ARTICLE IN PRESS

JCPS-00489; No. of pages: 11; 4C:

Available online at www.sciencedirect.com

ScienceDirect

Journal of Consumer Psychology xx, x (2015) xxx – xxx

JCP JOURNAL OF CONSUMER PSYCHOLOGY

Research Dialogue

Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power

Gary H. McClelland [a,*], John G. Lynch, Jr. [b], Julie R. Irwin [c], Stephen A. Spiller [d], Gavan J. Fitzsimons [e]

[a] *Department of Psychology and Neuroscience, University of Colorado—Boulder, United States*
[b] *Leeds School of Business, University of Colorado—Boulder, United States*
[c] *McCombs School of Business, University of Texas—Austin, United States*
[d] *Anderson School of Management, UCLA, United States*
[e] *Fuqua School of Business, Duke University, United States*

Received 30 April 2015; accepted 3 May 2015

Abstract

Considerable prior statistical work has criticized replacing a continuously measured variable in a general linear model with a dichotomy based on a median split of that variable. Iacobucci, Posovac, Kardes, Schneider, and Popovich (2015- this issue) defend the practice of "median splits" using both conceptual arguments and simulations. We dispute their conceptual arguments, and we have identified technical errors in their simulations that dramatically change the conclusions that follow from those simulations. We show that there are no real benefits to median splits, and there are real costs in increases in Type II errors through loss of power and increases in Type I errors through false–positive consumer psychology. We conclude that median splits remain a bad idea.

Keywords: Median splits; Statistical power; False–positive psychology

Introduction

Researchers can make Type I or Type II errors, rejecting a true null hypothesis, or failing to reject a false null hypothesis. In the same way, journals can make two kinds of errors, rejecting a paper that is later concluded to be insightful or publishing a paper that is later concluded not to be true. For instance, Gans and Shepherd (1994) reviewed famous economics papers that were rejected multiple times before being published and regarded as great. George Akerlof's (1970) "A Market for Lemons" paper was rejected by the *American Economic Review*, the *Journal of Political Economy*, and the *Review of Economic Studies*. Two said it was trivial, the other that it was too general to be true. Those journals made a Type II error. Ackerlof later won the Nobel Prize in economics for the work. In other cases, a prestigious journal publishes a sensational result that seems too good to be true and is later discredited, reflecting a Type I error. Prominent examples are cold fusion claims by Fleischmann and Pons (1989) and Bem's (2011) finding of correct prediction of events in the future (i.e. ESP). Both were followed by numerous failures to replicate, and in the case of Bem, detailed critiques of the statistical analysis by the editor who had accepted the original paper (Judd, Westfall, & Kenny, 2012).

The paper by Iacobucci, Posovac, Kardes, Schneider, and Popovich (2015- this issue, hereafter IPKSP) may fall within the latter category. These authors make conceptual arguments and present statistical simulations about the consequences of median splits of continuous independent variables in linear models. Later

* Corresponding author at: Dept of Psychology and Neuroscience, 345 UCB, University of Colorado Boulder, Boulder, CO 80309-0345, United States.
E-mail address: gary.mcclelland@colorado.edu (G.H. McClelland).

http://dx.doi.org/10.1016/j.jcps.2015.05.006
1057-7408/

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006

in this commentary, we point out technical errors in their statistical simulations. The actual programming code in Appendix A of IPKSP does not match the description in the text of their paper, and the result is that the simulations do not support the conclusions IPKSP wish to draw. Consequently, the bulk of the contribution of their paper must stand or fall on their conceptual arguments for the appropriateness of median splits, which we argue are often misguided. We first evaluate their conceptual arguments and present conceptual arguments of our own, then present our reanalysis and interpretation of their simulation results.

The topic of categorizing continuous predictor variables by splitting them at their median has been covered extensively, including in our own papers (e.g., Cohen, 1983; DeCoster, Iselin, & Gallucci, 2009; Fitzsimons, 2008; Humphreys, 1978; Humphreys & Fleishman, 1974; Irwin & McClelland, 2003; MacCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993). We know of no statistical argument in favor of median splits to counterbalance the chorus of statistical critiques against them. Because there is a danger that IPKSP may convince researchers to use median splits, we briefly present the arguments against their claims.

Our commentary will proceed as follows. First we will very briefly present the core statistical reasons why median splits are to be avoided. Second, we will review nonstatistical justifications for median splits presented by IPKSP—including the argument that median splits are "conservative"—and will show that there are ready answers for those justifications. Then we will discuss in more depth the statistical considerations for when median splits affect Type II errors, adversely affecting power. In our view, power is the most compelling reason to avoid median splits. We will address the conservatism defense in that section, where we will show that steps that lower the power of reports of significant findings in a journal increase the percent of published results that are Type I errors. Finally, we will address the discrepancies between the actual programming code in IPKSP's Appendix A and the descriptions in the body of IPKSP's paper and show how those discrepancies invalidate the conclusions drawn by IPKSP.

The statistical case against median splits in a nutshell

We highlight the statistical case against median splits in a simple design with a dependent variable Y and a single measured independent variable X. We later consider multiple independent variables in our reanalysis of IPKSP's simulations. Assume X is an indicator of some latent construct and that the observed X is linearly related to the underlying construct. By splitting the measured X at its median, one replaces X with a categorical variable X′ (e.g., 1 = greater than median, 0 = less than or equal to the median). There are four main consequences of this substitution, discussed in detail below:

a. This substitution introduces random error in the measure of the latent construct and all of the problems that adding error brings.
b. The analysis now is insensitive to the pattern of local covariation between X and Y within groups defined by the median split. All that matters is the mean difference.
c. This analysis involves a nonlinear transformation of the original X to a step function of the original X on the dependent variable Y. The use of a median split on X makes it impossible to test a substantive theoretical claim of a step function relation of latent X to dependent variable Y.
d. If one believes that there is a step function relation of latent X to the dependent variable Y, the threshold of that function is presumably general and not sample-dependent. A median split is sample-dependent.

a. Errors in variables

Introducing random error has two interrelated negative consequences. First, when there is a nonzero population correlation between X and Y, the correlation between the median split X′ and Y will be lower in expectation, though adding error can make the correlation higher in a subset of samples. Also, splitting at the median makes the measure of the latent construct underlying X noisier. Expected effect size goes down, and statistical power is a function of effect size.

