Available online at www.sciencedirect.com

ScienceDirect

Journal of Consumer Psychology 25, 4 (2015) 690 – 704

JCP JOURNAL OF CONSUMER PSYCHOLOGY

Research Article

The median split: Robust, refined, and revived☆

Dawn Iacobucci [a,*], Steven S. Posavac [a], Frank R. Kardes [b], Matthew J. Schneider [c], Deidre L. Popovich [d]

[a] *Vanderbilt University, Nashville, TN 37203, USA*
[b] *Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA*
[c] *Medill School of Journalism, Northwestern University, Evanston, IL 60201, USA*
[d] *Rawls College of Business, Texas Tech University, Lubbock, TX 79409, USA*

Received 17 June 2015; accepted 26 June 2015
Available online 3 July 2015

Abstract

In this rebuttal, we discuss the comments of Rucker, McShane, and Preacher (2015) and McClelland, Lynch, Irwin, Spiller, and Fitzsimons (2015). Both commentaries raise interesting points, and although both teams clearly put a lot of work into their papers, the bottom line is this: our research sets the record straight that median splits are perfectly acceptable to use when independent variables are uncorrelated. The commentaries do a good job of furthering the discussion to help readers better develop their own preferences, which was the purpose of our paper. In the final analysis, neither of the commentaries pose any threat to our findings of the statistical robustness and valid use of median splits, and accordingly we can reassure researchers (and reviewers and journal editors) that they can be confident that when independent variables are uncorrelated, it is totally acceptable to conduct median split analyses.

Keywords: Median split; Median-split; Dichotomization; Categorization

Introduction

In Iacobucci, Posavac, Kardes, Schneider, and Popovich (2015), we had documented the enormous popularity of median splits, in consumer research, psychology, and numerous other fields. We had acknowledged the traditional concerns regarding median splits regarding the loss of information and resulting power. More importantly, we sought to investigate the extent to which the more recently expressed concern about median splits held true, that using median splits may give rise to Type I errors. Our approach was more comprehensive than that of the literature to date because we designed full simulation studies rather than relying on an anecdotal data set.

We found that in the presence of multicollinearity, median splits could indeed result in Type I errors, though the effects were often negligible. The results of our studies were clean and unambiguous; in the absence of multicollinearity, median splits do not create misleading results. We made it clear that the findings were not attributable to the use of an ANOVA vs. the regression model, but rather due to the presence or absence of multicollinearity. If a researcher is running an experiment, such as a typical factorial (or other orthogonal design), then letting a median split serve as a factor is completely legitimate.

In our Discussion section, we mentioned that median splits were not likely to have caused problems in published articles and we explained why. We also explained that our statistical results hold for naturally occurring or experimenter-created groups. We demonstrated that our results held even in the presence of extremely non-normal distributions (e.g., quadratic,

☆ The authors are grateful to the Editor, the original submission Area Editor and the Research Dialog Area Editor, and the teams of commentators for their respective roles in this Research Dialog.
* Corresponding author.
E-mail addresses: Dawn.Iacobucci@owen.vanderbilt.edu (D. Iacobucci), Steve.Posavac@owen.vanderbilt.edu (S.S. Posavac), Frank.Kardes@uc.edu (F.R. Kardes), mattschneids@gmail.com (M.J. Schneider), Deidre.Popovich@ttu.edu (D.L. Popovich).

http://dx.doi.org/10.1016/j.jcps.2015.06.014
1057-7408/

natural log, bimodal, and uniform). We also entertained the notion of two median splits in a single study, that while such a practice might not seem advisable, in truth, it may well be less problematic than one might first think.

Finally, in our paper, we stated repeatedly and quite clearly that we were not intending to persuade researchers who like their continuous variables to begin dichotomizing. Rather, our study provides support for researchers who like working with median splits due to the beauty of their parsimony, and the ease with which they may be communicated. The findings of Iacobucci et al. (2015) support those researchers in their preferences for median splits.

