The Institute for Perception

For technical reports from The Institute for Perception or for information about short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reprinted from IFPress (2017) 20(2) 3,4

Claiming Superiority and Equivalence Simultaneously

Daniel M. Ennis

Background: The ASTM *Standard Guide for Sensory Claim Substantiation*[1] specifies the criteria for superiority, equivalence, and unsurpassed claims. A summary of these types of claims is outlined in Figure 1.

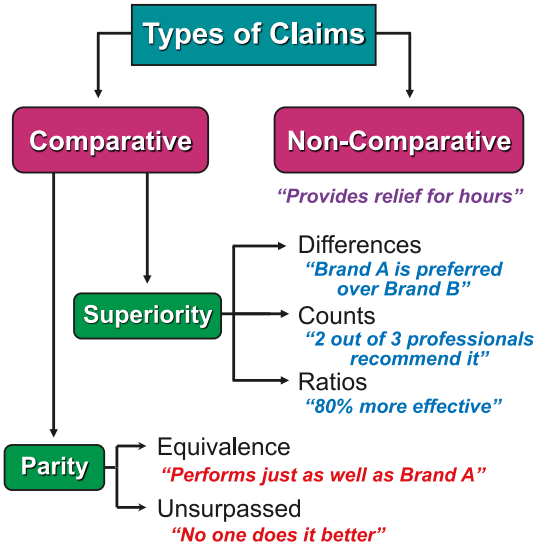

Figure 1. Types of advertising claims.

Equivalence claims are based on a binary choice between two products and involve two bounds set at 45% and 55%. Within these bounds, two products are considered to be equivalent. **Superiority claims** are established when a choice probability (usually preference, but not always) exceeds 50%. **Unsurpassed claims** are made when the choice probability exceeds 45%. The idea behind this latter choice is based on the definition of an unsurpassed claim. An unsurpassed claim combines the concept of equivalence, with a 45% lower bound, and superiority, which has an upper bound of 100%. Figure 2 illustrates these three types of claims

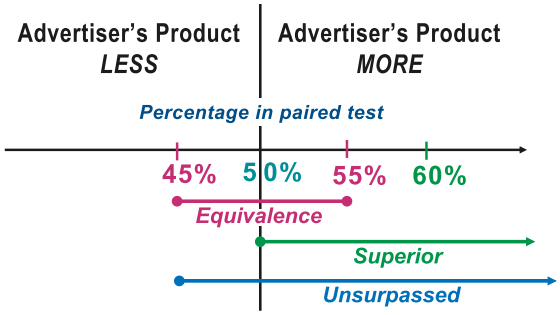

Figure 2. Bounds defining advertising claim types.

and their corresponding bounds. The fact that these three types of claims involve different specifications can lead to a paradoxical outcome in claims support. All of these claims are based on statistical testing involving the appropriate null hypothesis, usually conducted at the 95% confidence level (Type I error of 5%).

In this technical report we will consider the problem created by these different specifications and explore the likelihood of a claims conflict.

Scenario: You work in the claims substantiation division of a major manufacturer of plug-in fragrance products. Competitors in this category often compete regarding the longevity of their fragrance products. You know in advance that the fragrance strength of your product after 60 days does not exceed that of your competitor, but you hypothesize that its fragrance may be *equivalent* to your competitor. Based on the 45%:55% bounds for equivalence, you know from equivalence theory and tables[2,3] that your test will require a sample size in excess of 400. A power analysis, shown in Table 1, reveals that if the true choice probability is 50% then to achieve about 80% power would require a sample size of 900 consumers. From previous research, you also know that the true choice probability that your product will have greater fragrance intensity is a little less than 50%. For this reason you choose a sample size of 1200 which provides an 80% chance of supporting equivalence provided that the true choice probability is 48.6% as shown in Table 1.

n	P_c	Power	80% Detect
400	49.3	27.4	–
500	48.8	43.9	–
600	48.5	56.2	–
700	48.3	65.5	–
800	48.0	75.7	–
900	47.9	80.6	49.7
1000	47.7	86.3	49.1
1100	47.5	90.3	48.8
1200	47.4	93.1	48.6

Table 1. Power characteristics for equivalence test: $\alpha = 5\%$. P_c is the minimum observed lower choice percent required to claim equivalence. Power is the probability of claiming equivalence when the actual choice probability for the advertiser's product is 50%. 80% Detect is the actual lower choice probability that has an 80% likelihood of being declared equivalent.

Following a 60 day aging process for your product and your competitor, the aged products are tested in odor chambers under controlled conditions. The 1200 recruited consumers of the product category evaluate the chambers in pairs. They are instructed to report the chamber that has the greater fragrance intensity or whether they do not detect a difference. Table 2 reports the results after equal splitting of the no difference counts.

QUESTION: Fragrance Strength		
n	Advertiser / Competitor	Test ($p < 0.05$)
1200	569 / 631 (47.4% / 52.6%)	Equivalence: Advertiser 569 – 600
		Superiority: Competitor > 628

Table 2. Advertiser and competitor choice counts in a paired fragrance strength test.

Analysis of the data using equivalence tables[2] leads to the conclusion that the products are equivalent at the 95% confidence level. Table 2 shows that to declare equivalence would require that the lower choice count out of 1200 fall within 569 and 600 inclusive. Your marketing group proceeds to advertise the equivalence of the two plug-in products in fragrance strength after 60 days.

