TITLE: Revisiting the Power Pose Effect: How Robust Are the Results Reported by Carney, Cuddy and Yap (2010) to Data Analytic Decisions?

Marcus Credé

Department of Psychology

Iowa State University

Leigh A. Phillips

Department of Psychology

Iowa State University

ACCEPTED FOR PUBLICATION AT: SOCIAL PSYCHOLOGICAL AND PERSONALITY SCIENCE

Author Bios:

Marcus Credé is an Assistant Professor of Psychology at Iowa State University. His research focuses on non-cognitive predictors of performance in academic and work settings, the measurement of these variables and threats to the validity of statistical inference.

Leigh A. Phillips is an Assistant Professor of Psychology at Iowa State University. Her research focuses on the processes by which individuals develop and maintain healthy habits as well as the statistical methods used to conduct research in this domain.

Contact Information: Marcus Crede, W271 Lagomarcino Hall, Department of Psychology, Iowa State University, Ames, Iowa, 50010, e-mail: mcrede@iastate.edu

Abstract

The literature on the impact of expansive poses on biological and psychological variables is characterized by discrepant findings. These discrepant findings may, in part, be a function of differences in how data were analyzed. In this paper we use multiverse analysis to examine whether the findings reported in the original paper by Carney, Cuddy, and Yap (2010) are robust to plausible alternative data analytic specifications: outlier identification strategy; the specification of the dependent variable; and the use of control variables. Our findings indicate that the inferences regarding the presence and size of an effect on testosterone and cortisol are highly sensitive to data analytic specifications. We encourage researchers to routinely explore the influence of data analytic choices on statistical inferences and also encourage editors and reviewers to require explicit examinations of the influence of alternative data analytic specifications on the inferences that are drawn from data.

KEYWORDS: expansive pose, power pose, p-hacking, multiverse analysis, researcher degrees of freedom.

The claim of a positive impact of expansive body poses – often referred to as power poses - rests to a non-trivial degree on the widely cited and publicly well-known study described by Carney, Cuddy, and Yap (2010). These authors reported that participants who held a high-power pose experienced a significant increase in testosterone and a significant decrease in cortisol relative to participants who held a contractive (i.e., low power) pose. Carney et al. also reported that participants in the high-power pose condition were significantly more likely to engage in risky decision-making, and that they felt significantly more powerful and in-charge than participants in the low-power pose condition; two findings that they reported replicating in a second sample. This finding has been described in the second most viewed TED talk of all time (Cuddy, 2012) as well as in a best-selling book (Cuddy, 2015).

The claim that power poses hold these benefits has attracted substantial controversy in recent years because of failed attempts to replicate the findings reported by Carney et al. (2010) (e.g., Garrison, Tang, & Schmeichel, 2016; Ranehill et al., 2015). In response, proponents of the efficacy of power poses have argued that situational and methodological moderators may account for the inability of some researchers to replicate the initial power pose findings (see Carney, Cuddy & Yap, 2015). An alternative reason for this non-replication that we explore in this paper, is that the findings originally reported by Carney et al. (2010) are the result of p-hacking (Simmons, Nelson, & Simonsohn, 2011).

Broadly speaking, p-hacking refers to the fact that researchers have substantial decision latitude about how statistical analyses are conducted, and that the reported analytic approach is simply the one that resulted in an effect size estimate or inferential statistic that is most favorable for the research hypothesis. For example, researchers may decide to add participants until a desired level of statistical significance is reached for a particular inferential statistic and/or to add

or remove covariates from an analysis in order to maximize the observed effect size estimate or degree of statistical significance. Across different types of statistical analyses, researchers are asked to make many discrete decisions about how analyses are conducted, and for many of these decisions there is no universally acknowledged best practice. These arbitrary decisions include the manner in which missing data is treated, the identification and exclusion procedures for outliers, the decision to transform variables and the use of specific transformation procedures, the scoring procedures for inventories, the screening of data using attention-check indicators, the use and choice of covariates, and the reliance on specific estimation procedures. Because many of these decisions are independent of each other, the total number of possible analysis permutations for any one analysis can be very large.

Consider a simple case in which a researcher must decide between two strategies for dealing with missing data, three strategies for handling outliers, two strategies for dealing with careless-responders, three potential control variables (A, B, and C) that can be used in eight different possible combinations (i.e., no controls, A, B, C, AB, AC, BC, ABC), and two different statistical models. In this simple case, there are 192 (2 x 2 x 3 x 8 x 2) different ways of analyzing the data; a number that can rise even further if the many other analytic decisions that characterize many studies are also considered.

Simmons, Nelson and Simonsohn (2011) note that journals' preferences for statistically significant results and researchers' self-serving bias is likely to result in researchers presenting only the results from the most favorable constellation of data analytic decisions. Furthermore, Gelman and Loken (2013) argued that data analytic alternatives are problematic even when researchers do not engage in post-hoc exploration of different statistical decisions. That is, even in the case in which researchers may have decided a priori upon a particular data analytic

approach (and even when this approach was pre-registered), the inference drawn about the existence and size of an effect based on that particular approach represents only one inference from the set of possible inferences that might be arrived at using other plausible data analytic approaches. Of course, inferences about psychological phenomena should ideally be robust and insensitive to arbitrary data analytic decisions.

