Claim Substantiation: Common Issues In Study Design That Trip Advertisers Up

Jeff Warshafsky
Jennifer Yang

March 20, 2024

Proskauer

Speakers

Jeff Warshafsky
Partner
jwarshafsky@proskauer.com

Jennifer Yang
Senior Counsel
jyang@proskauer.com

Substantiation: An Overview

Sensory Testing

Clinical and Home Use Testing

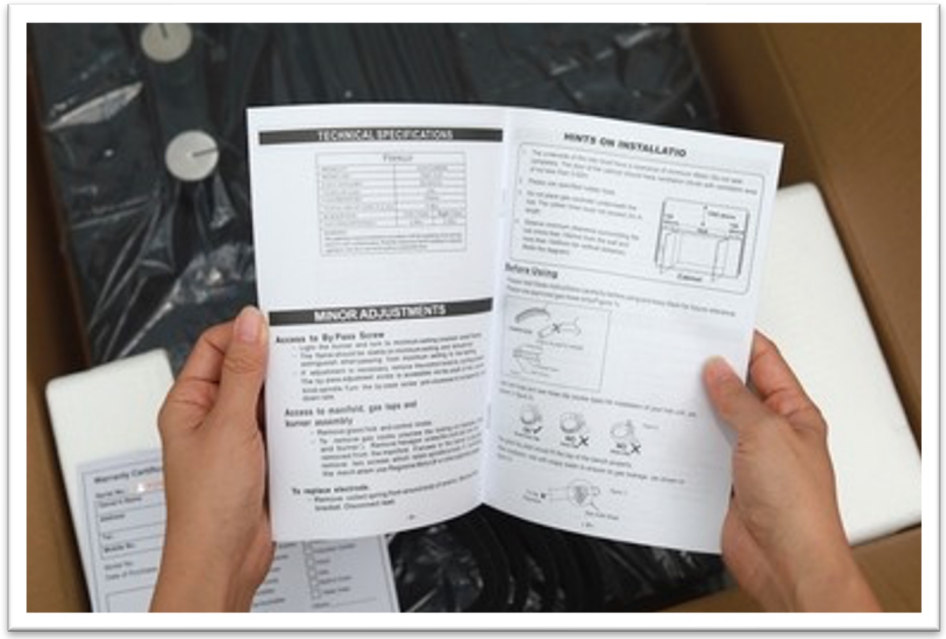
Analytical Testing

General Testing Principles

Which Products Do You Need to Test?

- The advertised product
- Any other products your ad references (even implicitly)
  - "It is well accepted that, when made in a comparative context, claims of superiority (or parity, or being 'unsurpassed') are best supported by head-to-head testing."
    - *Shell Oil Company (Shell V-Power NiTRO+ Premium Gasoline)*, Report #6065, NAD/CARU Case Reports (March 2017)
  - "It is well established by NAD precedent that whenever an advertiser makes a broad category-wide superiority claim, comparative testing against at least 85% of the relevant marketplace is needed to support the claim."
    - *SharkNinja Operating LLC (Shark Stratos™ Powered Lift-Away upright vacuum)*, Report #7151, NAD/CARU Case Reports (April 2023)
- An appropriate control product
- A representative sample of the relevant products

How Should You Test the Products?



- Replicate real-world conditions
  - "NAD determined that the conditions under which the advertiser's product demonstration experiment was conducted did not accurately reflect how body wash is used in real life."
    - *Unilever United States (Dove® Deep Moisture Body Wash)*, Report #5599, NAD/CARU Case Reports (June 2013)
- Follow product usage instructions
  - "A study which does not follow a product's use instructions in evaluating the product does not provide reliable evidence of the performance the product."
    - *3B Medical, Inc. (Lumin CPAP Cleaner)*, Report #6300, NAD/CARU Case Reports (August 2019)

How Should You Test the Products?

- Follow industry-standard test methods unless you have good reason not to
  - "While NAD agreed with the advertiser that it does not blindly adhere to industry standard testing, NAD will carefully scrutinize departures from industry standard testing where a particular industry standard test has long been established as the means for substantiating the claim at hand."
    - *SharkNinja Operating LLC (Shark Rotator Powered Lift-Away Speed Vacuum Cleaner)*, Report #6174, NAD/CARU Case Reports (April 2018)
- Avoid bias…

How Can You Minimize Bias?