Although we suspect that the commentaries as a whole would have added more value if a psychologist or consumer researcher favorable to median splits wrote one of the commentaries, the upside of our having received commentaries written by two teams with a track record of opposition to median splits is that readers can be confident that any possible objection to our results has been generated. Thus, taken together, we are delighted with the commentaries by Rucker, McShane, and Preacher (2015) and McClelland, Lynch, Irwin, Spiller, and Fitzsimons (2015), and this opportunity to clarify and reify the fact that median splits are a perfectly valid, and extremely useful analytical tool for researchers. The commentaries offer a range of opinions, from concepts that are absolutely correct on one hand (e.g., Rucker et al.'s points that, all other things being equal, Type I and Type II errors rise and fall in opposition, and that regressions in and of themselves do not support causal statements, or McClelland et al.'s remarks that a median split is known to reduce power), to erroneous on the other (e.g., McClelland et al.'s claim that our simulations were incomplete or incorrect with technical errors, and their false equivalence logical fallacy when appealing to the ESP literature). The styles of the two commentaries are rather different, with the first being deeper and focused, whereas the second is broader. We address the arguments in each commentary in turn, concurring and clarifying as appropriate.

Commentary by Rucker, McShane, and Preacher

Rucker et al. (2015) offer a number of well thought-out arguments about the treatment of continuous and median split variables, and we feel that when readers compare their perspective with ours, our goal of moving the field toward a more nuanced understanding of median splits is greatly facilitated. Recently, the field had been told to reject median splits because of concerns regarding Type I error based on an overly broad conclusion derived from a highly artificial and constructed data set. The main purpose in our paper was to show that concerns with Type I error are, in fact, bounded within certain methodological contexts. In perhaps the most common experimental scenario in which median splits are used, wherein one factor is an experimental manipulation and the other factor is a median split, our paper shows that Type I error is not increased by median splits. Rucker et al. seem to concede this point, but nevertheless having concerns with median splits, change the focus of the debate from Type I to Type II error. Although we see some of the issues raised in Rucker et al.'s commentary differently, we feel that they make a number of well-reasoned arguments regarding Type II error that help to increase the sophistication of the discussion regarding median splits.

Type I and Type II errors

If this discussion is to revolve around Type I and Type II errors, let us review the basics. Fig. 1 depicts two normal distributions. The distribution at the left, drawn in a solid line, is the distribution around the null hypothesis population mean, $\mu = 0$, the one being tested. The distribution at the right, drawn in a dashed line, is a distribution around a different population mean, $\mu = 1.5$. In the left-hand distribution, the critical regions are drawn at ± 1.96 for a Type I error rate of $\alpha = 0.05$ in a two-tailed test. If the null hypothesis is true and a calculated z exceeds ± 1.96, the researcher would make a Type I error. Type II errors reflect the opposites—the opposite reality and the opposite decision. If the null hypothesis is not true, but z falls short of ± 1.96, the researcher does not reject the incorrect null, committing a Type II error, the likelihood of which is depicted by the shaded area labeled β. Recall the standard label of the probability of committing a Type I error is α, and that for a Type II error is β.

Students of statistics are taught that there is an inverse relationship between Type I errors and Type II errors. It is not a simple relationship, as if α and β sum to some constant value, in part because a Type I error can only occur if the null hypothesis is true, and a Type II error can only occur if the null hypothesis is false, and of course these conditions cannot both hold simultaneously. In Fig. 2, we depict the two distributions with the use of a more conservative $\alpha = 0.01$ In changing critical values from 1.96 (for $\alpha = 0.05$) to 2.58 (for $\alpha = 0.01$), the Type I error probability has decreased. Note that the Type II error, the size of the area under the curve labeled β has increased in Fig. 2 compared with Fig. 1. Figs. 1 and 2 illustrate how the relationship between α and β holds, that as an α-level decreases, the β probability increases. (Conversely, as α increases, say from 0.05 to 0.10, then the likelihood of a Type II error, β, decreases.)