The Challenge: Your competitor has its own consumer research showing that their plug-in product smells stronger than your company's product after 60 days. They initiate an NAD[4] complaint and provide their research support. Since your attorneys expect the decision to rest heavily on technical arguments, your company agrees to participate. Otherwise it could lead to an expensive litigated case where the technical arguments may not be properly considered. In the process of participating, you are requested to submit your claim substantiation data. Your competitor then provides the NAD with an analysis of your data that shows that their product has significantly greater fragrance intensity and is therefore superior with regard to fragrance longevity after 60 days. This analysis is shown in Table 2. They demonstrate that superiority is established if their choice count equals or exceeds 628 out of 1200. The count for their product is 631, which exceeds 628. This analysis is also shown in Table 2.

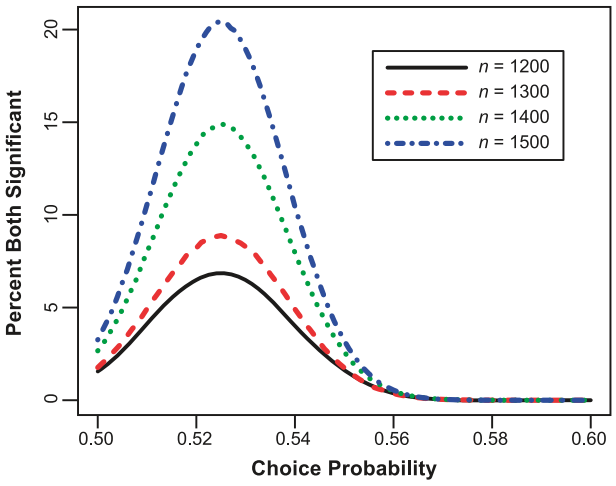

Figure 3. The joint probability of equivalence between two products and the superiority of a competitor's product.

Superiority and Equivalence: The paradox just described can arise in product tests designed to test for equivalence. How often can this situation arise? Figure 3 shows the joint probability that a competitor could claim superiority and an advertiser claim equivalence depending on the sample size and the extent to which the competitor is superior. Larger sample sizes increase the joint probability from about 7% at 1200 to about 20% at 1500 with competitor superiority at about 52.5%. Above this level of superiority, the probability of declaring equivalence decreases and below this level, the probability of superiority decreases. The circumstances leading to the paradox, as reported in Table 2, are fairly specific – the competitor's choice probability must be close to 52.5%, and the sample size must be 1200+. Otherwise support for equivalence and superiority at the same time is not likely to be observed.

A Consumer Relevant Bound for Superiority[5]: If there is general agreement that the bounds defining equivalence in binary choice experiments are 45%:55%, then these bounds may be used to define a region where consumers are indifferent to product differences. For this reason, it is reasonable to consider a lower bound of 55% to define superiority and a statistical test would use a null hypothesis of 55%. There is no general agreement in the product testing community for using such a standard because it would limit the opportunity to claim superiority, but it is worth considering its implications. Regarding the equivalence/superiority paradox it would provide a resolution. Any test result showing a competitor's choice probability above 55% would simultaneously not support equivalence. This is because the lower choice count would fall below 45% and the lower choice count for equivalence must fall between 45% and 50% at any sample size. Conversely, any test result supporting equivalence would not support competitor superiority because the competitor's choice count would not exceed 55%. Another way to resolve the paradox, at any given sample size, is to use the 50% null hypothesis for superiority, but require the point estimate (the value obtained from the experiment) to exceed 55% or any other specified upper bound. This would prevent a competitor from claiming superiority for small differences which may occur in very large sample sizes. Neither of these two approaches have been generally adopted in the analysis of binary choice data, but their merits should be discussed.

Response to the Challenge: In order to provide a persuasive argument for equivalence based on your data, it will be necessary to present a case for a consumer-relevant action standard. While you do not dispute the superiority of your competitor, your argument is that the difference, although significant at the 95% level, is not large enough to be consumer relevant. This argument may have more traction in an NAD hearing than in a litigated case where the awareness of nuanced technical issues may be less well developed. Since the scientific community in the field has not yet accepted a standard greater than 50% to support superiority claims, it will be necessary to establish a precedent.

Conclusion: Technically, it is entirely reasonable to support the equivalence of two products *and* the superiority of one product over another with the same binary choice data. To achieve consistency between the test conclusions, one should pay attention to establishing choice probabilities that take into account a consumer-relevant action standard and avoid seeming contradictions.

References

1. ASTM International. (2012). ASTM E1958-12 Standard guide for sensory claim substantiation. West Conshohocken, PA: ASTM
2. Ennis, D. M., Rousseau, B., and Ennis, J. M. (2014). *Tools and Applications of Sensory and Consumer Science.* (page 172), Richmond, VA: The Institute for Perception.
3. Ennis, D. M. and Ennis, J. M. (2009). Hypothesis testing for equivalence defined on symmetric open intervals. *Communications in Statistics – Theory and Methods.* **38**(11), 1792-1803.
4. The National Advertising Division® (NAD®) is a self-regulatory unit of the Advertising Self-Regulatory Council (ASRC).
5. Rousseau, B. (2015). Sensory discrimination testing and consumer relevance. *Food Quality and Preference,* **43**, 122-125.