Conduct testing independently

- "NAD noted that one of the tests on the Atoxelene Line Wand tested the product on employees of the company which has the potential to bias results, particularly where, as here, the test subjects were asked to assess the results themselves."
  - *Intraceuticals LLC (Atoxelene Skin Care Products)*, Report #5953, *NAD/CARU Case Reports* (May 2016)

Use a repeatable and reliable methodology

Randomly assign subjects to test groups

Avoid leading questions

Consider a control

- "[T]he study relies upon the survey results of subjective reactions of study participants, yet the study lacked a control group to serve as a basis for comparison making it difficult to know whether the changes in premenstrual symptoms perceived by the subjects were due to a placebo effect or due to the use of Pamprin Botanicals."
  - *Focus Consumer Healthcare (Pamprin Botanicals)*, Report #7247, *NAD/CARU Case Reports* (January 2024)

How Can You Minimize Bias?

Blind study participants (both subjects and investigators)

- "The Palsson study's results were further undermined by the study's lack of blinding, which introduces the potential for bias. Blinding is particularly important when the response criteria are subjective, such as the alleviation of pain or other sensory symptoms."
  - *i-Health, Inc. (DSM North America) (Culturelle® IBS Complete Support)*, Report #7080, *NAD/CARU Case Reports* (May 2022)

If you can't blind, have a good justification and implement other controls

- "NAD has previously acknowledged that some consumer tests can be difficult to blind. This is particularly true where the product has a unique appearance that allows it to be recognized by its shape and design. … In instances where products cannot reasonably be blinded, to help minimize the potential for bias, NAD has recommended that the study sample be balanced" and the advertiser should "take[] additional steps to limit brand references (e.g., covering branding on the vacuums, appliance storage bags and use manuals)."
  - *SharkNinja Operating LLC (Shark Rotator Powered Lift-Away)*, Report #6095, *NAD/CARU Case Reports* (July 2017)

What Does Your Testing Need to Prove?

- Statistically significant results
- Clinically meaningful results
  - "NAD was also concerned whether the results were not only statistically significant but also clinically meaningful. … The main objective should be to produce a clinically significant reduction in symptoms rather than a small but statistically significant reduction."
    - *The Procter & Gamble Company (Crest Sensitivity Treatment & Protection toothpaste)*, Report #5386, *NAD/CARU Case Reports* (September 2011)