Adding random error to one's measure of X creates "errors in variables" in regression models, a source of bias in estimated (standardized) coefficients. Since multiple regression models assume errorless measurement of the latent constructs underlying X, adding error via median split creates inconsistent estimates of the standardized coefficient (i.e., estimates that do not have expected value equal to the true parameter). We will demonstrate that this practice is hazardous, not "conservative" as IPKSP maintain. It is surprising to us that Iacobucci, Saldanha, and Deng (2007) have argued so eloquently about the negative consequences of ignoring errors in variables in statistical mediation analysis, but in the current paper IPKSP defend the deliberate adding of measurement error to an independent variable.

b. Ignoring information about local within-group covariation between X and Y

Consider a simple regression of Y on continuously measured X, and a reanalysis of the same data replacing X with X′ defined by a median split. The analysis using median splits is insensitive to the pattern of local covariation between Y and the continuous X within the above-median and below-median groups. The analysis using the continuously measured X is sensitive to that within-group covariation. As a thought experiment, imagine holding constant the univariate distributions of X and Y above and below the median, but scrambling the pairings of X and Y within the subsets of points above and below the median. Different scrambles produce widely different slopes of the regression of Y on continuous X, some significant, some not, but identical slopes of the regression of Y on X′. Thus, it is untrue that it is uniformly conservative to use the median split. In some cases the *t* statistics from the median split can be more significant than the *t* statistics from regressing Y on continuous X, and in most cases less significant. Such inconsistencies could allow unscrupulous researchers to pick whichever outcome was more favorable, as we discuss in more detail later.

c. Nonlinear transformation of X implies a step-function form of the X–Y relation

Median splits produce a nonlinear transformation of continuous X (that is linearly related to the latent X construct) to a crude step function relating latent X to X′. Thus, if continuous X was linearly related to Y, the use of a median split X′ is equivalent to revising one's prediction to be that the continuous X has a step function relation to Y, and that the step happens right at the median. Using the dichotomized X′ measure instead of the original X is the same as predicting that all values of X below the median lead to the same value of Y and all values of X above the median lead to a second value of Y.

Is this step function model an improvement or a distortion? That depends on the true relationship between X and Y *within* each group. As noted by DeCoster et al. (2009), if one believes the theoretical relationship between X and Y is a step function, it is inappropriate to dichotomize. With only two data points, it is impossible to test the substantive theoretical claim that the latent construct underlying X is categorical and binary. If latent X is categorical, one should test the assumption via polynomials, other forms of nonlinear regression or latent class models (DeCoster et al., 2009). With *k* discrete levels of a continuum, it is possible to test the substantive threshold claim that a) a dummy or contrast variable for whether a level is above or below the threshold is highly significant and b) a set of *k*-2 dummies or contrasts measuring variations in X above or below the threshold do not explain significant remaining variance (Brauer & McClelland, 2005; Keppel & Wickens, 2004, p. 104).

d. Sample dependence of the claimed step function X–Y relation

Median splits represent a special case of replacing a continuous X with a dichotomized X′. Suppose that one had strong theoretical reasons to believe that the function relating measured X and Y was a step function with threshold $X_{threshold}$. That threshold would presumably be the same no matter whether the sample of respondents in the study was drawn from a subpopulation likely to be high or low on the original X.

Median splits impose a sample-dependent threshold. There is no compelling theoretical argument underlying the implicitly claimed cut-point in a particular sample of data, when the cut-point is always the median of that particular sample. Spiller, Fitzsimons, Lynch, and McClelland (2013, p. 282) make a similar critique of relying on sample-dependent tests, citing Frederick's (2005) Cognitive Reflection Test with scores ranging 0 to 3. Results from a Massachusetts Institute of Technology (MIT) sample had M = 2.18, SD = .94, and results from a University of Toledo sample had M = .57, SD = .87. A "low" MIT score would be similar to a "high" University of Toledo score.

IPKSP offer nonstatistical arguments for using median splits in some situations and statistical arguments that turn on how median splits affect Type I and Type II errors. We first consider the nonstatistical arguments.

IPKSP's nonstatistical arguments for using median splits

Nonstatistical argument 1: Because median splits are popular and appear in the best journals they should be seriously considered as candidates for data analysis

We agree that median splits are popular. We would argue that the popularity of the practice hurts science. In their amusingly-titled chapter, "Chopped Liver? Ok. Chopped Data? Not OK," Butts and Ng (2009) bemoan this fact: "it should follow that if researchers are aware of the disadvantages of using chopped data and regard the practice as poor science, it should not occur with much frequency in articles published in high-quality journals."

As an example of the pitfalls of the popularity argument, Mani, Mullainathan, Shafir, and Zhao (2013a) published a high-profile paper in *Science* concluding that poverty impedes cognitive functioning. These authors reported experiments in which respondents were asked to say how they would cope with various financial shocks and then their cognitive functioning was measured. The key independent variables were a) the size of the shock manipulated between subjects and b) income. Income was measured continuously and subjected to median splits. Across three laboratory experiments, the key result was that measured cognitive functioning showed an interaction between the size of the shock and income. Thinking about coping with larger shocks inhibited subsequent cognitive functioning in the low-income group but not the high-income group. The same result did not obtain in a fourth study with nonfinancial scenarios.

The next issue of *Science* printed a criticism of those findings by Wicherts and Scholten (2013). They reported that when the dichotomized indicators were replaced by the original continuous variables, the critical interactions were not significant at $p < .05$ in any of the three core studies: p values were .084, .323, and .164. In a reply to Wicherts and Scholten, Mani, Mullainathan, Shafir, and Zhao (2013b) justified their use of median splits by citing papers published in *Science* and other prestigious journals that also used median splits. This "Officer, other drivers were speeding too" defense is often tried but rarely persuasive, especially here when the results of the (nonsignificant) continuous analyses were known. Though Mani et al. further noted their effect reached the .05 level if one pooled the three studies, we would guess that the editor poured himself or herself a stiff drink the night after reading Wicherts and Scholten's critique and the Mani et al. reply. It is hard to imagine that *Science* or many less prestigious journals would have published the paper had the authors initially reported the correct analyses with a string of three nonsignificant findings conventionally significant only by meta-analysis at the end of the paper. The reader considering the use of median splits should consider living through a similarly deflating experience. Splitting the data at the median resulted in an inaccurate sense of the magnitude of the fragile and small interaction effect (in this case, an interaction that required the goosing of a meta-analysis to reach significance), and a publication that was unfortunately subject to easy criticism.

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006

Nonstatistical argument 2: Median splits are useful for the expression of categorical latent constructs

IPKSP consider the argument that median splits are appropriate when the underlying X is theoretically categorical.

