International Journal of Communication 16(2022), 3268–3289 1932–8036/20220005

Quantifying the Evidential Value of Celebrity Endorsement: A *p*-Curve Analysis

SHIYUN TIAN
Sacred Heart University, USA

RUOYU SUN
University of Miami, USA

QIAN HUANG
Miami University, USA

JOHN PETIT
University of Miami, USA

Celebrity endorsements have long been used as a promotional tool in marketing communication. However, literature has documented inconsistent findings on the effects of celebrity endorsements compared to no endorsement or noncelebrity endorsements, suggesting a close examination about the reliability and robustness of celebrity endorsements is needed. This study conducted a *p*-curve analysis among two sets of published studies based on different comparison groups (celebrity endorsements vs. no celebrity endorsement; celebrity endorsements vs. noncelebrity endorsements) to investigate if both sets of studies contain an evidential value. The significantly right-skewed *p* curve suggests that both sets of published studies have some integrity. However, the studies that compared celebrity endorsements with no celebrity endorsements showed low statistical power. Theoretical and methodological implications for celebrity endorsement research were discussed.

Keywords: p-curve analysis, celebrity endorsements, advertising effectiveness, evidential value

Celebrity endorsement is a popular marketing strategy employed to influence consumers. It is defined as "an agreement between an individual who enjoys public recognition (i.e., a celebrity) and an entity (i.e., a brand) to use the celebrity for the purpose of promoting the entity" (Bergkvist & Zhou,

Shiyun Tian: tians@sacredheart.edu
Ruoyu Sun: ruoyu.sun@miami.edu
Qian Huang: huangq19@miamioh.edu
John Petit: johnpetit@miami.edu
Date submitted: 2020-05-25

Copyright © 2022 (Shiyun Tian, Ruoyu Sun, Qian Huang, and John Petit). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at http://ijoc.org.

2016, p. 644). Traditionally, a celebrity endorsement involves a celebrity conveying a promotional message in an advertisement intended to persuade consumers. A real-life example of a celebrity endorsement would be George Clooney calmly smiling into the camera, while praising the quality and taste of Nespresso coffee. In addition to explicit endorsements, celebrity endorsements can also be implicit (i.e., George Clooney merely appears with the Nespresso coffee; McCracken, 1989). As such, the intrinsic value of celebrity endorsements lies within the notion that the brand can profit from its association with the celebrity through improving brand awareness, brand liking, and purchase intention (Spry, Pappu, & Cornwell, 2011). In addition to advertising, celebrity endorsements are also integrated into health campaigns (Young & Miller, 2015), political communication (Pease & Brewer, 2008), and nonprofit organizations' advocacy (Wheeler, 2009).

Extensive literature has demonstrated that consumers who are exposed to the systematic pairing of a product with a highly regarded celebrity will develop more favorable responses toward the brand (e.g., Till, Stanley, & Priluck, 2008) than those exposed to other types of endorsers (i.e., an anonymous model) or no endorser (i.e., product only). Nevertheless, a series of studies have documented the ineffectiveness of celebrity endorsement in influencing cognitive, affective, and behavioral outcomes (e.g., Friedrich & Nitsch, 2019; Schouten, Janssen, & Verspaget, 2020). One recent meta-analysis found a zero overall effect of celebrity endorsements (Knoll & Matthes, 2017).

These contradictory findings cast doubt on the robustness of celebrity endorsement effects and raise an empirical need to examine methodological, theoretical, and empirical concerns surrounding the validity of celebrity endorsement. One possible concern that may arise is questionable research practices (QRPs; Simonsohn, Nelson, & Simmons, 2014a), including selectively reporting results, removing cases without particular reason, and conducting inappropriate data analyses. QRPs enable researchers to find statistically significant results for cases where the sample size is too small to reliably detect the effect they are studying, as well as for effects that do not exist in reality (Simonsohn et al., 2014a). In the presence of QRPs, results from estimated effect sizes will be inflated, thereby biasing the findings of meta-analyses once these results are included in such analyses (Simmons, Nelson, & Simonsohn, 2011). As a consequence, these practices adversely influence the theoretical soundness and replicability of studies testing the effectiveness of celebrity endorsement. Therefore, it is necessary to investigate whether the effect of celebrity endorsement is evidential.

In 2020, communication scholars proposed "An Agenda for Open Science in Communication," which pointed out that issues (i.e., replication crisis, publication bias, *p*-hacking) facing the neighboring disciplines may also exist in communication given the overlap of theories, methods, and publication practices in quantitative social science (Dienlin et al., 2020). Therefore, they called for *open communication research* to examine potential QRPs and the validity of research in our discipline. To assess the reliability and robustness of celebrity endorsement studies and respond to this call, the current research used a *p*-curve analysis to estimate the evidential value, statistical power, and selective reporting of celebrity endorsement literature. Specifically, the present research looked into published studies that have tested the effectiveness of celebrity endorsements by comparing celebrity endorsers with other types of endorsers or no endorser.

Questionable Research Practices and *p*-Curve Analysis

Over the years, quantitative social science research has held a false belief that statistical significance equals "real" effects (Dienlin et al., 2020). This notion is reflected in that significant results are more likely to be published (Rosenthal, 1979). Unfortunately, this encourages researchers to adopt QRPs to achieve statistically significant findings, such as selective or distorted reporting and running inappropriate analyses (Simonsohn, Nelson, & Simmons, 2014b). Simmons, Nelson, and Simonsohn (2011) termed these practices *p*-hacking. The significant results produced through *p*-hacking are neither robust nor reliable, raising concerns about whether a set of significant findings represent a real phenomenon or false positives misconstrued as statistical significance (Simmons et al., 2011).

Estimating this error can improve the rigor and reliability of the celebrity endorsement literature. Generally, the *p*-curve method analyzes the distribution of *p* values among published articles to examine whether the findings provide evidence for true effects, or whether they reflect an artifact of publication bias because of QRPs. This reasoning is based on the statistical evidence that studies with true effects where the null hypothesis is false are more likely to produce low *p* values ($p < .025$) than *p* values in the high range of significance ($.025 < p$s $< .05$; Lehmann, 1986). Therefore, the distribution of *p* values (i.e., the "*p*-curve") for a true effect should be right-skewed. Studies that investigate null effects will produce an equal distribution of *p* values, therefore generating a uniform *p* curve. In other words, a flat *p*-curve indicates that the body of literature lacks evidentiary value. The use of QRPs to squeeze Type I-error findings below the threshold of statistical significance is likely to produce *p* values in the upper range of significance (e.g., $.04 < p < .05$). Studies that are composed of *p*-hacked effects lack evidentiary value and consequently will produce a left-skewed *p* curve.

It is worth noting that *p*-curve analyses are not intended to replace meta-analyses; rather, they complement meta-analyses by providing additional statistical information. First, meta-analysis tests whether celebrity endorsements lead to different consumer responses compared with noncelebrity endorsements or no endorsement, and if so, whether these differences are consistent in the literature and strong enough to be meaningful. Using *p*-curve analyses, on the other hand, enables us to investigate the empirical and statistical integrity rather than the magnitude and direction of celebrity endorsements.

Second, studies generating significant effects in expected directions are more likely to be published; studies producing null or significant effects in opposite directions tend to remain unpublished. As research findings accumulate over time, *p*-hacking artificially increases Type I errors and may demonstrate that effects exist when, in reality, they do not (Rosenthal, 1979). Since published research is more likely to be examined in a meta-analysis, skewed results in the examined literature are likely to be reflected in the findings of the meta-analysis as well.