The Current Study

In order to explore the degree to which the findings reported by Carney et al. (2010) are sensitive to data analytic decisions, we use the publicly posted data from the primary Carney et al. study (see Fosse, 2016) and a recently developed approach to understanding the robustness of inferences to data analytic decisions. This approach, referred to as a "multiverse analysis" (Steegen, Tuerlinchx, Gelman, & Vanpaemel, 2016) simply provides a summary of the effect size estimates and associated p-values used in a null hypothesis testing approach across all plausible combinations of data analytic decisions. We begin by describing three data analytic decisions that needed to be made when analyzing the data collected by Carney et al. (2010).

Decision 1: Identification of Outliers

Carney et al. (2010) excluded three participants out of a total of 42 participants on the basis of a univariate outlier analysis. Specifically, participants whose cortisol or testosterone scores were more than three standard deviations above or below the sample mean were excluded. At least three alternative strategies are also plausible. First, the authors could have decided to include all observations, because all hormone data was collected via saliva samples and were therefore unlikely to be characterized by response errors such as random responding. Second, the authors could have identified univariate outliers for testosterone by first conditioning on gender,

as is recommended by some endocrinology researchers (e.g., Stanton, 2011), because testosterone exhibits very large gender differences and is produced differently in men and women. That is, separate means and standard deviations are computed for men and women, and outliers are identified using these gender specific means and standard deviations. Third, the authors could have identified testosterone and cortisol outliers using a multivariate criterion such as Mahalanobis Distances, a strategy that is sometimes recommended for ANOVA/ANCOVA analyses over using univariate outlier analysis (e.g., Burdenski, 2000).

Decision 2: Choice of Dependent Variable

Carney et al. (2010) rely on an ANCOVA model, with the post-manipulation hormone (testosterone or cortisol) level as the dependent variable and the pre-manipulation hormone level as a control variable. An alternative strategy – one followed by the unsuccessful replication attempt by Ranehill et al. (2015) - is to use the *change* in the hormone from pre-manipulation to post-manipulation as the dependent variable. Importantly, these two analytic approaches can result in different inferences (a phenomenon known as "Lord's Paradox"; Lord, 1967), because the two approaches answer subtly different questions. The first approach examines the effect of the power pose on post-manipulation hormones that is not explained by pre-manipulation hormones, while the second approach provides information on the influence of the power pose on the change in hormone levels.

Decision 3: Use of Control Variables

For the analysis involving testosterone as the dependent variable, the ANCOVA model relied on by Carney et al. (2010) included gender, pre-manipulation testosterone, pre-manipulation cortisol, and post-manipulation cortisol as control variables. Similarly, for the

analysis involving cortisol as the dependent variable, Carney et al. included gender, pre-manipulation cortisol, pre-manipulation testosterone, and post-manipulation testosterone as control variables. However, there are multiple alternative configurations of control variables that are also plausible—their precise configuration being partly determined by the choice of dependent variables. For example, when the dependent variable is the post-manipulation hormone level, the corresponding pre-manipulation hormone should always be included as a covariate to maximize statistical power and to avoid the confound of pre-existing differences in the hormone; however, whether to include or exclude pre-manipulation and post-manipulation levels of the other hormone is more ambiguous. Similarly, when the dependent variable is the change in hormone levels from pre-manipulation to post-manipulation, the pre-manipulation level of that hormone should not be included as a covariate. The reason for this is that controlling for one part of a change scores reduces the dependent variable simply to the other part of the change scores (Edwards, 2001). Finally, gender should be included as a covariate in this study, unless separate effect size estimates are computed for men and women – a strategy that is recommended for testosterone data by Stanton (2011) in his critique of the data analytic approach taken by Carney et al. (2010). Stanton argues for the separate computation of effects for men and women, because the biological generating mechanism for testosterone is different for men and women, and because women are well-known to have a different hormonal response than men in dominance situations (e.g., power posing).

The analyses described by Carney et al. (2010) involved other discrete decisions – such as the question of whether gender by treatment interactions should be examined (see Stanton, 2011, for a discussion); for the sake of simplicity, we limit our discussion and the multiverse analysis to the three aforementioned analytic decisions. Together the three analytic decisions

result in 36 potential data analytic configurations. The raw data for the findings described by Carney et al. has recently been made public (Fosse, 2016), thereby allowing an examination of how the effect size estimate for the power pose manipulation is dependent on the precise configuration of analytic decisions.

Results

For the sake of simplicity, we present two statistics for our multiverse analyses: (1) partial eta squared, as an estimate of the power pose treatment effect size, and (2) the associated p-values that researchers using a null hypothesis testing framework would rely on to arrive at inferences. We interpret the effect sizes using the standards for eta-squared proposed by Miles and Shevlin (2001): 0.01 is small, 0.06 is medium, and 0.14 is large.