> "In fact, there are numerous constructs that, while being measured on continuous rating scales, are conceptually more discrete, viz dichotomous (MacCallum et al., 2002). For example, locus of control (Srinivasan & Tikoo, 1992) is usually discussed with an emphasis on internal versus external, people are said to be low or high on "self-monitoring" (Becherer & Richard, 1978), and people are said to be low or high in their "need for closure" (Silvera, Kardes, Harvey, Cronley, & Houghton, 2005). Such personality typologies abound: introversion and extraversion, gender identity, type A and B personalities, liberal and conservative, and so forth. When researchers think in terms of groups, or study participants having relatively more or less of a characteristic, it is natural that they would seek an analytical method that is isomorphic, so the data treatment may optimally match the construct conceptualization." (p. 3)

We have two responses to this argument. First, the examples offered are older papers that treat a continuous variable categorically when the overwhelming majority of researchers subsequently using the same scales consider these same constructs to be continuous and not categorical (cf. Czellar, 2006; Disatnik and Steinhart (2015), Hoffman, Novak & Schlosser, 2003; Judge & Bono, 2001; Kardes, Fennis, Hirt, Tormala, & Bullington, 2007; Shah, Kruglanski, & Thompson, 1998; Webster & Kruglanski, 1994). We believe that when an author uses language such as "high and low" on some continuous construct, this terminology is often a linguistic convenience rather than a claim that the construct is categorical with categories of equal frequencies. For example, readers should ask themselves if they really believe that authors describing "liberals" and "conservatives" intend to imply only two levels of liberalism–conservatism.

Second, as noted above, if one believes that the theoretical X–Y relation is a categorical step function a) the threshold is unlikely to be at a sample-dependent value like the median, and b) a step function is a substantive theoretical claim that cannot simply be assumed; it should be tested using the statistical methods just mentioned. Median splits do not allow testing of thresholds and in fact make it impossible to see any sort of nonlinear relationship involving the variables that might be theoretically characterized as thresholds. If the continuous data are split into two categories then all that can be tested is the difference between those two categories, i.e., a line. There would be no way to tell from split data whether the original data had a linear or nonlinear relationship with Y. Leaving the data continuous allows for testing of whatever nonlinear relationship the researcher would like to test.

Nonstatistical argument 3: Median splits and ANOVA are easier to conduct and understand than regression/ANCOVA

IPKSP note that some say that dichotomization of a continuous X makes the analysis easier to conduct and interpret. Regression, ANOVA, and ANCOVA (the combination of regression and ANOVA) are of course identical at their core and are simply different-appearing instantiations of a general linear model. ANOVA may seem easier to conduct for people trained long ago because before computer packages, ANOVA was easier for students to compute by hand and with a calculator. Given this constraint, it made some sense that median splits were utilized to help researchers turn their regression data into ANOVA-friendly data (Cohen, Cohen, West and Aiken, 2003). This reason for median splits is no longer a good argument for splitting data. Graduate training in regression and ANCOVA has become ubiquitous. We have collectively trained literally hundreds of PhD students over the years in these techniques, including many who have gone on to publish in *JCP* and other top outlets for consumer research.

IPKSP go on to suggest that ANOVA is easier than regression to use, because regression results are, "more difficult to interpret because there are no post hoc tests specifying which values of the independent variable are significantly different from each other." We strongly disagree. If researchers want to test at particular focal values of X for significance then they can use spotlights (Fitzsimons, 2008; Irwin & McClelland, 2001; Spiller et al., 2013). If there are no focal values, it can be useful to report a "floodlight" analysis, reporting regions of the continuous X variable where the simple effect of manipulated Z is significant (Spiller et al., 2013). Johnson and Neyman (1936) originally proposed these tests, but these tests did not catch on when statistical computations were by hand. Andrew Hayes' ubiquitous PROCESS software now makes it trivial to find and report these regions as a follow-up to finding significant interactions between manipulated Z and continuously measured X.

Further, the argument that median split analyses are easier to conduct, report, and read breaks down once one acknowledges what the researchers must report to convince the reader that the analyses are not misleading. As we describe later, the researcher wishing to justify a median split of that covariate must compute correlations between the variable to be split and each factor and interaction. For a simple two-by-two ANOVA design with a single covariate, the researcher would need to compute three correlations, one with each factor and one with the interaction. As we will show later in this paper, what matters is the magnitude of these correlations, not their statistical significance as suggested by IPKSP. Our Fig. 3 later in this paper shows that simply by chance, some of these correlations would likely be large enough to cause serious bias. The researcher would have to prove to the reader that substantive and statistical conclusions do not differ between the median split analysis and the fully reported corresponding ANCOVA (cf. Simmons, Nelson, & Simonsohn, 2011). To us, this seems more difficult for researchers and readers alike than simply reporting the correct analysis using the continuously measured independent variables.

Nonstatistical argument 4: Median splits are more "parsimonious"

IPKSP argue that it is more "parsimonious" to use a 2-category indicator of one's latent X defined by a median split

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006

than to use the original continuous measure of X. The standard definition of parsimonious is that nothing unnecessary is added, that the simplest model or analysis technique is preferred. Philosophers from Aristotle to Galileo to Newton have agreed. Galileo remarked, "Nature does not multiply things unnecessarily; that she makes use of the easiest and simplest means for producing her effects; that she does nothing in vain, and the like" (Galileo, 1962, p. 397).

Adding an extra unnecessary step (requiring calculation of a statistic, the median) to take the data away from its original form is not more parsimonious. This use of the concept of parsimony is not in line with the usual scientific use of the word.

Statistical considerations in the use of median splits: Effects on Type II and Type I errors

Beyond their nonstatistical arguments, IPKSP make statistical and quasi-statistical arguments about the consequences of median splits. We articulate these arguments and present counter-arguments below.

IPKSP statistical argument 1: Median splits are "conservative"

Type II errors and conservatism

Much of the IPKSP paper rests on the argument that the use of median splits is conservative. As noted above, a primary problem with median splits is that they add error, and thus on average median splits reduce power. There is no way around this fact, statistically, and lowering power with no compensating benefit would be considered to be a bad thing by most researchers and all statisticians we know. IPKSP rebrand the reduction of power as a benefit, labeling it, "conservative."

Conservatism, in a statistical sense, simply means increasing the chance of Type II errors and decreasing the chance of Type I errors. Decreasing your alpha requirements to declare something significant (say, from .05 to .01) would make a test more conservative, with the cost in increased Type II errors having some offsetting benefit in fewer Type I errors. Splitting data is not conservative in the same way: it increases the chance of both types of errors because sometimes split data are significant when the continuous data would not be. If researchers pick the method that yields significance, then Type I errors will increase even as splitting, overall, reduces power.