So far, there are two quantitative meta-analyses in the literature testing the effectiveness of celebrity endorsements on consumer responses (Amos, Holmes, & Strutton, 2008; Knoll & Matthes, 2017), with the most recent one being published in 2017. This suggests that the literature does not need another meta-analysis. Furthermore, while Knoll and Matthes's (2017) meta-analysis relied on funnel plot asymmetry (funnel plots, Egger's regression test) to detect publication bias, researchers who criticized these methods argued that various explanations can also produce an asymmetric funnel plot (Pustejovsky & Rodgers, 2019). In addition, despite publication bias commonly occurring via statistical significance rather than effect size, both these tests adjust for selective reporting through estimating effect size (Simonsohn et al., 2014b). A *p*-curve analysis, by contrast, does not suffer from these limitations. Last, *p*-curve evaluates *p*-hacking and the *consequences* of publication bias, neither the funnel plots nor the Egger's regression test do so. As such, this study sets out to conduct a *p*-curve analysis to determine whether a set of findings in the published celebrity endorsement studies contains evidential value when *p*-hacking can be ruled out as an alternative explanation (Simonsohn et al., 2014a, 2014b), and therefore complements the traditional meta-analysis.

Celebrity Endorsements

The effects of celebrity endorsement have been examined from different perspectives (Knoll & Matthes, 2017). As suggested by the advertising effectiveness model (Lavidge & Steiner, 1961), the effects of celebrity endorsement primarily influence three psychological dimensions: cognitive, affective, and conative. Cognitive effects pertain to raising awareness and transfer knowledge of the endorsed brands by providing information and facts about the brands (Lavidge & Steiner, 1961). Compared with noncelebrity endorsements, celebrity endorsements are effective at directing consumers' attention and elicit a higher level of interest toward the endorsed brand (Hung, 2014; Wei & Lu, 2013). Furthermore, research has demonstrated that celebrity endorsements are effective at transferring cultural meanings to the endorsed object (Miller & Allen, 2012) and influencing consumers' perceptions of the perceived product quality and attributes (Biswas, Biswas, & Das, 2006; Jin & Phua, 2014). The affective effects of celebrity endorsement pertain to positive feelings and favorable attitudes toward the ads and the endorsed brands (Lavidge & Steiner, 1961). When compared with other types of endorsers, celebrities are often perceived as more credible, likeable, and attractive (La Ferle & Choi, 2005), and therefore more likely to elicit positive consumer attitudes. The behavioral effects of celebrity endorsement pertain to stimulating desire and behavioral manifestations (Lavidge & Steiner, 1961). As suggested by the theory of planned behavior (Ajzen, 1991), celebrity endorsements influence actual behavior by promoting positive attitudes toward the ad/endorsed brand, subsequently resulting in enhanced purchase intentions. Results of previous research have demonstrated that celebrity endorsements affected behavioral outcomes, such as purchasing endorsed products (Roozen & Claeys, 2010), supporting charitable causes (Wheeler, 2009), and voting for a political candidate (Pease & Brewer, 2008).

With these findings, researchers and practitioners seem to have taken the positive effects of celebrity endorsements for granted for a long time. Until recently, some doubts concerning the effectiveness of celebrity endorsement have been raised (Ambroise & Albert, 2020; Schimmelpfennig & Hunt, 2020). Indeed, many empirical studies have found that celebrity endorsements have led to negative consumer responses (i.e., negative brand attitudes, reduced acceptance of product claims; Boerman, Willemsen, & Van Der Aa, 2017; Dekker & Van Reijmersdal, 2013) or produced null effects (e.g., Biswas et al., 2006; Friedrich & Nitsch, 2019; Young & Miller, 2015). However, despite being important indicators of QRPs (Simonsohn et al., 2014b), the mixed results of celebrity endorsement effects have not been reflected in literature reviews on the subject, and thus receive very little attention from the entire field (Ambroise & Albert, 2020). Besides, a recent meta-analysis (Knoll & Matthes, 2017) challenged conventional thought by showing an almost zero overall effect of celebrity endorsement on consumer responses. Such findings suggest that endorsement by celebrities has hardly any effect on standard measures (e.g., awareness, attitude toward ad, purchase intention), except under certain conditions (e.g., moderating effects). Yet, many studies in the celebrity endorsement literature predict and confirm positive main effects of such endorsement, raising a need to examine whether these results are achieved by incorrectly rejecting the null hypothesis because of QRPs.

Of note, it is, of course, possible that mixed results are emerged due to heterogeneity of methodological practices or study design (e.g., different brands/celebrities, different measures). However, the overall reliability of any body of literature is determined by ruling out publication bias (Simonsohn et al., 2014b). Design considerations, systemic factors, and methodological practices have been known to contribute to the low replicability as undisclosed flexibility in experiment implementation and analysis may be involved (Nosek et al., 2015). Therefore, one cannot conclude that evidentiary value is present in the celebrity endorsement literature without accounting for selective reporting and *p*-hacking.

Finally, no prior studies have evaluated the statistical power of celebrity endorsement literature. As statistical power has been identified as a major factor influencing replicability and credibility of scientific research (Świątkowski & Dompnier, 2017), assessing whether celebrity endorsement research is underpowered or adequately powered is clearly warranted. Taken together, a close examination is imperative for the field to distinguish between QRPs and other factors that may contribute to the mixed results.

Possible Reasons for Mixed Results

First, QRPs may account for the mixed findings. As significant results in expected directions are more likely to be published (Rosenthal, 1979), researchers may engage in QRPs to achieve "desired" results. Consistent use of QRPs will lead to a research bias because it allows researchers to "discover" significant relationships between unrelated factors (Simmons et al., 2011). As researchers may quit conducting other analyses after obtaining statistically significant findings, *p*-hacking is disproportionately prone to generating "large" significant *p* values (i.e., *p* values close to .05). As a result, smaller effect sizes are more likely to be generated, becoming less accurate in evaluating the true effects of the study. Specifically, small sample sizes and *p* values close to a significance level of .05 may suggest the presence of *p*-hacking as observed effects under these two conditions may be false positives when the null hypothesis is true (Simonsohn et al., 2014b). In fact, some studies in the celebrity endorsement literature rely upon small sample sizes (e.g., Chen, Lin, & Hsiao, 2012; Till et al., 2008) and statistically significant *p* values close to .05 (e.g., Maronick, 2006; Pease & Brewer, 2008).

Second, variations in celebrity endorsement effects may be caused by different types of comparison groups across studies. Experimental celebrity endorsement research usually compares celebrity endorsers with either no endorser (i.e., product only or no product/celebrity pairing) or noncelebrity endorsers (i.e., an anonymous model). It is very likely that celebrity endorsers are more effective compared with a certain type of comparison targets than compared with another type of comparison targets. A meta-analysis (Knoll & Matthes, 2017) found that celebrity endorsements performed best compared with no endorsement, and this effect was decreased when compared with noncelebrity endorsements. Nevertheless, it is unclear whether the effects based on both comparison groups contain evidential value. If not, the findings of the meta-analysis are compromised.