Outlier Analysis

Carney et al. (2010) reported excluding three participants from the analyses because at least one of their hormone levels was more than three standard deviations above the sample mean. When separate means and standard deviations on testosterone are calculated for men and women, as recommended by Stanton (2011), only one univariate outlier is present in the data. An examination of Mahalanobis distances indicate no evidence for multivariate outliers (at alpha = .001) across the four hormone measurements, when data for men and women are examined separately. For this specific sample the four possible ways of identifying outliers that we discussed earlier therefore only result in three different sample sizes. We therefore present our multiverse analyses in a way that includes only discrete specifications for the outlier analysis: one excluding the outliers as in Carney et al. (2010), resulting in a sample size of 39, one that

identifies outliers based on gender-conditioned testosterone values resulting in a sample size of 41, and one that does not exclude any participants (i.e., N=42).

Multiverse Analysis for Testosterone

The multiverse analysis for testosterone (Table 1) provides 54 different estimates of the effect of the power pose treatment on testosterone, although 16 of these are redundant with other analyses due to invariant sample sizes for women across the three outlier identification specifications. Each of these estimates is the result of an ANCOVA (with high-power, low-power pose manipulation as the IV) with different combinations of the following analysis decisions: 1) determining gender specific/nonspecific outliers on single or total variables (i.e. all analyses had 39, 41, or 42 participants depending on identification of outliers); 2) defining the dependent variable as T2 hormone level or as change in hormone level from T1 to T2; 3) choosing control variables; and 4) handling gender (i.e. gender separate or combined analysis). In aggregate, our analyses indicate very substantial variability in the partial eta squared effect size estimate for the power pose manipulation. Indeed, partial eta squared estimates range from zero to a large effect of 0.192 and substantial variability in effect size estimates is evident across all three specification levels. That is, the choice of outlier analysis, the choice of dependent variable, and the choice of control variables all exhibited substantial influences on effect size estimates. Particularly low effect size estimates are evident for most of the analyses involving only female participants. Similar variability is also evident in the p-values associated with the effect sizes, such that researchers relying on a null hypothesis significance testing framework would make very different inferences about both the size and presence of a power-posing effect on testosterone. It is also noteworthy that all alternative effect size estimates for the total sample (N=42) are smaller than the effect size reported for testosterone by Carney et al. (2010). In many

instances the alternative estimates were substantially smaller. That is, the multiverse analysis shows that the reported effect size is not robust to data analytic decisions.

Multiverse Analysis for Cortisol

The multiverse analysis for cortisol (Table 2) also provides 54 estimates (38 unique) of the effect of the power pose treatment on cortisol. As in the analysis of testosterone, each of these estimates represents a different way of 1) identifying outliers, 2) defining the dependent variable, 3) combination of control variables, and 4) method of controlling for gender effects. In aggregate, this analysis indicates similarly high levels of variability as the testosterone analyses in both the effect size estimates and the associated p-values. Partial eta squared estimates ranged from zero to a large effect of 0.239.

As was the case for testosterone, all alternative effect size estimates for the total sample (N=42) are smaller than the effect size reported for cortisol by Carney et al. (2010), and in many instances the alternative estimates were substantially smaller.

Follow-Up Analysis

Although all of the examined analytic decisions influenced the inferences made about the effect of power-posing, perhaps the largest influence was evident in the separate analyses for men and women. In general, the estimates of the effect of the power-pose manipulation on testosterone was very strong for men and near zero for women, while the effect on cortisol was relatively strong for women and much weaker for men. In order to examine whether this gender effect was also evident for the behavioral and perceptual dependent variables examined by Carney et al. (2010), we re-analyzed this aspect of the data made available by Fosse (2016). In addition to the effects on hormones that we have already discussed, Carney et al. reported that

participants in the high-power pose condition were significantly more likely to engage in risky decision-making, and that they felt significantly more powerful and in-charge than participants in the low-power pose condition; two findings that they reported replicating in a second sample.

However, a re-analysis of both data sets[1] shows that the effect of the power-pose manipulation on risk-taking was indeed much stronger for *men* in both samples (χ^2 (1, n=16)= 4.75, p= .029, r= .54 in sample 1 and χ^2(1, n= 20)= 6.11, p= .013, r=.55 in sample 2) than for women (χ^2(1, n=26) = 0.72, r= -.17 in sample 1 and χ^2(1, n=29)= 0.68, p>.25, r= .15 in sample 2). A comparison of these effects using Cochran's test supported a significant difference (χ^2 (1, n=42)=3.64, p=.056 in sample 1) and χ^2 (1, n=49)=4.84, p=.03 in sample 2). The gender moderating effect was also evident for feelings of being in power and in charge: men in the high power-pose condition felt more powerful and in charge than men in the low power-pose condition (t(14)= 3.86, p=.002, Cohen's d=1.95 for sample 1, and t(18)=1.78, p=.09, Cohen's d=.80 for sample 2), while the same effect for women was much weaker (t(24)=1.16, p>.25, Cohen's d= .45 for sample 1, and t(27)= 1.21, p=.24, Cohen's d= .19 for sample 2).