Median splits and false–positive consumer psychology

The fact that a given sample of data might have a significant relationship between Y and X for X split at the median and not for continuously measured X implies that there is a significant risk of "false–positive" consumer psychology when authors are licensed to analyze their data either way and report whichever comes out to be more significant. In an influential article, Simmons et al. (2011) noted how "undisclosed flexibility in data collection and analysis allows presenting anything as significant." They focused on "p-hacking" by topping up subjects in a study until statistical significance is reached or collecting multiple covariates and adding them to the analysis of an experiment in different combinations. Gelman and Loken (2014) argue that this is producing a "statistical crisis in science" — when researchers' hypotheses can be tested in multiple ways from the same data set and what is reported is what works out as most supportive of their theorizing. We simulated the effects for 10,000 samples of $N = 50$ from a bivariate distribution with true correlation of 0, tested at $\alpha = .05$ for the continuous X–Y correlation and then at $\alpha = .05$ for the correlation between Y and median-split X'. If one picks and chooses, in 8% of all samples one or the other or both tests will significant.

As Gelman and Loken (2014) noted, it is not just unscrupulous researchers who fall into this trap. Well-meaning researchers see multiple alternatives as reasonable and decide a posteriori which seems most reasonable — with more thinking about alternatives when things don't work out. We are concerned that IPKSP risk giving researchers cover for more undisclosed flexibility in data analysis. This allowance just goes into the "researcher analysis degrees of freedom" issue that fueled the Simmons et al. (2011) "false–positive psychology" paper and the associated recent heightened concern about findings in the social sciences that do not replicate.

Bayes theorem and effects of low power

Bayes theorem is the normatively appropriate model for updating beliefs on the basis of evidence observed from sample data. Bayes theorem shows that less belief shift is warranted from a statistically significant finding the lower the power of the study. IPKSP (p. 2) note that:

> "In his book, *Statistics as Principled Argument*, Abelson (1995) repeatedly made the point that there are many misconceptions about statistics, and we might argue that misconceptions about median splits should be added to Abelson's list."

We do not believe that Abelson would agree if he were still alive. Abelson (1995) and Brinberg, Lynch, and Sawyer (1992) both rely on Bayes theorem to make the point that reducing power implies reducing the belief shift that is warranted from observing a statistically significant result. Consider hypothesis H that there is a relationship of a certain magnitude (say $r = .25$) between X and Y in the population and the null hypothesis H− that there is no association between X and Y. The expected prior odds of the relative likelihood of H and H− = P(H) / P(H−). Then one observes Datum D, a statistically significant empirical association between X and Y. The likelihood of observing D under hypothesis H is the statistical power of the test, P(D|H). The likelihood of observing D under H−, the null hypothesis, is one's Type I error rate alpha (P(D|H−). Bayes theorem says that the updated posterior odds ratio of H and H− is now the prior odds ratio times the relative likelihood of observing datum D given H versus H−. Specifically:

$$\frac{P(H|D)}{P(H-|D)} = \frac{P(H)}{P(H-)} \times \frac{P(D|H)}{P(D|H-)} \tag{1}$$

Eq. (1) says that the greater the power relative to Type I error rates, the greater the belief shift.

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006

Abelson (1995) articulates his five "MAGIC" criteria for persuading readers based on study findings: Magnitude, Articulation, Generality, Interestingness, and Credibility. The first of these is "magnitude." Chapter 3 of his book is devoted to the argument that results are more persuasive if they reflect bigger effect sizes. Thus, reducing expected effect size by use of median splits is not a decision to be "conservative" and persuasive. It is a decision to be less persuasive.

> "We propose that the rhetorical impact of a research result is a direct function of the raw effect size divided by the "cause size," a ratio that we call *causal efficacy*. A large effect from a small variation in the cause is the most impressive, whereas a small effect arising from an apparently large causal manipulation is the most anticlimactic and disappointing."
>
> [Abelson (1995), p. 48]

"Conservative" studies do not make for a conservative science

Thus far we have focused on how the researcher's choice to use a median split degrades the persuasiveness of his or her article. That's problematic, but arguably the author's own choice. But consider the aggregate implications of these arguments from the perspective of a journal editor. From a pool of statistically significant results (all observing D rather than D−), some subset is made up of true effects and the complementary subset is made up of Type I errors. The proportion of published significant results that are Type I errors is directly determined by the ratio $P(D|H)/P(D|H-) = (1 - \beta)/\alpha$, power divided by Type I error rate.

Ioannidis (2005) has pointed out that even in the absence of p-hacking or any bias in reporting, the probability that a statistically significant research finding is true is an increasing function of power, and therefore, of effect size. Assume a world where of all hypotheses researchers investigate, half are true and half are actually null effects in the population. Further, assume that papers are not published unless study findings are significant at $\alpha = .05$. Imagine two versions of that world, one where power to detect real effects is .80 and another where it is .40. When power is .80 and $\alpha = .05$, the ratio of likelihood of finding significant results when the null is false to when it is true is 16 to 1 — for every 17 significant results reported, 16 are real. When power is .4 and at $\alpha = .05$, the ratio of likelihood of finding significant results when the null is false to when it is true is 8 to 1 — for every 9 significant results reported, 8 are real. Editors who countenance median splits are making a choice to publish proportionately more Type I errors in expectation relative to the number of results that reflect a true effect in the population.

IPKSP statistical argument #2: The loss of power from median splits is minimal and easily offset

When there is a linear relationship between Y and latent X, the effect size (i.e. the r-squared value for the model) when correlating Y with X′ via median splits is, for normally distributed data, around .64 of the value when correlating Y with continuously measured X. That is, the split data have 64% ($2/\pi$) of the effect size that the original data had before dichotomization. Irwin and McClelland (2003) show that the damaging reduction in power persists even when the independent variable is not normally distributed. Rather than reporting the r-squared, IPKSP instead focus on the fact that the split coefficient is 80% of the original coefficient, perhaps causing a casual reader to underestimate the loss due to dichotomization.

One of the most disturbing aspects of IPKSP is the suggestion that losing power is fine, because researchers can simply increase sample size to make up for the loss. An estimate for normally distributed covariates is that sample size would need to be increased by $\pi/2 = 1.57$. Increasing sample size by 57% to accommodate for a median split is both costly and potentially unethical. IPKSP, making an argument that median splits are acceptable, approvingly cite two studies from the medical literature that used median splits in their analyses (Kastrati et al., 2011; Lemanske et al., 2010). We believe that these studies do not support IPKSP's point; rather, these studies illustrate why "just adding more participants" is an unwise solution to the power loss caused by median splits. These were medical experiments on actual patients, with true risks. Some participants in Kastrati et al. died; some children in Lamenske et al. were hospitalized with serious conditions. We believe it would be unconscionable to increase in sample size so that the researchers could use median splits because regression is somehow less convenient for them. Sadly, *none* of the split variables in those two studies had statistically significant relationships.