Third, consumer perceptions may have contributed to the varying effects of celebrity endorsements. The ways consumers perceive the attributes of celebrity endorsers appears to influence whether the celebrity could resonate well with consumers. Celebrity endorsements are often ineffective when participants do not identify with the celebrity (Schouten et al., 2020), and perceive the celebrity as less credible (Djafarova & Rushworth, 2017) and incongruent with the endorsed products (Roozen & Claeys, 2010). In addition, consumers' perceptions of the motivation behind the product claims also affect celebrity endorsements. According to the Persuasion Knowledge Model (Friestad & Wright, 1994), if participants attribute celebrity endorsements to financially motivated interests rather than genuine liking for the endorsed product, they tend to trust the advertisement less and respond unfavorably (Boerman et al., 2017).

In sum, it is crucial to examine whether the inconsistent findings of celebrity endorsement effects stem from types of comparison groups, potential QRPs, individual factors, or a Type II error (false negative). By analyzing the distribution of *p* values among a set of published articles, *p*-curve analysis allows us to investigate whether the results provide evidence for a true effect, or whether they reflect an artifact of publication bias because of potential QRPs (Simonsohn et al., 2014a). Estimating statistical power and research bias caused by QRPs can improve the reliability and robustness of the celebrity endorsement literature. We thus propose three general research questions:

RQ1: Do the findings of examined studies on celebrity endorsements contain evidential value? Do QRPs provide a likely explanation of these reported effects?

RQ2: What is the average power of included celebrity endorsement studies?

RQ3: Do p-curves differ in the two types of comparison groups?

Method

As the *p*-curve analysis examines whether a set of *significant* findings contains evidential value when QRPs can be precluded as an alternative explanation (Simonsohn et al., 2014a), the current study tests the distribution of *p* values in published research on celebrity advertising effectiveness that reported significant results only. We thus excluded unpublished studies and studies that resulted in effects that did not reach $p < .05$; this is because selective reporting primarily occurs in published research and that researchers are assumed to only *p*-hack to achieve significant results ($p < .05$).

The present research followed four steps. The first step involved an extensive search of all published articles related to celebrity endorsements before October 2020. The authors conducted independent searches of six advertising journals (*Journal of Advertising, Journal of Advertising Research, Journal of Current Issues & Research in Advertising, International Journal of Advertising, Journal of Interactive Advertising, Journal of Marketing Communications*) and nine marketing journals (*Journal of Consumer Research, Journal of Consumer Psychology, Journal of Marketing, Journal of Marketing Research, Psychology*

& Marketing, Marketing Letters, Journal of the Academy of Marketing Science, Marketing Science, International Journal of Research in Marketing). Online databases (e.g., Communication & Mass Media Complete, Business Source Premier) and Google Scholar were searched using the following keywords: (1) Terms for celebrity endorsements: "celebrity endorsement (s)" OR "celebrity endorser (s)" OR "athlete (s)"; (2) Terms for advertising effectiveness: "advertising effectiveness" OR "consumer responses" OR "consumer evaluations" OR "attitudes" OR "purchase intention" OR "brand-liking" OR "brand recall" OR "credible"; (3) Terms for research methods: "experiment" OR "manipulation."

Second, an inclusion rule was created to evaluate which studies to incorporate into separate analyses for nonendorsed comparison group (NE comparison group) and noncelebrity endorser comparison group (NC comparison group). The inclusion rule captured the research method, the manipulated independent variable, the types of comparison groups, the dependent variable, and the research hypothesis. In total, five inclusion rules were developed:

1. As the current research looks into published studies that focus on the effectiveness of celebrity endorsements by comparing celebrity endorsers with noncelebrity endorsers or no endorser, the studies included in this research should employ an experimental design. Each experiment in multiple-study articles was coded as a single study.
2. The manipulated independent variable was an endorser.
3. Studies that compared the effects of celebrity endorsement between celebrity endorsers (pairing a brand/product/service with a celebrity) and the control condition (presents brand/product/service only or no systematic pairing of a brand/product/service with a celebrity) were included in the NE comparison group. On the other hand, studies that compared the effects of celebrity endorsement between celebrity endorsers and noncelebrity endorsers (pairing of a brand/product/service with a noncelebrity endorser) were included in the NC comparison group. If the experiment was a three-cell design (celebrity endorser vs. noncelebrity endorser vs. no endorser), we relied on the study prediction to decide about which group that study belongs to. If the study intended to compare celebrities with noncelebrities and included a control group as a baseline condition, then it was included in the NC comparison group. If the study predicted that a celebrity endorsement was more effective than no endorsement and no endorsement was more effective than a noncelebrity endorsement, it was included in the NE comparison group.
4. Informed by Lavidge and Steiner's (1961) advertising effectiveness model, the dependent variable in this research was advertising effectiveness, which was measured from cognitive, affective, and behavioral perspectives.
5. Following the instructions of Simonsohn and colleagues (2014a), only statistically significant and directionally consistent studies should be included. Thus, the incorporated study should expect celebrity endorsements to be more effective than noncelebrity endorsements or no endorsement, and such prediction should be supported by the results.

Third, the researchers identified and collected the statistical results about celebrity advertising effectiveness from the included studies. Direct effects of celebrity endorsement on consumer responses from experimental designs were reported in our analysis. In a 2 × 2 factorial experimental design, if an effect was expected to attenuate, the *p* value of an interaction effect was selected. An attenuated interaction is defined as an effect that is smaller under one condition than under another condition. However, if an effect was predicted to reverse, the *p* values of both simple effects were selected. Based on our inclusion rule, as well as following the instruction of Simonsohn and colleagues (2014a, 2015) about the coding of statistical results, a total of 12 studies[1] yielding 12 significant results were included in the NE comparison group, and 16 studies yielding 25 significant *p* values were included in the NC comparison group.

Study information is presented in Disclosure Tables.[2] If the author(s) did not report appropriate test statistics, we attempted to calculate test statistics based on the mean, standard deviation, and sample size of each experimental condition. As a result, these studies were excluded from the analysis, although reported in Disclosure Tables. Lastly, the *p* curve was generated and analyzed by using existing online analysis software available at http://p-curve.com (Simonsohn et al., 2014a). The statistical results were then converted to corresponding *p* values (Simonsohn et al., 2014a).

Computing the p Curve

Two methods are used to test the evidential effect: binomial test and continuous test (Simonsohn, Simmons, & Nelson, 2015). The binomial test compares *p* values into two sets ($p < .025$ and $p > .025$). The continuous test compares *p* values of studies on the continuum of *p* values. Two criteria are used to evaluate the results from each method: full *p* curve and half *p* curve. For the full *p* curve, the distribution is drawn based on studies showing *p* values less than .05, while for the half *p* curve, the distribution is drawn based on the studies showing *p* values less than .025. That is, a series of studies will have evidential value if a) the half *p* curve has a $p < .05$ right-skew test, *or* b) both the full and half *p* curves have $p < .1$ right-skew tests. In this sense, combining full and half *p* curves into a single analysis will eliminate false-positive conclusions of evidential value without decreasing statistical power (Simonsohn et al., 2015).

