues[b] | 3,288 | 19 |
| Purchase record issues[c] | 3,071 | 217 |
| Excessive searches (>$99^{th}$ pct.) | 3,049 | 22 |
| Browsing time <30 s | 2,366 | 683 |
| **Final sample size** | **2,366** | – |

[a] Users who answered "No, I cannot install the extension or do not use Google Chrome."
[b] 5 users selected a product outside the experiment; 14 users lacked any product detail page browsing record.
[c] 212 users repeated the shopping task; 5 users had incomplete purchase records.

# Web Appendix C: Claim Ratings

In the main experiment, respondents were shown a set of efficacy claims and asked:

> *Based on these results, how certain are you that this skincare product effectively achieves these functions?*

Responses were recorded on a seven-point Likert scale (1 = "not at all certain", 7 = "very certain"). Each participant rated three claims. The first (benchmark) claim was identical for everyone, after which each participant was randomly assigned to evaluate two of the remaining four claims.

- **Benchmark claim**
  100% showed improvement in firmness and elasticity*
  96% showed reduced lines and wrinkles*
  93% showed improvement in texture*
  **Based on a clinical study with 31 participants after 8 weeks.*

- **Clinical study claim**
  93% reported skin became very smooth*
  86% reported skin was full of vitality*
  82% reported diminished appearance of fine lines*
  **Based on a clinical study with 202 participants after 1 month of regular use.*

- **Consumer perception study claim**
  Identical to the clinical-study claim, but with the reference:
  **Based on a consumer perception study with 202 participants after 1 month of regular use.*

- **Small sample claim**
  Identical to the clinical-study claim, but with the reference:
  **Based on a clinical study with 35 participants after 1 month of regular use.*

- **High efficacy claim**
  Identical to the clinical-study claim, but the efficacy figures are 100%, 98%, and 98%.