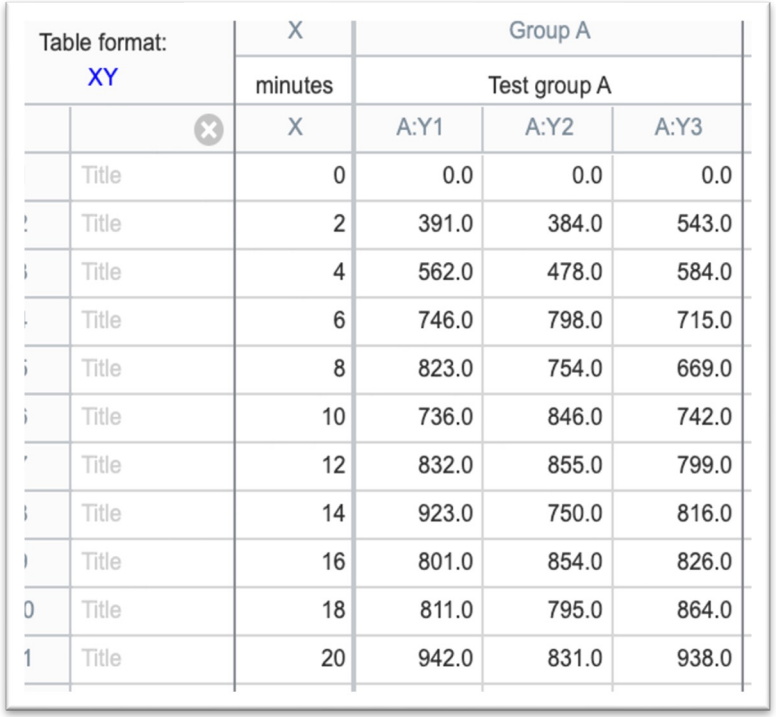


Table format: XY		X	Group A		
		minutes	Test group A		
		X	A:Y1	A:Y2	A:Y3
	Title	0	0.0	0.0	0.0
	Title	2	391.0	384.0	543.0
	Title	4	562.0	478.0	584.0
	Title	6	746.0	798.0	715.0
	Title	8	823.0	754.0	669.0
	Title	10	736.0	846.0	742.0
	Title	12	832.0	855.0	799.0
	Title	14	923.0	750.0	816.0
	Title	16	801.0	854.0	826.0
0	Title	18	811.0	795.0	864.0
1	Title	20	942.0	831.0	938.0

Ensure Your Testing is a "Good Fit" for Your Claim

Population should be representative of your target audience

- "When conducting testing to support an advertising claim, testing on a study population that is representative of the target population to which the claim is targeted is a hallmark of sound methodological design. … A study population does not have to be representative of the population of consumers to whom a claim is directed in every respect, but for study results to be reliable, the study population must be representative of the target population with respect to those characteristics most relevant to the claim."
  - *Oral Essentials, Inc. (Lumineux Whitening Strips)*, Report #7235, *NAD/CARU Case Reports* (October 2023)

Results should match your advertising claim

- "The study concluded that Fillerina 'is able to provide an improvement in the appearance of chronoaged skin in subjects showing mild-to-moderate clinical signs of skin aging.' This conclusion is far more tempered than the challenged claims promising dramatic and long-lasting improvements in wrinkles and sagging skin."
  - *Qf Systems, LLC (Fillerina Dermo-Cosmetic Replenishing Gel)*, Report #6373, *NAD/CARU Case Reports* (June

Ensure Your Testing is a "Good Fit" for Your Claim

Justify any "bridging" or "extrapolation"

- "In reviewing the advertiser's evidence, NAD was concerned that the advertiser relied on testing for its Double Roll product, even though the claims in this category were for its Mega Roll product. … Despite the advertiser's argument that any difference between the Double Roll and Mega Roll product is immaterial, the challenger's evidence suggests that the Double and Mega Roll products in fact differ in their physical characteristics."
  - *The Procter & Gamble Company (Charmin Ultra Strong and Charmin Ultra Soft Products)*, Report #5960, *NAD/CARU Case Reports* (May 2016)
- "NAD recognizes that there may be instances when health performance claims can be substantiated without clinical studies on the specific product advertised. In such cases, however, an advertiser must provide reliable evidence demonstrating that it is scientifically sound to extrapolate the conclusions drawn from other studies and data and apply them to the performance claimed for the advertised product."
  - *Nootrobox, Inc. (Advertising for Nootropics)*, Report #5995, *NAD/CARU Case Reports* (August 2016)

Testing Principles Applied to Health Claims

Competent & Reliable Scientific Evidence

"Health-related claims generally should be supported by competent and reliable scientific evidence, as defined by the FTC, and includes, 'tests, analyses, research, or studies that (1) have been conducted and evaluated in an objective manner by experts in the relevant disease, condition, or function to which the representation relates; and (2) are generally accepted in the profession to yield accurate and reliable results.'"

Fitness Cubed, Inc. (Cubii Seated Elliptical Trainer), Report #7145, NAD/CARU Case Reports (February 2020).

Substantiating Health Claims

- Randomized, controlled human clinical trials
- Double-blinded
- Control group vs. treatment group
- Independently-conducted replicated study
- Designed to yield clinically meaningful and statistically significant results

Focus Consumer Healthcare (Pamprin Botanicals), NAD Case #7247 (Jan. 2024)