The null hypothesis of the *p* curve analysis is the distribution of *p* values associated with the analyzed studies forms a flat line. Rejecting this null hypothesis means that the observed studies show evidential effect. If no evidential effect was found, we checked whether the observed distribution of *p* values from the analyzed studies show a flatter distribution than studies with a small sample size (showing the power of 0.33), which means we would find only a one-third chance that the study would

---

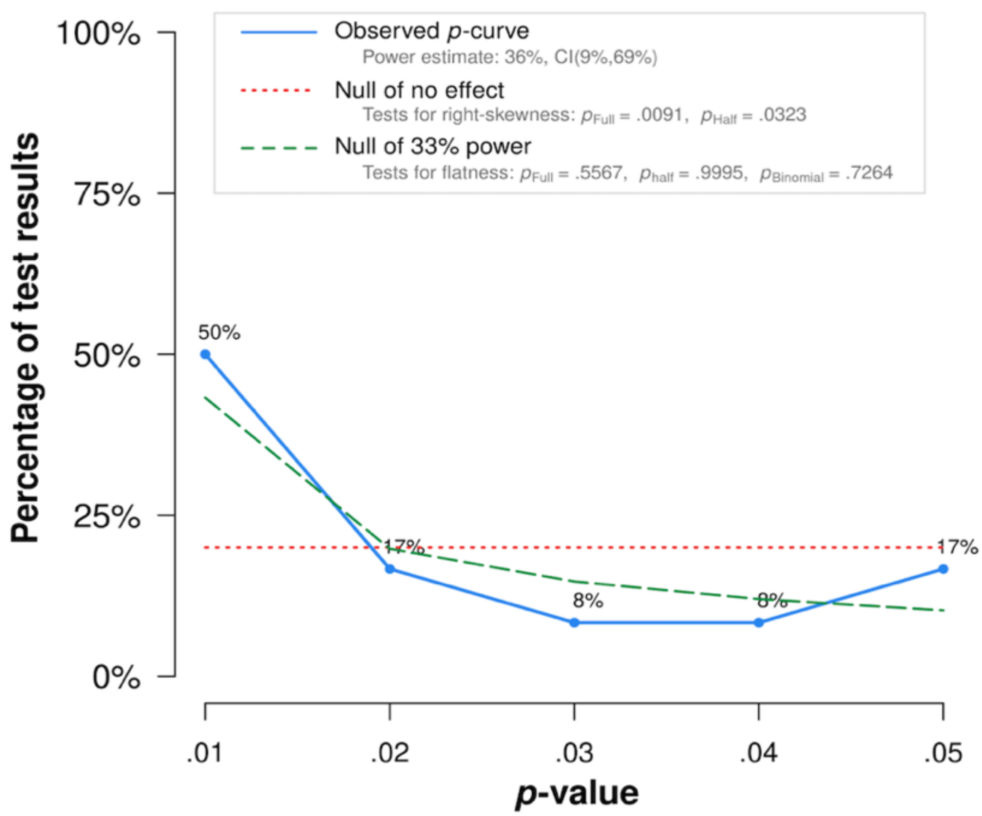
[1] It should be noted that because of different inclusion rules and coding of statistical results, the sample size of the current p-curve analysis is different from Knoll and Matthes's (2017) meta-analysis. Twenty-two studies included in Knoll and Matthes's (2017) meta-analysis were excluded in this p-curve analysis. Four studies were not included because of the statistically insignificant findings (e.g., Park, Turner, & Pastore, 2008). Six studies were excluded because those studies asked only general research questions about the effects of celebrity endorsements but did not propose a specific prediction (e.g., Friedman, Termini, & Washington, 1976). Twelve studies were excluded because those studies did not examine the celebrity endorsement effects this p-curve analysis plans to test. For example, Yoon and Kim (2016) examined whether the celebrity endorser moderates the effect of self-congruity on source credibility.

[2] Disclosure tables for both groups, references of the analyzed studies and all other materials have been made publicly available via Open Science Framework and can be accessed at https://osf.io/zve4x/?view_only=0451eaa9c5a64b1d945e4f7d30d62aa9.

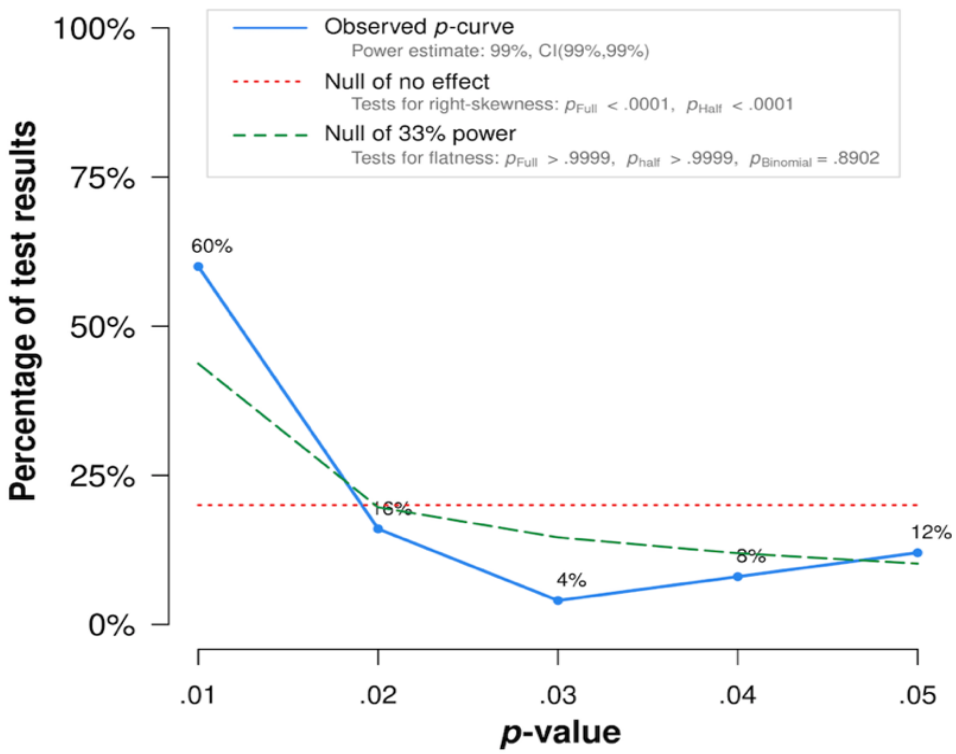
show this effect if the study is replicated. The null hypothesis for 33% power assumes that the small effect is null, saying that the evidential effect is adequate. It is possible that ambitious *p*-hacking still exists even if the *p* value is less than .05 (Simonsohn et al., 2015).

Results

A *p*-curve analysis was conducted for each group. In NE comparison group, 2 of the 14 results were excluded from the overall analysis as their corresponding *p* values were greater than .05: $t\,(132) = 1.78$; $t\,(132) = 1.64$. Thus, a final sample of 12 studies yielding 12 statistically significant results was included in the *p*-curve analysis, among which 9 studies had *p* values smaller than .025. In NC comparison group, one of the 26 results was excluded because of the insignificant *p* value: $t\,(100) = 1.33$. As a result, a sample of 16 studies yielding 25 significant results was included in NC comparison group for analysis, among which 20 were smaller than .025 (calculations for each test entered into the *p* curve was shown in a table uploaded in the Open Science Framework).



a.



b.

Figure 1. The results figure of p curve.

Note. a: NE comparison group; b: NC comparison group. The results display confidence intervals (CIs) for the estimate of the average power of the graphed studies. The blue line is the observed *p* curve, the red dotted line is a reference to test for right skewness, and the green dotted line is to test whether studies with a small sample size showed no effect.