Admittedly, the stakes in consumer psychology experiments are typically not that extreme. However, in our field as well, there are ethical issues involved with routinely using 57% more participants than necessary. Requiring people to run more participants in their studies simply to avoid using multiple regression instead of ANOVA wastes the time of volunteers in course-related subject pools (cf. Smith, 1998), wastes money when paying participants, and potentially depletes the common resource of willingness to participate in surveys. In any case, researchers owe it to the participants who have graciously provided data to analyze those data using the best and most powerful statistical methods. Losing power is bad, and deliberately losing power via median splits is neither effective nor efficient use of research resources.

Simulations

We have examined IPKSP's simulations and compared them to the code shown in the Appendix A from their paper. Our examination revealed serious problems with the simulations. In some instances the code does not match its description in the paper and in other instances the aggregated reporting of the results substantially underestimates the deleterious effects of median splits. We present the highlights of our analysis of the simulations here and provide extensive details, including revisions and extensions of the figures in an online technical appendix. We consider the following important issues in their simulation results.

Interactions

A major problem with IPKSP's claim of support for splitting a single continuous covariate in ANCOVA designs is that no simulation results are presented for the effect of such splitting on interactions. IPKSP's Fig. 3 purports to show the effects on the interaction term but their sampling from a multivariate normal distribution completely precludes any possibility of generating non–null interactions. Aiken and West (1991, p. 181) prove that if Y, X_1, and X_2 have a multivariate normal distribution, then the coefficient for the interaction term must be zero. They summarize (emphasis in the original):

> This result seems surprising. It says that when *two predictors X and Z and a criterion Y are multivariate normal, the covariance between the product XZ and Y will be zero*. Does this mean that there is necessarily no interaction if X, Z, and Y are multivariate normal? Yes. Turning the logic around, if there exists an interaction between X and Z in the prediction of Y, then, necessarily, the joint distribution of X, Z, and Y is *not* multivariate normal.

How then do IPKSP provide their Fig. 3, which purports to be the effect on the standardized partial regression coefficient for the interaction term when one variable is split as a function of the correlation between the independent variables? IPKSP (p. 4) state for Study 1: "A multiple regression model was used to analyze the three variables, and the estimates for β_1, β_2, and β_3 (for the interaction term) were obtained."

However, an examination of the SAS code provided by IPKSP reveals they only estimated and recorded the additive effect coefficients for the continuous and median split analysis along with some of the p-values. They neither recorded nor aggregated results for the interaction coefficient. Had they done so, they would have found that, inconsistent with their Fig. 3, the mean coefficient for the interaction whether in the continuous or median split analysis was zero.

A reader of IPKSP might believe that their Fig. 3 came from the same simulation runs as their Figs. 1 and 2. However, the code in IPKSP's Appendix A reveals that instead of computing the interaction as the product of X_1 and X_2 in their original simulations, IPKSP created an additional simulation in which they sampled a third variable X_3 from a multivariate normal distribution.[1] For the continuous estimates (upper curve in their Fig. 3), this third variable X_3 is simply labeled as an interaction[2] although the mean correlation between this "interaction" and the product of X_1 and X_2 is 0 when it should be 1. That is, rather than analyzing Y as a function of X_1, X_2, and $X_1 * X_2$, IPKSP analyzed Y as a function of X_1, X_2, and X_3. For the split estimates (the lower curve in their Fig. 3), *SplitX*$_1$ was calculated by splitting continuous X_1 into a $\{0, 1\}$ variable. Rather than analyzing Y as a function of *SplitX*$_1$, X_2, and *SplitX*$_1 * X_2$, IPKSP analyzed Y as a function of *SplitX*$_1$, X_2, and *SplitX*$_1 * X_3$.[3] Coefficient estimates for this last term neither represent the true interaction nor can they be meaningfully compared to the continuous estimates. Thus, IPKSP's Fig. 3 does not depict results about interaction terms and should be ignored entirely.

The simulations underlying IPKSP's Fig. 4 explicitly built in a null effect for the interaction. Hence, IPKSP present no information about the effects of median splits on the estimate of true interactions. If they had simulated an actual interaction, what would they have found? Busemeyer and Jones (1983) showed that interactions are very fragile due to measurement error in the independent variables and monotonic transformations. Median splits introduce unnecessary measurement error and are a heavy-handed monotonic transformation. McClelland and Judd (1993) show that even without those problems there is precious little statistical power for detecting interactions involving continuous variables, especially ones having normal distributions. Mani et al. (2013a) provide an empirical example of surprising effects on interactions caused by median splits. IPKSP favorably cite Farewell, Tom, and Royston's (2004) analysis of prostate cancer data, but even they warn, "this example illustrates the potential pitfalls of trying to establish interactions of treatment with a continuous variable by using cutpoint analysis." IPKSP present no simulations or other information to alleviate concerns about the many dire warnings against using median splits when interactions might be involved.

Simulations versus derivations

The simulation results in IPSKP's Figs. 1 and 2 are unnecessary because they are easily derivable. This observation is not a criticism in itself. Instead, we use the derivations both to extend their results to parameter values that IPKSP did not consider and to provide a more detailed examination of the effects of splitting a continuous variable. Cohen et al. (2003, p. 68) present the basic formulas for computing standardized partial regression coefficients from correlations:

$$\beta_{Y1.2} = \frac{r_{Y1} - r_{Y2}\ r_{12}}{1 - r_{12}^2} \quad \text{and} \quad \beta_{Y2.1} = \frac{r_{Y2} - r_{Y1}\ r_{12}}{1 - r_{12}^2}. \tag{2}$$

It is well known that performing a median split of a normally-distributed variable reduces its squared correlation with other variables to $2/\pi \approx 0.64$ of what it would have originally been without splitting. That is, a model predicting Y from the median split of X_1 suffers a loss of explanatory power of 36% compared to a model predicting Y from continuous X. Adding the factor $\sqrt{2/\pi}$ to r_{Y1} and r_{12} in the above equations provides expected values for standardized partial regression coefficient when normally-distributed X_1 but not X_2 is split at its median.

[1] This can be seen in IPKSP's Appendix A as "call vnormal(x,,sigma,nn,);" in conjunction with the first three sets of modifications.

[2] This can be seen in IPKSP's Appendix A as "interact = x[,4];".