"Clinically Tested" and Other Health-Related Claims

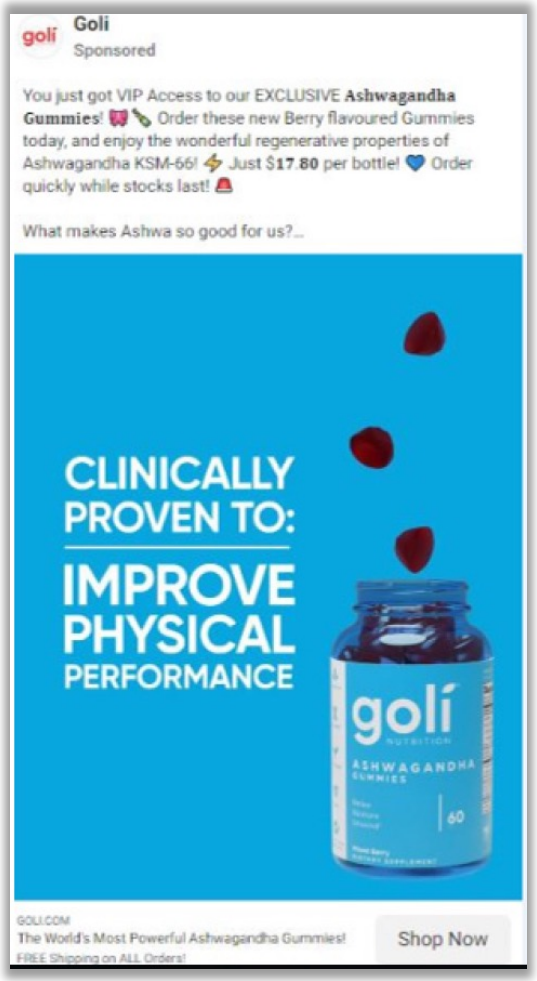
- Advertiser's Study
  - Phase 1:
    - Participants filled out baseline survey about severity of common menstrual cycle symptoms and had blood drawn.
    - Took Pamprin Botanicals before and during period.
    - Completed survey on fifth day of period and did another blood draw.
  - Phase 2:
    - Participants took Pamprin Botanicals and Pamprin OTC before and during period.
    - Completed survey on fifth day of period and did another blood draw.

Focus Consumer Healthcare (Pamprin Botanicals), NAD Case #7247 (Jan. 2024)

- Study was insufficiently reliable to support establishment and health-related claims.
  - Subjects were permitted to take OTC pain relievers during Phase I, which could impact survey results related to pain and bloating.
  - Study relied on survey results of subjective reactions of study participants, but had no control group to determine whether results were due to placebo effect.
  - Objectively measurable results from blood tests failed to show statistically significant results for biomarkers tied to inflammation.

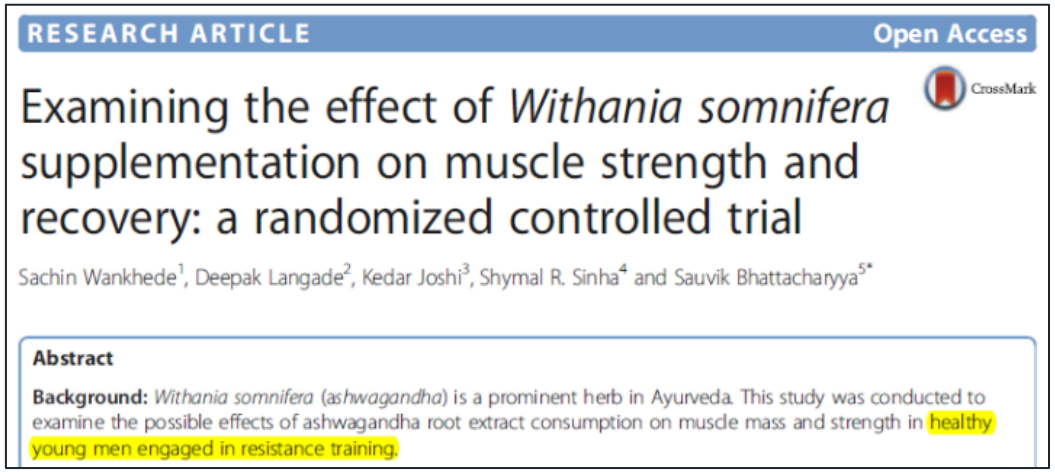
Goli Nutrition Inc. (Goli Ashwagandha Gummies), NAD Case #7059 (April 2022)