Further, a re-analysis of the data from Ranehill et al. (2015)–who failed to replicate most of the power pose effects reported by Carney et al. (2011)–shows a similar pattern for feelings of power and confidence; effects were much higher for men (t(79)= 3.41, p=.001, Cohen's d= 0.76) than for women (t(78)= 0.76, p> .25, Cohen's d= 0.17). A meta-analysis of the effects across the three studies shows that the effect for men (k=3, N=117, Cohen's d= .87) was significantly stronger (Z=2.45, p=.01) than the effect for women (k=3, N=135, Cohen's d = .22). This evidence for the moderating role of gender is particularly noteworthy, because power poses have often been emphasized as being effective for women (Cuddy, 2012; 2013). For example, in a discussion of how power posing can help women "lean in" in business settings Cuddy (2013)

argued that: "….standing in a bathroom stall like Wonder Woman before a stressful meeting — has the potential to substantially improve women's ability to lean in – to take risks, face fears and barriers, and to endure the stressors inherent to the kinds of changes Sandberg recommends".

Discussion

Our paper makes two broad contributions to the literature. First, we have illustrated that the original findings regarding the benefits of expansive, "power poses" are highly sensitive to the specific configuration of plausible data analytic choices made by the researchers and that those reported by Carney et al. (2010) and discussed in Cuddy (2012) were the strongest effects of all possible effects, a majority of which were small effects or near-zero effects, and strongly moderated by gender. As such, our findings should help to clarify the apparent discrepancies between the original findings and subsequent unsuccessful efforts to replicate these findings. That is, our results suggest that the data described by Carney et al. (2010), like the data from various unsuccessful replication attempts, are not supportive of a robust effect for power poses. It should, of course, also be noted that some of the authors who reported a failure to replicate the power pose effect also only presented findings for one particular configuration of data analytic choices and that these configurations did not necessarily match those used by Carney et al. For example, Ranehill et al. (2015) did not exclude outliers, did not include the covariates used by Carney et al., and used changes in hormone levels as the dependent variable rather than post-manipulation hormone levels controlling for pre-manipulation hormones. As such it is possible that some of the failed replication attempts selectively presented results for the least favorable combination of data analytic choices. However, our own re-analysis of the data reported by Ranehill et al. (reported in Supplementary Material) suggests that the failure to replicate the power pose in that data is robust to the types of data analytic choices described in this paper.

Our second contribution is broader in nature and relates to the manner in which we hope researchers will explore their data and present their findings. First, we hope that researchers will become better aware of how data analytic choices can dramatically influence the inferences they draw from their findings. Popper (1963) and others, such as Feynman (1974) and Greenwald (1986), encourage us to not only report but to seek out disconfirming evidence in order to advance our field as rapidly as possible. Examining the role of analytic choices on our statistical inferences represents one way of seeking out and presenting such potentially disconfirming evidence. As such, our findings may encourage greater caution in how findings are interpreted, how they are integrated into our current understanding, and how they are used as the building blocks for future research. Of course, this requires a greater willingness on the part of journals to publish results that are not entirely robust to all analytic approaches. Second, we hope that journals will allow researchers to report multiverse analyses in order to illustrate to readers how robust or sensitive findings are to data analytic strategy. Explicitly modeling the effect of data analytic choices should ameliorate reader's concerns about possible p-hacking and thereby increase the faith that readers have in reported findings. This is likely to be particularly important at a time when we are becoming increasingly aware of the damaging effect that questionable research practices such as p-hacking and hypothesizing-after-results-known (Bosco, Aguinis, Field, Pierce, & Dalton, 2015; Kerr, 1998) and the metamorphosis of reported findings between dissertation and journal versions of the same data (the Chrysalis Effect; O'Boyle, Banks, & Gonzalez-Mule, 2014) have on the credibility of our discipline (see also John, Loewenstein & Prelec, 2012). To this end, we would also encourage reviewers and editors to not only request multiverse analyses from authors, but to also think about plausible alternative analytic strategies that the authors may not have considered. Lastly, we hope that systematic efforts to model the

effect of analytic choices may, at times, also yield meaningful theoretical insights, as we attempted to illustrate with our example on the moderating role of gender on the impact of power posing interventions on psychological variables such as risk-taking behaviors and feelings of being powerful and “in-charge”.

References

Banks, G.C., O'Boyle Jr., E.H., Pollack, J.M., White, C.D., Batchelor, J.H., Whelpley, C.E., Abston, K.A., Bennett, A.A., & Adkins, C.L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management, 42,* 5-20.