[3] This can be seen in IPKSP's Appendix A as "interact = x[,4]#x[,5]; regx = interc||x[,5]||x[,3]||interact;", where x[,5] represents *SplitX*$_1$, x[,3] represents X_2, x[,4] represents X_3, and interact is recalculated in the first statement to be X_3 * *SplitX*$_1$.

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006

Unrepresentative sampling

Having the exact formulas for the expected values allows a re-examination of the research situations generated by IPKSP's sampling strategy. IPKSP sampled from a multivariate normal distribution varying r_{Y1}, r_{Y2}, and r_{12} each over the set of values {0, 0.1, 0.3, 0.5, and 0.7} for a total of 125 conditions. We are ignoring the sampling over *n*, as did they, because the formulas provide expected values independent of *n*. Although the factorial sampling of the correlation values used by IPKSP may seem appealing, it has the disadvantage of creating many unusual, unrepresentative combinations. For example, one of their sampled conditions is $r_{Y1} = 0.7$, $r_{Y2} = 0$, $r_{12} = 0.7$. Using the above formulas the expected estimates in this case are $\beta_1 = 1.37$ (greater than 1.0 because of the collinearity) and $\beta_2 = -0.96$ (strongly negative despite the zero-order correlation being zero). Although possible, this combination of parameter values is rare and not representative of typical research. Why are we interested in the effect of median splits in such atypical conditions? In 54 (42%) of the 125 sampled conditions, one or the other standardized partial regression coefficients are negative, despite having non-negative zero-order correlations. More importantly, as we shall illustrate later, IPKSP's averaging across these positive and negative values makes the effects of splitting appear more benign than the reality revealed in the disaggregated results.

Estimates of β_1 in Study 1

IPSKP's Fig. 1 displays estimates of the standardized partial regression coefficient for X_1 with and without the median split. Despite having performed their simulations over a set of five values for the independent variable intercorrelation (0, 0.1, 0.3, 0.5, and 0.7), they report only three (0, 0.3, and 0.5). Our derivations in the technical appendix show that the ratio of the split estimate to the original estimate as well as the increment in squared correlation depend only on the correlations of the independent variables with each other and not with the dependent variable. The graphs of these relationships in our Fig. 1 show an increasingly rapid loss in parameter size and explained variation as the intercorrelation increases. Even with no correlation, splitting produces a sharp loss of 20% in the size of the parameter estimate and an even larger loss of 36% of increment of the squared correlation to about 64% of what it would have been, as represented by the dashed line. These initial penalties are stiff and these penalties rapidly increase for both the parameter estimate and the increment in squared correlation as the intercorrelation among independent variables increases.

As an example, consider the plausible case where $r_{Y1} = 0.3$, $r_{Y2} = 0.3$, $r_{12} = 0.3$. The parameter estimate for the continuous data is 0.23 and is reduced to 0.178 by splitting. The very small increment in squared correlation of 0.048 for the continuous analysis loses 38.5% of its value to 0.03 by splitting. Few researchers in the social sciences can afford a *minimum* loss of 36% of their already small increments in explained variation. It is useful to quantify just how sizable

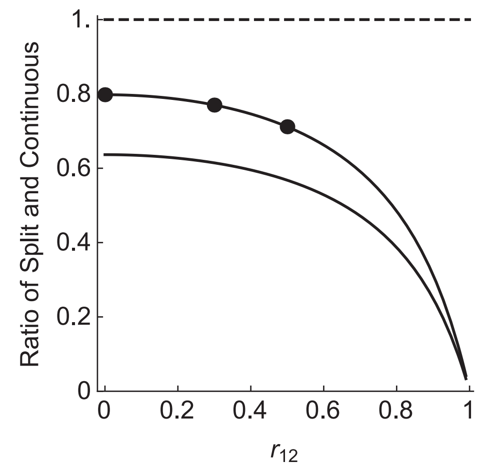

Fig. 1. Ratio of split to continuous results for parameter estimate of β_1 (top curve) and its increment in squared correlation as a function of the predictor intercorrelation. Dashed line represents no splitting of the continuous variable.

these reductions are. When $r_{12} = 0.3$, splitting an independent variable and using $\alpha = 0.05$ for significance testing is equivalent (in expectation) to doing the analysis of the continuous variable but using $\alpha = 0.01$. This is substantial and unnecessary loss of power, which becomes rapidly worse as the intercorrelation between independent variables increases beyond 0.3.

Estimates of β_2 in Study 1

IPKSP's Fig. 2 displays the standardized partial regression coefficient for X_2 with and without the median split of X_1. The aggregation across disparate conditions in their Fig. 2 presents an unrealistically benign view of the effects of splitting one variable on the estimate of the other variable. IPKSP's Fig. 2 shows in the aggregate what the authors refer to as a slight "lifting" of the estimate of β_2. However, we prove in the technical appendix that whenever

$$r_{12} < \frac{r_{Y1}}{r_{Y2}}, \tag{3}$$

splitting the first predictor *increases* the estimate of the coefficient for the second predictor compared with what it would have been without the median split. Conversely, the opposite inequality implies that splitting the first predictor *decreases* the estimate of the coefficient for the second predictor. IPKSP's sampling scheme included more of the former than the latter so the weighted average displayed in their Fig. 2 shows a slight increase. Disaggregating results according to the inequality reveals major distortions in the estimate of β_2 as the predictor correlation increases.

Consider the special but realistic case for which the two zero-order correlations are approximately equal; that is, $r_{Y1} = r_{Y2}$. Then the ratio of the two correlations is 1.0 and the intercorrelation r_{12} is necessarily less than 1.0 so the effect of splitting the first predictor will be to enhance the estimate of the second predictor, with the enhancement increasing as the correlation increases. Simultaneously, increasing predictor intercorrelation implies decreasing estimates of the first predictor's coefficient. The combined effect is dramatic, as illustrated in the graph of the ratio β_2/β_1 in our Fig. 2. The ratio of the true values is 1 but at a minimum when the predictors are independent, the ratio when the first predictor is split is 1.25. For correlations of 0.3, 0.5, and 0.7, the ratio increases to 1.45, 1.71, and 2.32, respectively. Thus, splitting one variable when predictor correlations are equal would lead a researcher to misjudge the relative magnitude of the two predictors.