- Categories:
  - Weight loss and weight management
  - Sexual function
  - Physical performance (endurance & muscle)
- Types of Claims:
  - Ingredient claims
  - Product claims
  - “Clinically proven” claims



Example #1: Flaws in Goli's Studies
Poor Fit for Goli's Claims

aid for many years. In this study, oral administration of a high concentration root extract of *Ashwagandha* led to increased VO_2max, enhanced cardiorespiratory endurance, and improved QOL in healthy athletic adults. Findings of this study suggest that *Ashwagandha* root extract improves the cardiovascular dynamics by increasing the VO_2max levels thereby enhancing the cardiorespiratory endurance, and also brings an improvement in QOL in healthy adults. However, due to the limited cross-section of the population considered in this study, the findings may not generalize to all populations. Further studies are needed to validate these findings.



RESEARCH ARTICLE Open Access

Examining the effect of *Withania somnifera* supplementation on muscle strength and recovery: a randomized controlled trial

Sachin Wankhede[1], Deepak Langade[2], Kedar Joshi[3], Shymal R. Sinha[4] and Sauvik Bhattacharyya[5*]

Abstract

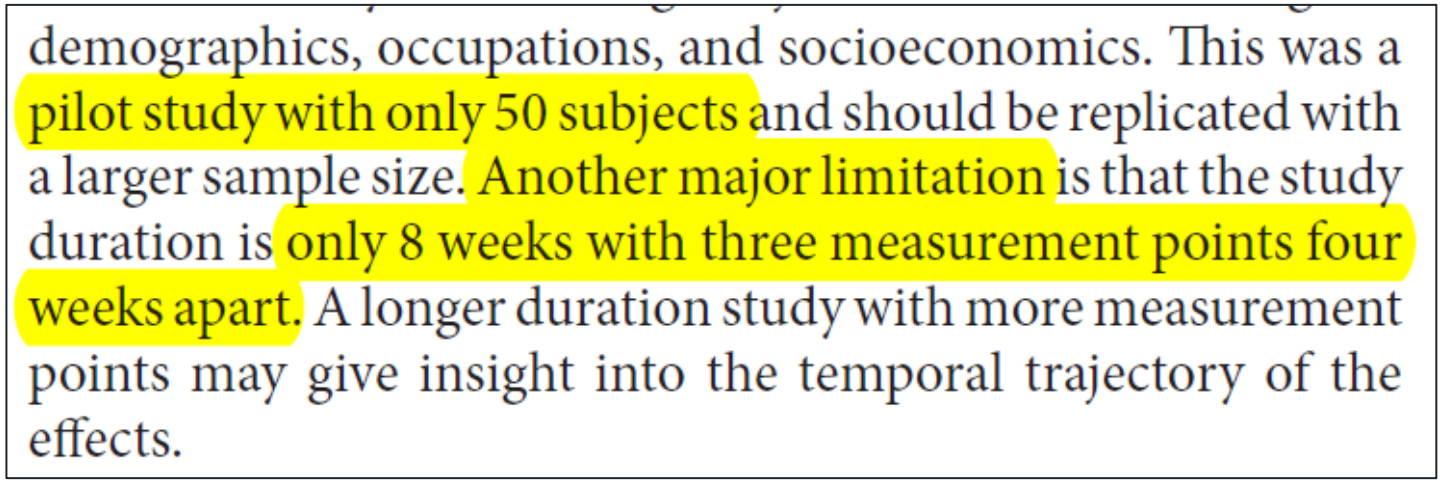
Background: *Withania somnifera* (*ashwagandha*) is a prominent herb in Ayurveda. This study was conducted to examine the possible effects of ashwagandha root extract consumption on muscle mass and strength in healthy young men engaged in resistance training.