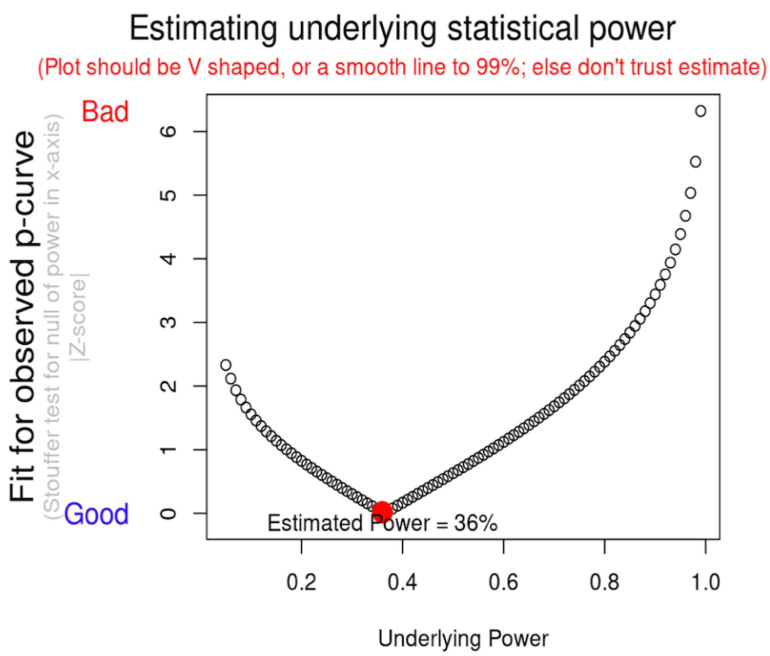
In NE comparison group, we cannot reject the null hypothesis that the observed *p* curve was flat (p = .07) according to the results of the binomial test. However, since both the full *p* curve (z = −2.36, p = .01) and the half *p* curve (z = −1.85, p = .03) were smaller than .05, indicating that the studies included in the current *p*-curve analysis contained evidential value (Simonsohn et al., 2014a,2015), suggesting the existence of celebrity endorsement effects (see Figure 1). Furthermore, the binomial test of 33% power revealed that the evidential value of studies in the NE comparison group was not absent (p = .73) with insignificant full *p* curve (z = .14, p = .56) and half *p* curve (z =3.28, p > .99). We therefore concluded that the curve did not indicate that the evidential value was inadequate (Simonsohn et al., 2014b, 2015). That is to say, the included studies have adequate statistical power to detect the effect of celebrity endorsement.

A separate *p*-curve analysis was conducted for the studies in NC comparison group. The binomial test supported the alternative hypothesis that the *p* curve was right-skewed (p = .002). Both the full (z = −15.66, p < .001) and the half curves (z = −16.79, p < .001) were significant. The results suggest that the effect of celebrity endorsement exists. In addition, the binomial test of 33% power revealed that the evidential value of studies in the NC comparison group was not absent (p = .89) with insignificant full (z = 10.73, p >.99) and half curves (z = 15.12, p > .99), indicating that the included studies have adequate statistical power to detect the impact of celebrity endorsement. Hence, we concluded that the evidential value of studies in NC comparison group was not inadequate.

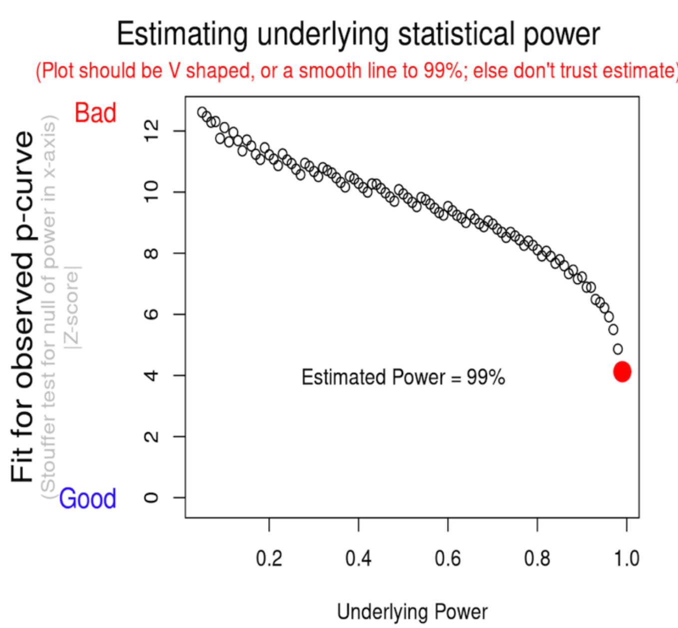
Furthermore, moderately ambitious p-hacking can be detected by analyzing whether the distribution of p values between .00 and .025 (i.e., the half p curve) is right-skewed as the half curve is formed by significant results that have a lower probability of mistaking p-hacking as evidential values (Simonsohn et al., 2015). Here both half-curves for both groups were significantly right-skewed, indicating evidential values in studies of both groups.

While both groups of studies showed evidential value of celebrity endorsement in customers' reactions, the average power of NE comparison group studies (36% under a 90% CI[9%, 69%]) was lower than it is of NC comparison group (99%, under a 90% CI[99%, 99%]). In that sense, for NE comparison group studies, when the true effect size is zero at the .05 level, the lower end of the confidence interval for power was 9%. This means that if we assume that 9% is the level of power, we would observe a p curve that is right-skewed, or more right-skewed, as indexed by the Stouffer combination of the resulting *pp* values only 5% of the time. The other end of the confidence interval (69%) means that if power were that high, we would see a p curve as flat, or flatter, 95% of the time. Therefore, if the same studies are replicated and analyzed again, it is likely that 36% of them will replicate with significant results at the 5% significance level (Simonsohn et al., 2014b). That means nearly 64% of the replications will likely not find significant results.

For NC comparison group studies, we could expect that when the true effect size is zero at the .05 level, both the lower and the higher ends of the confidence interval were 99%, indicating that if the same studies are replicated and analyzed again, it is likely that 99% of them will replicate with significant results at the 5% significance level. In other words, NC comparison group studies are adequately powered (see Figure 2).



a.



b.

Figure 2. Diagnostic plot for power estimation.

Note. a: NE comparison group; b: NC comparison group.

Discussion

The current *p*-curve analysis examined potential publication bias and *p*-hacking among two different groups of studies that tested the effectiveness of celebrity endorsement on consumer responses. The results of our research suggest that the examined body of literature on celebrity advertising has true effects. That is to say, these findings in the literature are very likely to be valid: celebrity endorsement is more effective than no endorsement or noncelebrity endorsement in influencing consumers' cognitive (i.e., increasing brand recall), affective (i.e., eliciting favorable attitudes toward the endorsed products), and behavioral (i.e., stimulating desire and actions to buy endorsed products) responses. Despite these findings, NE comparison group (celebrity endorsement vs. no endorsement) presented relatively low power of the studies, suggesting that these findings might not be easily replicable.

The *p* curve of NC comparison group illustrates the true effects of celebrity endorsements on consumer responses compared with noncelebrity endorsements. It is valid that celebrity endorsers are more effective than noncelebrity endorsers in producing favorable outcomes when brands use celebrities that are congruent with the endorsed products and target consumers. Furthermore, the statistical power of NC comparison group analysis is 99%, which is reassuring to this particular celebrity endorsement scholarship. Studies included in NC comparison group followed a relatively homogeneous study design. Celebrities were usually carefully selected in the pretest to ensure that they were generally liked by the public and were perceived as congruent with the endorsed products. The stimuli employed by these experiments were designed as a real advertisement mostly in a print format (e.g., magazine). Noncelebrity endorsement, on the other hand, was manipulated exactly the same as celebrity endorsement except that the endorser is not a celebrity (i.e., an employee). Future research comparing the effects of celebrity and noncelebrity endorsers that followed the same study design and procedure would be highly likely to successfully replicate these findings.