In the case where the predictor correlations with the dependent variable are unequal, i.e., $r_{Y1} \neq r_{Y2}$, we prove in the technical appendix that the estimate for the second coefficient when the other is split becomes a weighted average of the coefficients from the continuous analysis. That is:

$$\beta^*_{Y2.1} = w\beta_{Y1.2} + (1-w)\beta_{Y2.1}. \quad (4)$$

The exact form of w is not as important as recognizing that the estimate of $\beta^*_{Y2.1}$ when splitting X_1 is always a *confounding* of the original two coefficients for the continuous analysis, and the confounding works in the same way as poor experimental design. As an example consider the plausible case $r_{Y1} = 0.5$, $r_{Y2} = 0.3$, $r_{12} = 0.3$. Then $w = 0.18$ and $\beta^*_{Y2.1} = w\beta_{Y1.2} + (1-w)\beta_{Y2.1} = 0.18(0.45) + (1-0.18)(0.165) = .22$. In other words, splitting X_1 increases the estimate of the coefficient for X_2 from 0.165 to 0.22 by averaging in 18% of the coefficient for X_1. This is substantial and unnecessary confounding that should not be acceptable in scientific publications.

Study 2

IPSKP's Fig. 4 reports simulation results for a very narrow set of conditions: variable A is continuous and has an effect of varying magnitude whereas variable B, a two-level factor, and the interaction A x B have null effects. IPKSP's simulations reveal negligible effects on the average *p*-values for B and A x B when the continuous variable A is split at its median. However, there are problems with the simulations and how they are presented in their Fig. 4.

First, statistical power is of more interest than average *p*-values for evaluating analysis procedures. If effect sizes are small there is little hope of finding significant results whether or not a variable is split, and if effect sizes are large one can find significant results even if the data are abused by splitting. Of greater interest is the power for typical effect and sample sizes. As we have done above, consider an effect size[4] of $r_{YA} = 0.3$ and a sample size of $n = 100$. We used a simulation in R equivalent to the SAS code provided by IPKSP to find mean *p*-values of 0.032 for the continuous analysis of A and 0.082 when A is split at its median.[5] More importantly, the power for the continuous analysis equals 0.86, greater than the minimum power of 0.80 recommended by Cohen, whereas the power when A is split is only 0.68, below Cohen's recommendation. Given a real effect, we expect the continuous analysis to (correctly) produce a significant effect about 20% of the time more frequently than the split analysis. Thus, it is at the moderate effect and sample sizes most likely to be encountered in consumer psychological research that research is most likely to be adversely affected by the conservatism of median splits.

Second, and more importantly, the context underlying their Fig. 4 sets a low bar (i.e., whether there are any changes in *p*-values for null effects) in an unrepresentative research context. It would be quite unusual in an ANCOVA design with a single continuous covariate A, for the two-level factor B to have no anticipated effect on the outcome variable Y. Note also that the prior results showing distortions and major decreases in the parameter estimate and effect sizes when predictors are correlated precludes the use of two continuous covariates because they are likely to be correlated to some degree. IPKSP's analysis showed that splitting a continuous variable A at its median could have substantial effects on the other parameter estimate for the other variable, in this case B, when A and B were correlated. Nothing in that analysis required that B be continuous so any correlation between A and B risks distorting the analysis of B and the A x B interaction, likely artificially enhancing them.

Even though with random assignment to conditions a researcher would expect a zero population correlation between A and B, problems might still arise because it is not the *population* correlation but the actual correlation in the *sample* of data that determines whether there will be distortion in the estimates of B. It is irrelevant whether this correlation in the sample is statistically significant because any sample correlation will distort the parameter estimates. Our Fig. 3 displays the sampling distributions for the correlation coefficient for sample sizes of 50 (shallowest), 100, and 200 (steepest) when the population correlation is zero. Clearly, there is ample

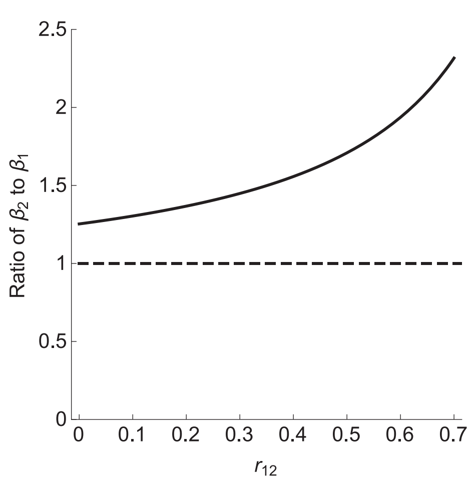

Fig. 2. Ratio of estimated coefficients when splitting X_1 at its median (solid line) versus leaving X_1 continuous when $r_{Y1} = r_{Y2} \neq 0$.

[4] We were unable to reproduce IPKSP's translation of correlations to the mean differences in standard deviations used as indices for the graphs in their Fig. 4. Instead of using those mean differences, we report our results in terms of the correlations used to generate variable A.

[5] This value is lower than reported in IPSKP's Fig. 4, which appears to be about 0.11. Running their SAS code also produced a value equal to about 0.08 so it appears the value in Fig. 4 is a transcription error.

opportunity for there to be a sample correlation that would cause problematic distortions in the estimates of the ANOVA factors when a continuous covariate is split at its median. Because, as shown above, the bias in the other estimates is a weighted average of the estimates from a continuous analysis, the danger is greater the larger the effect size of A. Hence, a recipe for an unscrupulous researcher wanting to enhance the effect size of factor B is to find a strong covariate A and split it at its median.

Summary

Contrary to the arguments of IPKSP, there is no compelling reason to split continuous data at their median. Splitting a continuous variable at its median introduces random error to the independent variable by creating a sample-dependent step function relating Y to latent X.

IPKSP argue that in some contexts it is appropriate to use median splits because they are popular. We describe the harrowing negative consequences of relying on their popularity.

IPKSP argue that analyses using median splits are easier to conduct, report, and understand than analyses using the original metric. We describe how a full accounting of the necessary conditions to safely use a median split is more onerous (and includes conducting and reporting the continuous analysis).

IPKSP argue that median splits are useful to test categorical latent constructs. Yet their own examples are continua, not categories, and if there were a substantive claim of such theory-derived thresholds, the functional form of such analyses would require testing.

Most critically, IPKSP argue that median splits are not problematic because they are "conservative," that is, because they merely make it more difficult to detect a true effect. Yet authors who choose to reduce statistical power by using median splits reduce the persuasive impact of their own findings (Abelson, 1995). Further, by licensing authors to use median splits or continuous analyses at their discretion, IPKSP open the door for an additional researcher degree of freedom and cherry-picking of the more favorable result (Gelman & Loken, 2014; Simmons et al., 2011). It is easy to show analytically that even without such cherry-picking, the lower the power of statistically significant findings in the literature base, the higher the proportion of results in the literature that will be false or overstated (Ioannidis, 2005). Cavalierly lowering power by median splits creates a less reliable literature base.