Bosco, F.A., Aguinis, H., Field, J.G., Pierce, C.A., & Dalton, D.R. (2015). Harking's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology*. Advance online publication. DOI: 10.1111/peps.12111

Carney, D. R., Cuddy, A. J. C., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science, 21,* 1363–1368. doi:10.1177/0956797610383437

Carney, D.R., Cuddy, A.J.C., & Yap, AY. (2015). Review and summary of research on the embodied effects of expansive (versus contractive) nonverbal displays. *Psychological Science, 26,* 657-663.

Cuddy, A. (June 2012). Your body language shapes who you are. [Video file]. Retrieved from http://www.ted.com/talks/amy_cuddy_your_body_language_shapes_who_you_are

Cuddy, A.C.J. (2013). Want to lean in? Try a power pose. *Harvard Business Review*. Retrieved online from: https://hbr.org/2013/03/want-to-lean-in-try-a-power-po-2

Cuddy, A.C.J. (2015). *Presence: Bringing Your Boldest Self to Your Biggest Challenges*. New York: Little, Brown & Company.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4,* 265–287.

Fosse, N. E. (2016). Replication Data for "Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance" by Carney, Cuddy, Yap (2010). Harvard Dataverse. Retrieved from http://dx.doi.org/10.7910/DVN/FMEGS6

Garrison, K. E., Tang, D., & Schmeichel, B. J. (2016). Embodying power: A preregistered replication and extension of the power pose effect. *Social Psychological and Personality Science, 7*, 623-630.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no fishing expedition or "p-hacking" and the research hypothesis was posited ahead of time. Technical report, Department of Statistics, Columbia University. Retrieved January 9, 2016, from www.stat. columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Greenwald, A.G., Pratkanis, A.R., Leippe, M.R., & Baumgardner, M. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93*, 216-229.

John L.K., Loewenstein G., & Prelec D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524–532.

Kerr, N.L. (1998). HARKing: Hypothesizing after the results are known**.** *Personality and Social Psychology Review, 2*, 196-217.

Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304-305.

Miles, J. and Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers*. Sage: London.

O'Boyle Jr., E.H., Banks, G.C., & Gonzalez-Mule, E. (2014). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management.* Advance online publication, doi: 10.1177/0149206314527133

Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge: London, England.

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R.A. (2015). Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. *Psychological Science*, *26*, 653-656. doi: 0956797614553946

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***,** 1359–1366.

Stanton, S.J. (2011). The essential implication of gender in human behavioral endocrinology studies. Frontiers in Behavioral Neuroscience, 5(9), doi: 10.3389/fnbeh.2011.00009

Steegen, S., Tuerlinchx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through multiverse analysis. *Perspectives on Psychological Science, 11*, 702-715.

Footnotes

1. The data for sample 2 was provided by Carney via personal communication with the first author.

Table 1: Multiverse Analysis for the Effect of Power-Posing on Testosterone

		Outlier Identification: Entire Sample (N=39)		Outlier Identification: Test. Conditioned on Gender (N=41)		Outlier Identification: Multivariate or No-Exclusion (N=42)	
Gender Effect	Control Variables	DV: T2 Test.	DV: Δ in Test.	DV: T2 Test.	DV: Δ in Test.	DV: T2 Test.	DV: Δ in Test.
Combined	Gender		.047 (p= .19)		.019 (p=.39)		.036 (p=.23)
Combined	Gender, T1 Test.	.029 (p=.31)		.042 (p=.21)		.055 (p=.15)	
Combined	Gender, T1 Cort.		.045 (p= .21)		.017 (p=.43)		.018 (p=.42)
Combined	Gender, T1 Test., T1 Cort.	.037 (p=.26)		.040 (p=.23)		.043 (p= .21)	
Combined	T1 Cort., T2 Cort.		.089 (p=.07)		.038 (p=.23)		.037 (p=.24)
Combined	Gender, T1 Test., T1 Cort., T2 Cort.	**.123 (p=.039)**		.099 (p=.06)		.102 (p=.051)	
Men Only	No Controls		.192 (p=.13)		.047 (p=.44)		.096 (p=.24)
Men Only	T1 Test.	.000 (p=.96)		.073 (p=.35)		.101 (p=.25)	
Men Only	T1 Cort.		.184 (p=.17)		.121 (p=.22)		.063 (p=.37)
Men Only	T1 Test., T1 Cort.	.026 (p=.64)		.104 (p=.28)		.083 (p=.32)	
Men Only	T1 Cort., T2 Cort.		.162 (p=.22)		.141 (p=.21)		.057 (p=.41)
Men Only	T1 Test., T1 Cort., T2 Cort.	026 (p=.657)		.125 (p= .26)		.086 (p= .33)	
Women Only	No Controls		.005 (p=.73)		.005 (p=.73)		.005 (p=.73)
Women Only	T1 Test.	.019 (p=.51)		.019 (p=.51)		.019 (p=.51)	
Women Only	T1 Cort.		.005 (p=.75)		.005 (p=.75)		.005 (p=.75)
Women Only	T1 Test., T1 Cort.	.023 (p= .48)		.023 (p= .48)		.023 (p= .48)	
Women Only	T1 Cort., T2 Cort.		.077 (p=.19)		.077 (p=.19)		.077 (p=.19)
Women Only	T1 Test., T1 Cort., T2 Cort.	.167 (p=.053)		.167 (p=.053)		.167 (p=.053)	

Note: Test. = Testosterone, Cort. = Cortisol, T1= Pre-Manipulation, T2= Post-Manipulation, Entries are partial eta-squared values and (in parentheses) the associated p-value. The entry in bold is the effect for the analyses originally reported in the Carney et al. (2010) paper. Blank entries mean that the analyses would not be recommended for reasons described in the text. The number of women was constant across the three outlier strategies.