Given that basic frame, let us now add the notion of power to the mix, recall it to be the likelihood of rejecting the

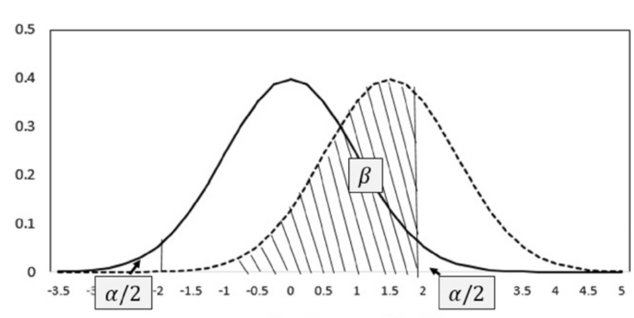

Fig. 1. Standard normal distribution, $\alpha = 0.05$, critical z = 1.96.

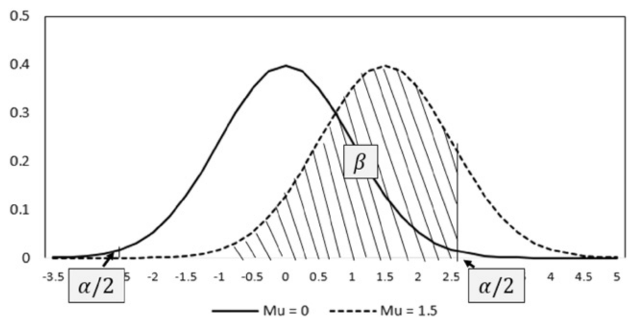

Fig. 2. Standard normal distribution, α = 0.01, critical z = 2.58.

null hypothesis when the null is indeed false. Power is denoted $(1 - \beta)$ as the probabilistic opposite of the likelihood of committing a Type II error, β. Power captures rejecting the null when we should, and β reflects not rejecting the null when we should. Power is a reflection of many factors, some of which are under the researcher's control and others of which are less so. For example, for many statistical tests, power is enhanced when the effect size is large, e.g., when the observed mean is very different from the hypothesized mean under the null, or when two sample means are very different from each other. For many statistical tests, power is also a direct function of sample size. Even without resorting to so-called big data, most researchers gather samples that are as large as they can, given participant pool hours, or budgetary or temporal constraints, at least sensibly within reason, e.g., as large as the size of samples that they have used in the past or that approximate the norm in the literature reporting studies of similar phenomena. (Almost never do researchers report a power calculation in an effort to estimate requisite sample size prior to collecting the data, but obviously this step would be useful.)

What's a researcher to do? Larger samples enhance power and decrease Type II errors

The reason it is important to understand the relationships among Type I errors (α), Type II errors (β), and power $(1 - \beta)$, is that we all know that researchers can make trade-offs among them. Say a researcher were to use the traditional Type I error rate of $\alpha = 0.05$. And say the researcher desired to use a median split but was concerned about the 20% reduction in effect size and power. The researcher can compensate by drawing a somewhat larger sample to help enhance power. If power $(1 - \beta)$ increases, the probability of a Type II error, β must decrease. Thus, by increasing sample size, the researcher achieves the excellent goal of overcoming the reduction in power caused by the median split, thereby accordingly also decreasing the likelihood of making a Type II error (i.e., the concern of Rucker et al.).

Our purposeful focus on Type I errors

Iacobucci et al. (2015) had focused on Type I errors because that had been the status quo lingering concern in the literature. As presented in Iacobucci et al. (2015), the traditional concerns with median splits had been their resulting smaller effect size and loss of power. The more recent concern had circulated that the use of a median split might actually increase Type I errors. Given that the first concerns had been known for decades, and many researchers had been willing to pay the price of slightly reduced power, our paper naturally focused more on the newer issue of possible Type I errors. Our simulations demonstrated that without the accompanying problem of multicollinearity, median splits resulted in no more Type I errors than a regression on a continuous variable would, and no more Type I errors than their nominal level, namely α.