Example #2: Flaws in Goli's Studies
Lack of Clinically Meaningful Results

weight reduction group 1 year after the intervention [19]. Consequently, the minimum requirement to improve health hazards in overweight and obese individuals (by WHO classification) can be considered to be 5% weight reduction in Caucasian. In contrast, we hypothesised that a minimum requirement in Japanese population is 3% weight reduction to get similar effects to those in Caucasian populations, and we proved our hypothesis in the minimum

Mean changes in body weight are shown in Table 2. The body weight for both the treatment and placebo groups was found to be reduced during the 8-week period of the study. After 4 weeks of treatment, a mean reduction of 2.14% and 1.09%, from baseline was observed in the treatment and placebo groups, respectively, but the difference in reduction in the 2 groups was not statistically significant after 4 weeks ($P = .0503$). However, at the end of 8 weeks, the reduction of body weight for the treatmentand placebo groups was 3.03% and 1.46%, respectively. The data collected after 8 weeks of

Example #3: Flaws in Goli's Studies
Procedural Failures



demographics, occupations, and socioeconomics. This was a pilot study with only 50 subjects and should be replicated with a larger sample size. Another major limitation is that the study duration is only 8 weeks with three measurement points four weeks apart. A longer duration study with more measurement points may give insight into the temporal trajectory of the effects.

Takeaways: Key Mistakes to Avoid

01

Studies must be a good fit for your advertising claims

- Use a heterogeneous sample that is representative of your target consumers
- Do not employ overly broad exclusion categories
- Your study must be focused on the correct endpoints

02

Studies must reflect clinically meaningful results

- Statistical significance does not always translate to consumer relevance
- Observed results should be attributable to the treatment, alone

03

Studies must use procedures that are accepted in the field

- Control for key variables
- Use a sufficiently large sample
- Employ the appropriate statistical analyses
- Watch out for methodological errors

Sensory Testing

What Constitutes a Sensory Claim?

"A sensory claim is, in essence, a claim about **how consumer[s] themselves react to a product** . . . [T]he messages conveyed to consumers [by a sensory claim] are about how individuals react to, perceive, or sense the products. For such claims, **sensory testing or consumer opinion testing is appropriate**."

"However, when a claim is about the **tangible, objective results that a consumer can expect** a product to provide, **more objective testing is appropriate**."

Kimberly-Clark Corporation (Huggies Natural Care Wipes), Report #5866, NAD/CARU Case Reports (July 2015)

Substantiating Sensory Claims with Percentage Agreement

- 206-subject in-home-use test:
  - **66% reported improvement** in target nail.
  - Of the 85 subjects who had a reference nail and were doctor-diagnosed or picture-approved, **61% saw at least some improvement**.
- NAD: Claim was **supported**.

Advantice Health (Kerasal Fungal Nail Renewal), Report #6421, NAD/CARU Case Reports (October 2020)

Substantiating Sensory Claims with Percentage Agreement

"Improved comfort"

- 1,205-subject consumer-use test:
  - **70% of subjects agreed** with attribute indicating improvement.
- NAD: Claim was **not supported**.

"[T]here was **no indication as to why 70 percent** (as opposed to any other figure) **should be dispositive**."

Reckitt Benckiser LLC (Amopé® GelActiv Insoles), Report #6097, NAD/CARU Case Reports (July 2017)

When Do You Need Experimental Controls?

56-75% of subjects agreed with statements about the product helping to reduce the appearance of fine lines, wrinkles, and age spots. No placebo/control in the study.