Table 2: Multiverse Analysis for the Effect of Power-Posing on Cortisol

		Outlier identification: Entire Sample (N=39)		Outlier Identification: Test. Conditioned on Gender (N=41)		Outlier identification: Multivariate or Conditioned on Gender (N=42)	
Gender Effect	Control Variables	DV: T2 Cort.	DV: Δ in Cort.	DV: T2 Cort.	DV: Δ in Cort.	DV: T2 Cort.	DV: Δ in Cort.
Combined	Gender		.004 (p=.71)		.013 (p=.47)		.002 (p=.79)
Combined	Gender, T1 Cort.	.08 (p=.09)		.079 (p=.08)		.061 (p=.12)	
Combined	Gender, T1 Test.		.007 (p=.63)		.007 (p=.62)		.003 (p=.75)
Combined	Gender, T1 Test., T1 Cort.	.073 (p=.11)		.087 (p=.07)		.078 (p=.09)	
Combined	T1 Test., T2 Test.		.011 (p=.54)		.022 (p=.37)		.000 (p=.90)
Combined	Gender, T1 Cort., T1 Test., T2 Test.	**.155 (p=.02)**		.129 (p=.03)		.135 (p=.02)	
Men Only	No Controls		.014 (p=.70)		.061 (p=.37)		.019 (p=.61)
Men Only	T1 Cort.	.044 (p=.51)		.103 (p=.26)		.019 (p=.62)	
Men Only	T1 Test.		.015 (p=.71)		.124 (p=.21)		.008 (p=.76)
Men Only	T1 Cort., T1 Test.	.000 (p=.98)		.101 (p=.29)		.027 (p=.58)	
Men Only	T1 Test., T2 Test.		.015 (p=.72)		.178 (p=.15)		.000 (p=.98)
Men Only	T1 Cort., T1 Test., T2 Test.	.000 (p=.97)		.122 (p=.27)		.111 (p=.04)	
Women Only	No Controls		.003 (p=.79)		.003 (p=.79)		.003 (p=.79)
Women Only	T1 Cort.	.094 (p=.14)		.094 (p=.14)		.094 (p=.14)	
Women Only	T1 Test.		.000 (p=.95)		.000 (p=.95)		.000 (p=.95)
Women Only	T1 Cort., T1 Test.	.108 (p=.12)		.108 (p=.12)		.108 (p=.12)	
Women Only	T1 Test., T2 Test.		.001 (p=.90)		.001 (p=.90)		.001 (p=.90)
Women Only	T1 Cort., T1 Test., T2 Test.	.239 (p=.02)		.239 (p=.02)		.239 (p=.02)	

Note: T1 Test. = Testosterone, Cort. = Cortisol, T1= Pre-Manipulation, T2= Post-Manipulation, Entries are partial eta-squared values and (in parentheses) the associated p-value. The entry in bold is the effect for the analyses originally reported in the Carney et al. (2010) paper. Blank entries mean that the analyses would not be recommended for reasons described in the text. The number of women was constant across the three outlier strategies.

Table S1: Data Analytic Robustness Analysis for the Effect of Power-Posing on Testosterone in Data Reported By Ranehill et al. (2015).