Although the *p*-curve results of NE comparison group demonstrate the true effects of celebrity endorsers compared with no endorser, the average power of the included studies is relatively low. That is to say, it is less likely that many of them would be replicated. Interestingly, the findings of Knoll and Matthes's (2017) meta-analysis showed that celebrity endorsers perform best when compared with no endorser. The low statistical power therefore raises skepticism about the true distribution of effect sizes found in meta-analyses. Dienlin and colleagues (2020) called that evidence concerning the replicability in communication is needed. Our findings respond to this call by showing that there might exist a replication problem of studies included in NE comparison group. Several factors may contribute to the low power of the effects of celebrity endorsement on consumer responses when compared with no endorsement.

First, the sample size of the included studies may cause the issue of poor statistical power. According to Cohen (1988), a study design and test combination could be underpowered to examine the effect sizes of interest. A study with a small sample size often leads to inadequate power. Almost half of the included studies (5 of 13 studies) in NE comparison group analysis have fewer than 25 participants in each experimental condition. For example, in study 3 of Till and colleagues (2008), only 10–14 participants were included in each group. Likewise, in study 2 and study 4 of Chen and colleagues (2012), only 20 subjects were involved in each condition. The statistically low power of the included experiments demonstrates the possibility of undetected effects of celebrity endorsement when compared with no endorsement.

Furthermore, scholars repeatedly documented that researchers were consistently running underpowered research with the inadequately small sample size for decades, and this practice has continued in recent years (e.g., Matthes et al., 2015; Sedlmeier & Gigerenzer, 1992). Nelson, Simmons, and Simonsohn (2018) speculated that *p*-hacking is the reason that various underpowered studies could be statistically significant, as "researchers did not learn from experience to increase their sample sizes precisely because their underpowered studies were not failing" (p. 515). We cannot make such a conclusion at this moment, but the underpowered studies with a small sample size in communication scholarship deserve our attention. We therefore advocate that sample sizes smaller than 25 participants per condition should report all observed effect sizes.

Second, experimental celebrity endorsement research manipulates product/celebrity pairings primarily in two ways. Some studies manipulated celebrity endorsement through asking participants to view an advertisement that features a celebrity who is endorsing a product (e.g., Pease & Brewer, 2008). On the other hand, some other studies adopted the "forward conditioning procedure" to manipulate celebrity endorsement in which only the images of the focal product and the focal celebrity were shown to the participants instead of an advertisement (e.g., Miller & Allen, 2012; Till et al., 2008). Participants were instructed to view a slide show presenting multiple pairings of an image of the focal product followed by a picture of the focal celebrity (conditioning trial), along with randomized images of the fillers (i.e., filler celebrities/brands, abstract paintings). Although the literature suggests that both types of celebrity endorsement manipulation were effective, especially on brand attitude (Knoll & Matthes, 2017), the variations between how the endorsements were presented to the participants may have influenced the validity of the manipulation and ultimately affected consumer responses. For example, explicitly showing the pairing of the celebrity and the product, such as highlighting the endorsement relationship in the headline or body text of the print ad (i.e., "Professional Athletes Agree"; Petty, Cacioppo, & Schumann, 1983), may be more salient for consumers to process endorsement information, which in turn affects subsequent outcomes. Although the forward conditioning procedure is well established in testing the effect of product/celebrity pairing, researchers noted that the ecological validity of this approach may be questionable (Miller & Allen, 2012), and thus may hinder the detection of the true phenomenon.

Third, repeatedly exposing consumers to an advertisement has been considered as an important way to produce desired effects on target consumers (Pechmann & Stewart, 1988). However, the number of repetitions or conditioning trials in the experiment of celebrity endorsement is associated with different theoretical mechanisms (e.g., Bergkvist, 2017) and may affect study results. Too few repetitions may not be enough to produce effects on target consumers. Ten repetitions were commonly used in previous research (e.g., Kim, Allen, & Kardes, 1996; Miller & Allen, 2012). However, past research findings have also shown that favorable brand attitudes can be obtained with only one exposure (e.g., Kim, Lim, & Bhargava, 1998). On the other hand, viewing too many images may cause fatigue among participants (Till et al., 2008). In the forward conditioning procedure, multiple conditioning trials plus several filler images often include dozens of pictures. Many studies asked participants to view 80 pictures in a slide show (e.g., Kim et al., 1996; Miller & Allen, 2012). In this sense, participants are likely to feel bored, lose patience, and thus not fully concentrate on the stimuli, which ultimately affects the validity of the manipulation.

Furthermore, utilizing filler images is crucial in the forward conditioning procedure as it reduces the potential for demand-artifact interpretations and decreases hypothesis guessing (Kim et al., 1996). While some studies employed both distracter products and celebrities as fillers (e.g., Miller & Allen, 2012; Till et al., 2008), some other studies employed only distracter products or abstract paintings without including any filler celebrities (e.g., Chen et al., 2012). As celebrities are often good at attracting attention, the only celebrity used in the conditioning trial would still possibly stand out in a series of images even with multiple fillers, making participants aware of the study purpose. Failing to include filler celebrities thus casts doubt on the effect of celebrity endorsement as demand artifact can still be an alternative explanation.

The selection of the brand may also influence the effect of celebrity endorsement. To increase the ecological validity of the experiment, real brands were employed in celebrity endorsement experiments. However, to investigate how celebrity endorsement influences consumers' attitudes toward the brand, care should be taken when selecting the brand. A popular brand that is not affectively neutral may influence the study results (Till et al., 2008); it is difficult to tease apart whether the favorable attitude comes from the celebrity or the brand itself. We are not arguing that it is wrong to use popular and mature brands. In fact, we encourage researchers to use well-established brands in their experiments to simulate real-world scenarios. However, we suggest researchers conduct experiments with care as brands that are not affectively neutral can influence consumers' affective responses.

Another important observation is that some included studies in the current *p*-curve analysis selectively reported statistical results and conducted inappropriate statistical analyses. We found some of the included experiments had selectively reported only statistically significant results (e.g., Freiden, 1982; Wheeler, 2009). For example, for a 5 × 2 study design, only one omnibus *F*-test result was reported when the prediction made by the author compared fit celebrity endorsers with the other four types of endorsers (Wheeler, 2009). No pairwise comparison results were provided to clarify if the fit celebrity endorser was indeed better than all four other endorsers as predicted. Some other research reported only regression coefficients for a two-cell design rather than *t*-test or *F*-test results (e.g., Jackson, 2008; Jackson & Darrow, 2005). In addition to reporting bias, some of the included research actually implemented inappropriate analyses. For instance, one study predicted an attenuated interaction effect, whereas the results of a series of *t*-test were used to make conclusions rather than the results of the two-way ANOVA (Aureliano-Silva, Lopes, Freire, & da Silva, 2015). These signs of QRPs are consistent with Matthes and colleagues' (2015) findings that published studies in top journals of communication show evidence of QRPs. Conducting inappropriate analysis increases the likelihood of false positives (Simonsohn et al., 2015), thereby impairing the true distribution of effect sizes (Jennions, Møller, & Hunt, 2004). The QRPs, together with the prevalence of small sample sizes in the NE comparison group, may explain why our findings of this group present both evidential value and low statistical power. Ambitious *p*-hacking and inappropriateness in analysis and implementation may have artificially exaggerated the evidential value.