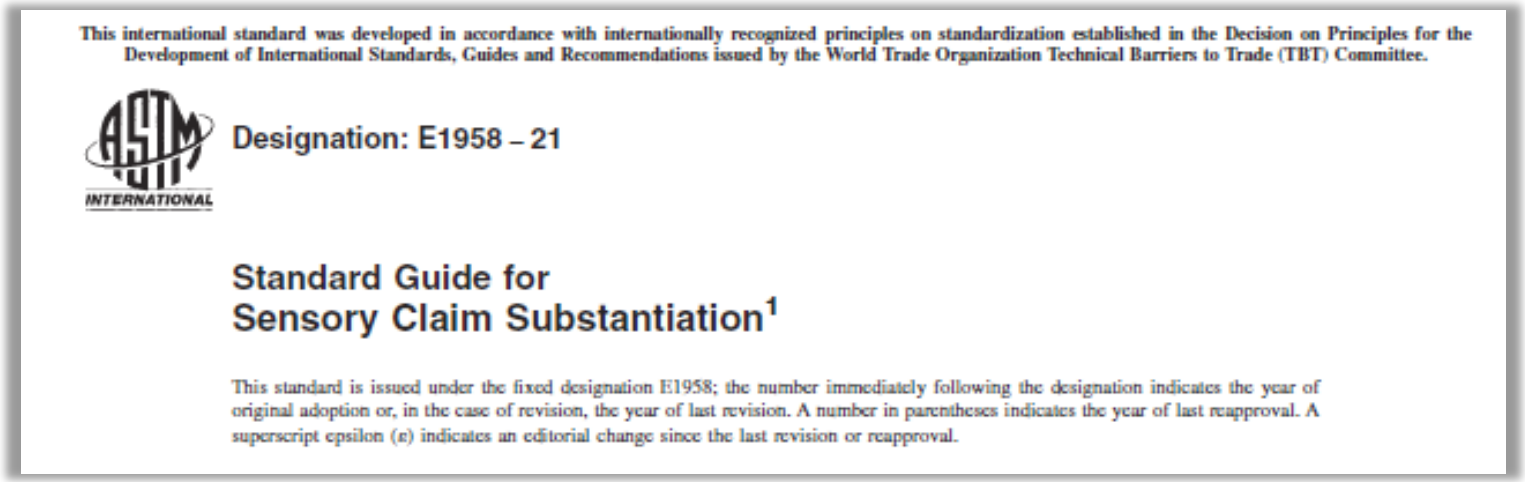
Claim was substantiated.

66-75% of subjects agreed with statements about the product's efficacy. No placebo/control in the study.

Survey deemed unreliable.

Proskauer

ASTM 1958



This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.

Designation: E1958 – 21

Standard Guide for Sensory Claim Substantiation[1]

This standard is issued under the fixed designation E1958; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

Other Considerations

- Blinding
- Yea-saying
- "No Opinion" Option
- Multiplicity
- Statistical Analysis
- Distribution of Results
- Expert vs. Consumer Panels

Statistical Issues

“Statistically Significant”

- Is the result due to chance?
- Type I error: false positive
  - p-value = probability that you will reject the null hypothesis when it is true
  - Usually set threshold at 0.05
  - 5% chance of a false positive = 95% confidence level
- Type II error: false negative
  - Driven by p-value and power level (including sample size, measurement error, and size of effect)
  - 80% power level usually considered acceptable
- Lower p-value means less Type I error but greater risk of a Type II error (and vice versa)

What does it mean if a study finds no statistically significant difference?

- There is no difference OR there is a difference, but it was not detected (i.e., false negative).
- It does not mean that the products/samples tested are equivalent:
  - "The advertiser argued that where there was no statistical difference between TBCC and copper sulfate that one could conclude that TBCC and copper sulfate were equal in performance. However, this conclusion is incorrect. The studies submitted to NAD were all designed to determine if there was a difference between TBCC and copper sulfate. Where no statistical difference is found, the only conclusion that can be made is that the test was not sufficiently powered to detect a difference. Parity or bioequivalence testing involves a different statistical framework."
    - *Novus International, Inc. (Mintrex and MAAC Organic Copper Supplements for Livestock)*, Report #5597, NAD/CARU Case Reports (May 31, 2013)