		Outlier Identification: Full Sample (N=198, 102 Males)		Outlier Identification: Multivariate Outliers, Conditioned on Gender (N=190, 101 Males)	
Gender Effect	Control Variables	DV: T2 Testosterone	DV: Δ in Testosterone	DV: T2 Testosterone	DV: Δ in Testosterone
Combined	Gender	N/A	.008 (p=.205)	N/A	.002 (p=.510)
Combined	Gender, T1 Testosterone	.008 (p=.224)	N/A	.002 (p=.517)	N/A
Combined	Gender, T1 Cortisol	N/A	.008 (p=.223)	N/A	.002 (p=.509)
Combined	Gender, T1 Testosterone, T1 Cortisol	.007 (p=.230)	N/A	.003 (p=.495)	N/A
Combined	T1 Cortisol, T2 Cortisol	N/A	.005 (p=.322)	N/A	.004 (p=.413)
Combined	Gender, T1 Testosterone, T1 Cortisol, T2 Cortisol	.004 (p=.362)	N/A	.002 (p=.503)	N/A
Men Only	No Controls	N/A	.010 (p=.328)	N/A	.003 (p=.568)
Men Only	T1 Testosterone	.009 (p=.338)	N/A	.003 (p=.573)	N/A
Men Only	T1 Cortisol	N/A	.008 (p=.386)	N/A	.003 (p=.567)
Men Only	T1 Testosterone, T1 Cortisol	.008 (p=.375)	N/A	.004 (p=.554)	N/A
Men Only	T1 Cortisol, T2 Cortisol	N/A	.006 (p=.430)	N/A	.003 (p=.599)
Men Only	T1 Testosterone, T1 Cortisol, T2 Cortisol	.007 (p=.423)	N/A	.003 (p=.591)	N/A
Women Only	No Controls	N/A	.007 (p=.418)	N/A	.002 (p=.688)
Women Only	T1 Testosterone	.004 (p=.559)	N/A	.004 (p=.579)	N/A
Women Only	T1 Cortisol	N/A	.007 (p=.421)	N/A	.002 (p=.709)
Women Only	T1 Testosterone, T1 Cortisol	.003 (p=.579)	N/A	.003 (p=.597)	N/A
Women Only	T1 Cortisol, T2 Cortisol	N/A	.001 (p=.781)	N/A	.003 (p=.611)
Women Only	T1 Testosterone, T1 Cortisol, T2 Cortisol	.000 (p=.838)	N/A	.003 (p=.625)	N/A

Note: T1: Pre-Manipulation, T2: Post-Manipulation, Entries are partial eta-squared values and (in parentheses) the associated p-value.

Table S1 Continued: Data Analytic Robustness Analysis for the Effect of Power-Posing on Testosterone in Data Reported By Ranehill et al. (2015).

		Outlier Identification: Univariate Outliers, not Conditioned on Gender (N=187, 95 Males)		Outlier Identification: Univariate Outliers, Testosterone Conditioned on Gender (N=186, 96 Males)	
Gender Effect	Control Variables	DV: T2 Testosterone	DV: Δ in Testosterone	DV: T2 Testosterone	DV: Δ in Testosterone
Combined	Gender	N/A	.004 (p=.412)	N/A	.005 (p=.340)
Combined	Gender, T1 Testosterone	.002 (p=.525)	N/A	.005 (p=.354)	N/A
Combined	Gender, T1 Cortisol	N/A	.003 (p=.425)	N/A	.005 (p=.353)
Combined	Gender, T1 Testosterone, T1 Cortisol	.002 (p=.501)	N/A	.005 (p=.342)	N/A
Combined	T1 Cortisol, T2 Cortisol	N/A	.003 (p=.499)	N/A	.003 (p=.466)
Combined	Gender, T1 Testosterone, T1 Cortisol, T2 Cortisol	.001 (p=.677)	N/A	.002 (p=.539)	N/A
Men Only	No Controls	N/A	.006 (p=.452)	N/A	.008 (p=.378)
Men Only	T1 Testosterone	.005 (p=.496)	N/A	.008 (p=.382)	N/A
Men Only	T1 Cortisol	N/A	.006 (p=.465)	N/A	.008 (p=.388)
Men Only	T1 Testosterone, T1 Cortisol	.006 (p=.476)	N/A	.009 (p=.368)	N/A
Men Only	T1 Cortisol, T2 Cortisol	N/A	.002 (p=.648)	N/A	.003 (p=.579)
Men Only	T1 Testosterone, T1 Cortisol, T2 Cortisol	.002 (p=.653)	N/A	.004 (p=.554)	N/A
Women Only	No Controls	N/A	.001 (p=.746)	N/A	.002 (p=.707)
Women Only	T1 Testosterone	.001 (p=.881)	N/A	.003 (p=.599)	N/A
Women Only	T1 Cortisol	N/A	.001 (p=.754)	N/A	.001 (p=.728)
Women Only	T1 Testosterone, T1 Cortisol	.001 (p=.808)	N/A	.003 (p=.616)	N/A
Women Only	T1 Cortisol, T2 Cortisol	N/A	.001 (p=.800)	N/A	.000 (p=.853)
Women Only	T1 Testosterone, T1 Cortisol, T2 Cortisol	.004 (p=.564)	N/A	.000 (p=.880)	N/A

Note: T1: Pre-Manipulation, T2: Post-Manipulation, Entries are partial eta-squared values and (in parentheses) the associated p-value.

Table S2: Data Analytic Robustness Analysis for the Effect of Power-Posing on Cortisol in Data Reported By Ranehill et al. (2015).