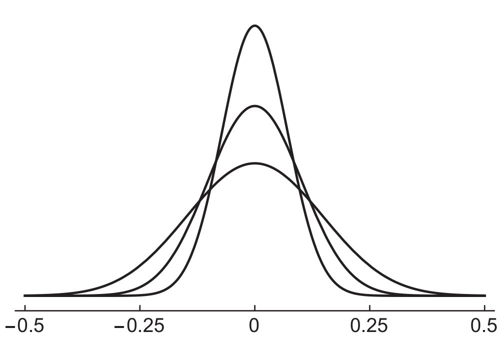

Fig. 3. Sampling distributions for correlation coefficient r for sample sizes of 50 (shallowest), 100, and 200 (steepest).

Finally, the publication of IPKSP's article depends on their simulations. These simulations are flawed. The text of IPKSP describes the simulations as bearing on models with interactions, but coding errors in the simulation of the interaction of a categorical and a continuous independent variable indicate that readers are learning about the effects of median splits in a model with three additive independent variables. IPKSP did not simulate the interaction. Further, the simulations needlessly aggregate across different situations with different effects. By pooling across multiple simulation conditions, IPKSP combine cases where the effect is underestimated with those where it is overestimated, leading to a misleading overall result of "not too bad." This error can be shown analytically.

According to IPSKP, their "main contribution is giving the green light to researchers who wish to conduct a median split on one of their continuous measures to be used as one of the factors in an orthogonal experimental design, such as a factorial, and then use ANOVA to model and communicate results" (p11). We see no such green light, and many red flags. IPKSP state unequivocally that, "there is no material risk to science posed by median splits pertaining to Type II errors" (p4). We hope that we have made clear in this commentary why we could not disagree more with this conclusion.

Appendix A. Mathematical Derivations

Mathematical derivations for this article can be found online at http://dx.doi.org/10.1016/j.jcps.2015.05.006.

References

Abelson, R. P. (1995). *Statistics as principled argument.* New York: Psychology Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage Publications.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. http://dx.doi.org/10.1037/a0021524.

Brauer, M., & McClelland, G. H. (2005). L'utilisation des contrastes dans l'analyse des données: Comment tester les hypothèses spécifiques dans la recherche en psychologie? *L'Année Psychologique*, *105*(2), 273–305.

Brinberg, D., Lynch, J. G., Jr., & Sawyer, A. G. (1992). Hypothesized and confounded explanations in theory tests: A Bayesian analysis. *Journal of Consumer Research*, *19*(2), 139–154.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, *93*(3), 549–562.

Butts, M. M., & Ng, T. W. (2009). Chopped liver? OK. Chopped data? Not OK. In C. E. Lance, & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 361–386). New York: Taylor & Francis.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006

Czellar, S. (2006). Self-presentational effects in the Implicit Association Test. *Journal of Consumer Psychology*, *16*(1), 92–100.

DeCoster, J., Iselin, A. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*(4), 349–366.

Disatnik, D., & Steinhart, Y. (2015). Need for cognitive closure, risk aversion, uncertainty changes, and their effect on investment decisions. *Journal of Marketing Research*, *52*(June), 349–359.

Farewell, V. T., Tom, B. D. M., & Royston, P. (2004). The impact of dichotomization on the efficiency of testing for an interaction effect in exponential family models. *Journal of the American Statistical Association*, *99*(467), 822–831.

Fitzsimons, G. J. (2008). Editorial: Death to dichotomizing. *Journal of Consumer Research*, *35*(1), 5–8.

Fleischmann, M., & Pons, S. (1989). Electrochemically induced nuclear fusion of deuterium. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, *261*(2), 301–308.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25–42.

Galileo, G. (1962). *Dialogue Concerning the Two Chief World Systems*. Translated by S. Drake, Foreword by Albert Einstein. Berkeley: University of California Press.

Gans, J. S., & Shepherd, G. B. (1994). How are the mighty fallen: Rejected classic articles by leading economists. *The Journal of Economic Perspectives*, *8*(1), 165–179.

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*(6), 460.

Hoffman, D. L., Novak, T. P., & Schlosser, A. E. (2003). Locus of control, web use, and consumer internet regulation. *Journal of Public Policy and Marketing*, *22*(1), 41–57.

Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology*, *70*(6), 873–876.

Humphreys, L. G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-difference variables. *Journal of Educational Psychology*, *66*(4), 464–472.

Iacobucci, D., Posovac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*. http://dx.doi.org/10.1016/j.jcps.2014.12.002 (this issue).

Iacobucci, D., Saldanha, N., & Deng, X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 139–153.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), 696–701.

Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research*, *38*(1), 100–109.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, *40*(August), 366–371.

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, *1*, 57–93.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*(1), 80.

Kardes, F. R., Fennis, B. M., Hirt, E. R., Tormala, Z. L., & Bullington, B. (2007). The role of the need for cognitive closure in the effectiveness of the disrupt-then-reframe influence technique. *Journal of Consumer Research*, *34*(3), 377–385.

Kastrati, A., Neumann, F. -J., Schulz, S., Massberg, S., Byrne, R. A., Ferenc, M., et al. (2011). Abciximab and heparin versus bivalirudin for non-ST elevation myocardial infarction. *The New England Journal of Medicine*, *365*(21), 1980–1989.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Berkeley: University of California Press.

Lemanske, R. F., Mauger, D. T., Sorkness, C. A., Jackson, D. J., Boehmer, S. J., Martinez, F. D., et al. (2010). Step-up therapy for children with uncontrolled asthma receiving inhaled corticosteroids. *The New England Journal of Medicine*, *362*(11), 975–985.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013a). Poverty impedes cognitive function. *Science*, *341*(6149), 976–980.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013b). Response to comment on "Poverty impedes cognitive function". *Science*, *342*(6163), 1169-e.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*(1), 181–190.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.

Shah, J. Y., Kruglanski, A. W., & Thompson, E. P. (1998). Membership has its (epistemic) rewards: Need for closure effects on in-group bias. *Journal of Personality and Social Psychology*, *75*(2), 383.

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.

Smith, N. C. (1998). Presidential session summary: Ethics in consumer research. In J. W. Alba, & J. W. Hutchinson (Eds.), *Advances in consumer research*, *Volume 25*. (pp. 68). Provo, UT: Association for Consumer Research.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, *50*(2), 277–288.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

Wicherts, J. M., & Scholten, A. Z. (2013). Comment on "poverty impedes cognitive function". *Science*, *342*(6163), 1169-1169.

Please cite this article as: McClelland, G.H., et al., Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power, *Journal of Consumer Psychology* (2015), http://dx.doi.org/10.1016/j.jcps.2015.05.006