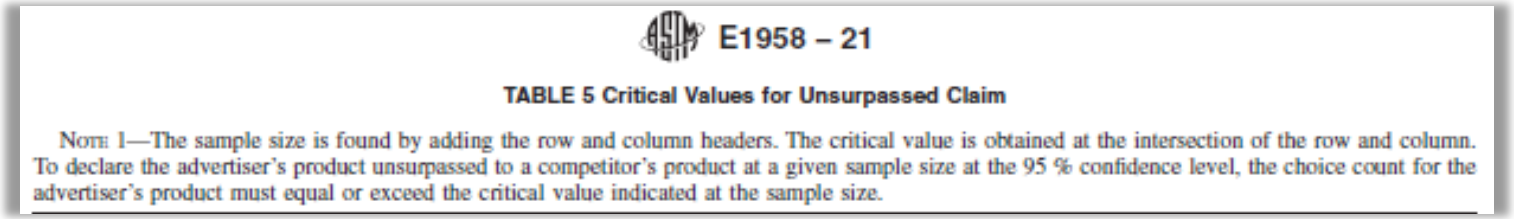
Equivalence and Unsurpassed Claims

- Set an equivalence/non-inferiority margin.
  - This typically needs to be **clinically or consumer meaningful**.
- Equivalence: Establish that your product and the comparator fall within the equivalence margin—are no different—at the 95% confidence interval.
- Unsurpassed claims: Establish that your product is within or above the non-inferiority margin—at least as good—versus the comparator at the 95% confidence interval.
- Beware that larger sample sizes are usually needed for these tests.

ASTM E1958 Methodology for Sensory Claims

Accepted methodology:

- "NAD has upheld the use test methodologies based on this ASTM standard as support for a preference or taste preference claim." *ConAgra Foods, Inc. (Marie Callender's Frozen Three Meat and Four Cheese Lasagna)*, NAD Case #5446 (April 2012)



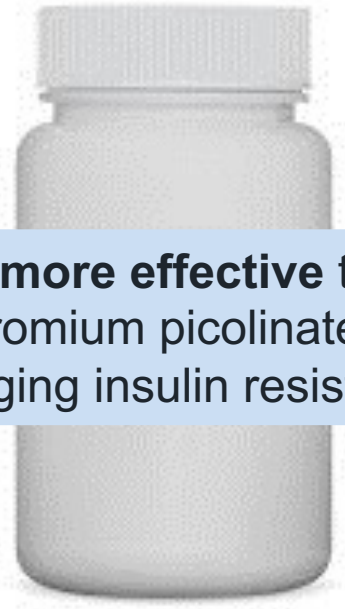
E1958 – 21

TABLE 5 Critical Values for Unsurpassed Claim

NOTE 1—The sample size is found by adding the row and column headers. The critical value is obtained at the intersection of the row and column. To declare the advertiser's product unsurpassed to a competitor's product at a given sample size at the 95 % confidence level, the choice count for the advertiser's product must equal or exceed the critical value indicated at the sample size.

Ratio Claims

ZYCHROME™



"**2x more effective** than chromium picolinate in managing insulin resistance"

"Because the p-values between the Chromax and Zychrome groups here were 0.06, the scientific standard for statistical significance has not been met and the advertiser has not provided competent and reliable scientific evidence to support its claim that "2x more effective than chromium picolinate in managing insulin resistance."

InterHealth Nutraceuticals, Inc. (Zychrome Dietary Supplement), Report #5569, NAD/CARU Case Reports (April 2013)

Ratio Claims

"NAD agreed that using the comparison to the water control to calculate the 3X improvement ratio results in a higher ratio than directly comparing the performance of the products.

Consider a simplified example with three treatments, A B, and X (the basis of comparison), with mean scores of 96, 92, and 90 respectively. The difference between the means of A and X is 6, while the difference between the means of B and X is 2. To state that A is "3X better" than B because of the ratio of differences from X is 3 is entirely misleading because it ignores the magnitude of the underlying measures, i.e., how far each is from zero"

The Procter and Gamble Company (Olay Body Wash), Report #7013, NAD/CARU Case Reports (March 2022)

P-Hacking

- P-Hacking: Post-hoc analysis of data to find a positive (statistically significant) result
  - "NAD has previously questioned the validity of a post-hoc statistical analysis when offered for advertising claim support because it deviates from the underlying study's protocol and ultimately undermines the reliability of the results."
    - *Similac Human Milk Fortifier (Similac® Human Milk Fortifier™)*, Report #5867, NAD/CARU Case Reports (July 2015)
- Establish your statistical analysis plan at the outset.

Multiplicity of Endpoints

- Multiplicity: Testing for numerous endpoints increases the odds of a Type I error.
  - Each p-value is a chance for a false positive.

- Correct for multiplicity using statistical measures
  - Most conservative method = Bonferroni Correction
    - Increase significance level based on the number of p-values
    - E.g., for 20 p-values, require significance at p=.0025 (instead of 0.05)

Q&A

Proskauer