		Outlier Identification: Full Sample (N=192, 99 Males)		Outlier Identification: Multivariate Outliers, Conditioned on Gender (N=184, 98 Males)	
Gender Effect	Control Variables	DV: T2 Cortisol	DV: Δ in Cortisol	DV: T2 Cortisol	DV: Δ in Cortisol
Combined	Gender	N/A	.004 (p=.364)	N/A	.000 (p=.783)
Combined	Gender, T1 Cortisol	.004 (p=.370)	N/A	.000 (p=.867)	N/A
Combined	Gender, T1 Testosterone	N/A	.003 (p=.424)	N/A	.000 (p=.842)
Combined	Gender, T1 Testosterone, T1 Cortisol	.004 (p=.385)	N/A	.000 (p=.872)	N/A
Combined	T1 Testosterone, T2 Testosterone	N/A	.001 (p=.652)	N/A	.000 (p=.865)
Combined	Gender, T1 Cortisol, T1 Testosterone, T2 Testosterone	.001 (p=.702)	N/A	.000 (p=.921)	N/A
Men Only	No Controls	N/A	.003 (p=.588)	N/A	.001 (p=.774)
Men Only	T1 Cortisol	.001 (p=.727)	N/A	.000 (p=.827)	N/A
Men Only	T1 Testosterone	N/A	.002 (p=.629)	N/A	.001 (p=.793)
Men Only	T1 Cortisol, T1 Testosterone	.001 (p=.706)	N/A	.001 (p=.803)	N/A
Men Only	T1 Testosterone, T2 Testosterone	N/A	.000 (p=.853)	N/A	.000 (p=.887)
Men Only	T1 Cortisol, T1 Testosterone, T2 Testosterone	.000 (p=.967)	N/A	.000 (p=.964)	N/A
Women Only	No Controls	N/A	.008 (p=.404)	N/A	.000 (p=.974)
Women Only	T1 Cortisol	.012 (p=.296)	N/A	.000 (p=.897)	N/A
Women Only	T1 Testosterone	N/A	.008 (p=.410)	N/A	.001 (p=.775)
Women Only	T1 Cortisol, T1 Testosterone	.013 (p=.266)	N/A	.001 (p=.824)	N/A
Women Only	T1 Testosterone, T2 Testosterone	N/A	.004 (p=.545)	N/A	.000 (p=.990)
Women Only	T1 Cortisol, T1 Testosterone, T2 Testosterone	.011 (p=.327)	N/A	.000 (p=.917)	N/A

Note: T1: Pre-Manipulation, T2: Post-Manipulation, Entries are partial eta-squared values and (in parentheses) the associated p-value.

Table S2 Continued: Data Analytic Robustness Analysis for the Effect of Power-Posing on Cortisol in Data Reported By Ranehill et al. (2015).

		Outlier Identification: Univariate Outliers, not Conditioned on Gender (N=181, 92 Males)		Outlier Identification: Univariate Outliers, Testosterone Conditioned on Gender (N=180, 93 Males)	
Gender Effect	Control Variables	DV: T2 Cortisol	DV: Δ in Cortisol	DV: T2 Cortisol	DV: Δ in Cortisol
Combined	Gender	N/A	.004 (p=.387)	N/A	.006 (p=.289)
Combined	Gender, T1 Cortisol	.003 (p=.458)	N/A	.005 (p=.351)	N/A
Combined	Gender, T1 Testosterone	N/A	.003 (p=.481)	N/A	.006 (p=.325)
Combined	Gender, T1 Testosterone, T1 Cortisol	.003 (p=.489)	N/A	.005 (p=.356)	N/A
Combined	T1 Testosterone, T2 Testosterone	N/A	.003 (p=.501)	N/A	.004 (p=.409)
Combined	Gender, T1 Cortisol, T1 Testosterone, T2 Testosterone	.001 (p=.655)	N/A	.002 (p=.567)	N/A
Men Only	No Controls	N/A	.008 (p=.386)	N/A	.010 (p=.346)
Men Only	T1 Cortisol	.006 (p=.460)	N/A	.007 (p=.407)	N/A
Men Only	T1 Testosterone	N/A	.007 (p=.443)	N/A	.009 (p=.362)
Men Only	T1 Cortisol, T1 Testosterone	.006 (p=.464)	N/A	.008 (p=.392)	N/A
Men Only	T1 Testosterone, T2 Testosterone	N/A	.005 (p=.528)	N/A	.005 (p=.489)
Men Only	T1 Cortisol, T1 Testosterone, T2 Testosterone	.003 (p=.631)	N/A	.003 (p=.603)	N/A
Women Only	No Controls	N/A	.001 (p=.807)	N/A	.003 (p=.630)
Women Only	T1 Cortisol	.000 (p=.844)	N/A	.002 (p=.683)	N/A
Women Only	T1 Testosterone	N/A	.002 (p=.699)	N/A	.006 (p=.469)
Women Only	T1 Cortisol, T1 Testosterone	.002 (p=.686)	N/A	.006 (p=.457)	N/A
Women Only	T1 Testosterone, T2 Testosterone	N/A	.004 (p=.538)	N/A	.003 (p=.594)
Women Only	T1 Cortisol, T1 Testosterone, T2 Testosterone	.005 (p=.509)	N/A	.004 (p=.571)	N/A

Note: T1: Pre-Manipulation, T2: Post-Manipulation, Entries are partial eta-squared values and (in parentheses) the associated p-value.