Recommendations for Future Celebrity Endorsement Research

There are several steps that may improve replication efforts in celebrity endorsement literature and scientific research in general. First, conducting a priori power analysis to determine the appropriate sample size required for the hypothesized effect sizes is crucial, as implementing studies with adequate power helps to detect the real effects of celebrity endorsement, especially when compared with no endorsement. Second, preregistration of experiments can help to improve replicability. This means preregistering as many details as possible about the hypotheses, how data will be analyzed, and what specific dependent variables will be measured. None of the studies analyzed in this *p* curve preregistered their research. Third, it is necessary to report complete statistical results and conduct appropriate statistical analyses in accordance with the study hypotheses, as they are imperative to help us understand the true effects of celebrity endorsement and maintain the integrity of scientific research. In addition, data sharing improves replicability as it urges researchers to be transparent in analyzing and reporting data, while also making it easier for other scholars in the field to assess the literature.

Careful considerations should also be made in selecting the study stimulus. The selected product should be affectively neutral and congruent with the selected celebrity. In addition, researchers who intend to replicate celebrity endorsements need to distinguish between information-rich stimuli that clearly show the relationship between the celebrity and the brand, and stimuli that only include pictures of the celebrity and the brand. Recent research has suggested that consumers process different stimuli in different ways, which in turn leads to different advertising outcomes (e.g., Bergkvist, 2017). A successful replication is determined by ensuring appropriate stimuli, theoretical mechanisms, and relevant outcomes.

Limitations and Future Work

This study has several limitations that can be addressed by future *p*-curve studies. First, the individual results coded in the present *p*-curve analysis only focused on the comparison between celebrity endorsements and no/noncelebrity endorsements. Past research has found that factors such as source credibility and endorser-product congruence can influence celebrity endorsements (e.g., Kamins & Gupta, 1994; Spry et al., 2011). Future research can examine the true effects of celebrity endorsements from other theoretical accounts, such as the match-up hypothesis. Second, the effectiveness of celebrity endorsement is understood broadly in this research. The included studies tested the effect of celebrity endorsement on various outcomes such as brand recall and brand attitude. We acknowledge that these outcomes, although interrelated with each other, may independently be influenced by celebrity endorsement (Vakratsas & Ambler, 1999). Future research may conduct separate analyses for these outcomes to investigate if evidential value still exists with different dependent variables.

Furthermore, like any other statistical approach, *p*-curve analysis is not free of flaws. First, a *p*-curve analysis can only provide information about the sample, and not the whole population (Simmons & Simonsohn, 2017). Therefore, the more studies a *p*-curve analysis includes, the more generalizability the findings of the *p*-curve analysis. Second, researchers have pointed out that *p*-curve analysis performs poorly when the effect sizes across studies are not homogeneous (McShane, Böckenholt, & Hansen, 2016). However, such limitations may not be a concern to this study as the main purpose of this research is to evaluate the average true effect of the analyzed studies, rather than all studies which could ever be attempted (Nelson et al., 2018). In such a case, Simonsohn and colleagues (2014b) have shown that the *p*-curve analysis is robust even when the effect sizes of the included studies are not homogeneous. As such, we believe the results of the present *p*-curve analysis are robust. The approach of *p*-curve analysis is still developing, and we look forward to testing its abilities to detect the *p*-hacking, file-drawer effect, and other forms of QRPs.

Conclusion

The present study used the *p*-curve analysis to examine whether the published findings concerning the effect of celebrity endorsements on consumer responses are true effects. The analyses of two separate groups revealed that the submitted studies demonstrated some evidentiary value, though the literature testing the effect of celebrity endorsement compared with no endorsement is severely underpowered. Our results and our reviews of the included experiments suggest that there are certain indications of QRPs in the included studies. With that being said, we are far from making the conclusion that QRPs are prevailing in celebrity endorsement. Our purpose is to raise awareness of potential QRPs in our discipline. Furthermore, we strongly recommend communication scholars to take statistical power seriously such that the field can continue to build upon adequately powered literature. Aligning with the purpose of conducting *p*-curve analyses, we also encourage researchers to report complete statistical results following established styles (e.g., APA, Chicago) and to use preregistration to improve the transparency of their study design and hypothesis testing. Following these steps could significantly reduce the occurrence of *p*-hacking, which is crucial for us to better understand the true effects of celebrity endorsement as well as various theories in our field.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. doi:10.1016/0749-5978(91)90020-T

Ambroise, L., & Albert, N. (2020). Celebrity endorsement: Conceptual clarifications, critical review, and future research perspectives. *Recherche et Applications en Marketing, 35*(2), 97–122. doi:10.1177/2051570719876198

Amos, C., Holmes, G., & Strutton, D. (2008). Exploring the relationship between celebrity endorser effects and advertising effectiveness: A quantitative synthesis of effect size. *International Journal of Advertising, 27*(2), 209–234. doi:10.1080/02650487.2008.11073052

Aureliano-Silva, L., Lopes, E. L., Freire, O. B. D. L., & da Silva, D. (2015). The brand's effect on the evaluation of advertising endorsed by celebrities: an experimental study. *Brazilian Business Review, 12*(4), 57–78. doi:10.15728/bbr.2015.12.4.3

Bergkvist, L. (2017). Celebrity trait transference: When brands pick up endorsers' personality traits. *International Journal of Advertising, 36*(5), 663–681. doi:10.1080/02650487.2017.1339658

Bergkvist, L., & Zhou, K. Q. (2016). Celebrity endorsements: A literature review and research agenda. *International Journal of Advertising, 35*(4), 642–663. doi:10.1080/02650487.2015.1137537

Biswas, D., Biswas, A., & Das, N. (2006). The differential effects of celebrity and expert endorsements on consumer risk perceptions. The role of consumer knowledge, perceived congruency, and product technology orientation. *Journal of Advertising, 35*(2), 17–31. doi:10.1080/00913367.2006.10639231

Boerman, S. C., Willemsen, L. M., & Van Der Aa, E. P. (2017). "This post is sponsored": Effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook. *Journal of Interactive Marketing, 38*(1), 82–92. doi:10.1016/j.intmar.2016.12.002

Chen, C. Y., Lin, Y. H., & Hsiao, C. L. (2012). Celebrity endorsement for sporting events using classical conditioning. *International Journal of Sports Marketing & Sponsorship, 13*(3), 46–56. doi:10.1108/ijsms-13-03-2012-b005

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dekker, K., & Van Reijmersdal, E. A. (2013). Disclosing celebrity endorsement in a television program to mitigate persuasion: How disclosure type and celebrity credibility interact. *Journal of Promotion Management, 19*(2), 224–240. doi:10.1080/10496491.2013.769473

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., . . . Huskey, R. (2020). An agenda for open science in communication. *Journal of Communication, 71*(1), 1–26. doi:10.1093/joc/jqz052

Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior, 68*, 1–7. doi:10.1016/j.chb.2016.11.009

Freiden, J. B. (1982). An evaluation of spokesperson and vehicle source effects in advertising. *Current Issues and Research in Advertising, 5*(1), 77–87. doi:10.1080/01633392.1982.10505323

Friedman, H. H., Termini, S., & Washington, R. (1976). The effectiveness of advertisements utilizing four types of endorsers. *Journal of Advertising, 5*(3), 22–24. doi:10.1080/00913367.1976.10672647

Friedrich, K., & Nitsch, C. (2019). Celebrity political endorsement and young voters in Europe: A five-country comparison on celebrity support effectiveness in the European elections. *International Journal of Communication, 13*, 4874–4894.

Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research, 21*(1), 1–31. doi:10.1086/209380

Hung, K. (2014). Why celebrity sells: A dual entertainment path model of brand endorsement. *Journal of Advertising, 43*(2), 155–166. doi:10.1080/00913367.2013.838720

Jackson, D. J. (2008). Selling politics: The impact of celebrities' political beliefs on young Americans. *Journal of Political Marketing, 6*(4), 67–83. doi:10.1300/J199v06n04_04

Jackson, D. J., & Darrow, T. I. A. (2005). The influence of celebrity endorsements on young adults' political opinions. *Harvard International Journal of Press/Politics, 10*(3), 80–98. doi:10.1177/1081180X05279278

Jennions, M. D., Møller, A. P., & Hunt, J. (2004). Meta-analysis can "fail": Reply to Kotiaho and Tomkins. *Oikos, 104*(1), 191–193. doi:10.1111/j.0030-1299.2004.12265.x

Jin, S. A. A., & Phua, J. (2014). Following celebrities' tweets about brands: The impact of twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. *Journal of Advertising, 43*(2), 181–195. doi:10.1080/00913367.2013.827606

Kamins, M. A., & Gupta, K. (1994). Congruence between spokesperson and product type: A matchup hypothesis perspective. *Psychology & Marketing, 11*(6), 569–586. doi:10.1002/mar.4220110605

Kim, J., Allen, C. T., & Kardes, F. R. (1996). An investigation of the mediational mechanisms underlying attitudinal conditioning. *Journal of Marketing Research, 33*(3), 318–328. doi:10.2307/3152128

Kim, J., Lim, J. S., & Bhargava, M. (1998). The role of affect in attitude formation: A classical conditioning approach. *Journal of the Academy of Marketing Science, 26*(2), 143–152. doi:10.1177/0092070398262005

Knoll, J., & Matthes, J. (2017). The effectiveness of celebrity endorsements: A meta-analysis. *Journal of the Academy of Marketing Science, 45*(1), 55–75. doi:10.1007/s11747-016-0503-8

La Ferle, C., & Choi, S. M. (2005). The importance of perceived endorser credibility in South Korean advertising. *Journal of Current Issues & Research in Advertising, 27*(2), 67–81. doi:10.1080/10641734.2005.10505182

Lavidge, R. J., & Steiner, G. A. (1961). A model for predictive measurements of advertising effectiveness. *Journal of Marketing, 25*(6), 59–62. doi:10.2307/1248516

Lehmann, E. L. (1986). *Testing statistical hypotheses*. New York, NY: John Wiley.

Maronick, T. J. (2006). Celebrity versus company president as endorsers of high risk products for elderly consumers. *Journal of Promotion Management, 11*(4), 63–80. doi:10.1300/J057v11n04_06

Matthes, J., Marquart, F., Naderer, B., Arendt, F., Schmuck, D., & Adam, K. (2015). Questionable research practices in experimental communication research: A systematic analysis from 1980 to 2013. *Communication Methods and Measures, 9*(4), 193–207. doi:10.1080/19312458.2015.1096334

McCracken, G. (1989). Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research, 16*(3), 310–321. doi:10.1086/209217

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11*(5), 730–749. doi:10.1177/1745691616662243

Miller, F. M., & Allen, C. T. (2012). How does celebrity meaning transfer? Investigating the process of meaning transfer with celebrity affiliates and mature brands. *Journal of Consumer Psychology, 22*(3), 443–452. doi:10.1016/j.jcps.2011.11.001

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69*(1), 511–534. doi:10.1146/annurev-psych-122216-011836

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. doi:10.1126/science.aab2374

Park, M., Turner, B. A., & Pastore, D. L. (2008). Effective public service advertisements to attract volunteers for the Special Olympics: An elaboration likelihood perspective. *Sport Management Review, 11*(2), 165–192. doi:10.1016/S1441-3523(08)70108-6

Pease, A., & Brewer, P. R. (2008). The Oprah factor: The effects of a celebrity endorsement in a presidential primary campaign. *The International Journal of Press/Politics, 13*(4), 386–400. doi:10.1177/1940161208321948

Pechmann, C., & Stewart, D. W. (1988). Advertising repetition: A critical review of wearin and wearout. *Current Issues and Research in Advertising, 11*(1–2), 285–329. doi:10.1080/01633392.1988.10504936

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research, 10*(2), 135–146. doi:10.1086/208954

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods, 10*(1), 57–71. doi:10.1002/jrsm.1332

Roozen, I., & Claeys, C. (2010). The relative effectiveness of celebrity endorsement for print advertisement. *Review of Business and Economic Literature, 55*(1), 76–90.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. doi:10.1037/0033-2909.86.3.638

Schimmelpfennig, C., & Hunt, J. B. (2020). Fifty years of celebrity endorser research: Support for a comprehensive celebrity endorsement strategy framework. *Psychology & Marketing, 37*(3), 488–505. doi:10.1002/mar.21315

Schouten, A. P., Janssen, L., & Verspaget, M. (2020). Celebrity vs. influencer endorsements in advertising: The role of identification, credibility, and product-endorser fit. *International Journal of Advertising, 39*(2), 258–281. doi:10.1080/02650487.2019.1634898

Sedlmeier, P., & Gigerenzer, G. (1992). Do studies of statistical power have an effect on the power of studies? *Psychology Bulletin, 105*(2), 389–406. doi:10.1037/10109-032

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. doi:10.1177/0956797611417632

Simmons, J. P., & Simonsohn, U. (2017). Power posing: P-curving the evidence. *Psychological Science, 28*(5), 687–693. doi:10.1177/0956797616658563

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. doi:10.1177/1745691614553988

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. doi:10.1037/a0033242

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology. General, 144*(6), 1146–1152. doi:10.1037/xge0000104

Spry, A., Pappu, R., & Cornwell, T. B. (2011). Celebrity endorsement, brand credibility and brand equity. *European Journal of Marketing, 45*(6), 882–909. doi:10.1108/03090561111119958

Świątkowski, W., & Dompnier, B. (2017). Replicability crisis in social psychology: Looking at the past to find new pathways for the future. *International Review of Social Psychology, 30*(1), 111–124. doi:10.5334/irsp.66

Till, B. D., Stanley, S. M., & Priluck, R. (2008). Classical conditioning and celebrity endorsers: An examination of belongingness and resistance to extinction. *Psychology & Marketing, 25*(2), 179–196. doi:10.1002/mar.20205

Vakratsas, D., & Ambler, T. (1999). How advertising works: What do we really know? *Journal of Marketing, 63*(1), 26–43. doi:10.2307/1251999

Wei, P. S., & Lu, H. P. (2013). An examination of the celebrity endorsements and online customer reviews influence female consumers' shopping behavior. *Computers in Human Behavior, 29*(1), 193–201. doi:10.1016/j.chb.2012.08.005

Wheeler, R. T. (2009). Nonprofit advertising: Impact of celebrity connection, involvement and gender on source credibility and intention to volunteer time or donate money. *Journal of Nonprofit & Public Sector Marketing, 21*(1), 80–107. doi:10.1080/10495140802111984

Yoon, D., & Kim, Y. K. (2016). Effects of self-congruity and source credibility on consumer responses to coffeehouse advertising. *Journal of Hospitality Marketing & Management, 25*(2), 167–196. doi:10.1080/19368623.2014.1001932

Young, K., & Miller, Y. D. (2015). Keeping it natural: does persuasive magazine content have an effect on young women's intentions for birth? *Women & Health, 55*(4), 447–466. doi:10.1080/03630242.2015.1022690