In focusing their commentary on Type II errors, Rucker et al. essentially cede the issue of Type I error equality. That is, there is no more to be said about Type I errors in the presence of median splits because our studies showed that without multicollinearity they do not occur. Yet turning to the issue of Type II errors is reminiscent of the old debate surrounding theory testing and effects estimation. Type I errors are intimately connected to theory testing. Calder, Phillips, and Tybout (1981) made the clear case that science is about theory testing and building, even more to the point, "*falsification procedures*" (p. 198, italics in original), per Popper (2002a). In the statistical machinery that is used for scientific progress, falsification in theory testing and building is quite pointedly directed at constructing a null hypothesis, which the mechanics of statistical theory attempts to reject. A null hypothesis is posited and tested, and if rejected, that premise in the overall argument is rendered invalid, bringing down the larger theory, of which the hypothesis was a part. Of course the complexity of bridging the philosophical and the statistical is that a theory is comprised essentially of alternative, not null, hypotheses, whereas the statistical machinery is conducted on null hypotheses. Specifically, a valid symbolic logical argument is constructed with the premise, $p \rightarrow q$ (that is, if p then q), followed either by affirming the antecedent (and p, therefore q), or by denying the consequence (not q, therefore not p). In statistics, we translate p and q as follows: "If the null hypothesis is true, then in our data we shall see $\mu_1 = \mu_2$. In the data we learn that $\mu_1 \neq \mu_2$. Therefore we conclude the null hypothesis is not true; we falsify the null." A theory often postulates expectations of differences between groups so rejecting a null of equality offers a tentative step toward substantiating the theorizing. A statistical test is conducted, and if the null is rejected, the scientist evaluates that premise in the theoretical argument (which served as the alternative hypothesis in the statistical machinery) as "reasonable, so far." Subsequent stages of progress require the scientist to modify the theory and resubmit the new premises into an even stronger theory, to be tested subsequently. Naturally with a focus on scientific progress, theory testing and building, researchers and editors and journal reviewers care about whether that scientific task can be achieved properly; hence an emphasis on Type I errors. Also note of course that Calder et al. (1981, p. 199) further mention that theory testing is foundational to any subsequent efforts of theory application and effects estimation.

A researcher's freedom of choice

One final concern that arose as we read Rucker et al. (2015) was their comment that researchers are not just hurting themselves (e.g., with lack of power) by using a median split but that somehow they are hurting the field. Of course one cannot argue with not wanting to hurt the field. Unfortunately in this case, that goal is being attached to a choice of analysis that is a matter of simple preference. We believe, indeed for the good of the field, that researchers should be allowed to retain their free will and freedom of choice among their statistical techniques. Science is, after all, empirical. Let the data speak. If median splits were not fruitful, or if they led to irreplicable results, researchers would have ceased using them decades ago. If other researchers find regressions more enlightening, they should continue to use their analytical tool of choice. Our position is that a paternalistic stance against median splits is simply not empirically justified. Researchers should be able to choose from among their preferred models and analytical tools. To draw a parallel, behavioral researchers who currently prefer regression could be reminded that regression is but one extremely simple form of a structural model that is subsumed into the fuller, richer structural equations modeling framework (Aaker & Bagozzi, 1979; Anderson & Gerbing, 1988; Fornell & Larcker, 1981; Iacobucci, 2009, 2010; Iacobucci, Saldanha, & Deng, 2007). In the same way that regression is a perfectly valid technique for many applications, and should not be judged as uniformly inferior to more comprehensive structural equations modeling, median splits are also a valid technique in the methodological contexts we have specified.

Finally, several minor points need clarification:

- In their section headed "Costs," the paragraph beginning "1. Replicating prior research," note that we investigated the behavior of one median split, not two simultaneously.
- Their next point, "2. In their first simulation" is a bit misleading—we demonstrate that no spurious effects occur if there is no multicollinearity, and we show that in the presence of multicollinearity, the spurious effects are tiny.
- Shortly thereafter in a paragraph headed "Increased Type II error," the statement "A similar loss of power from 80% to just under (sic) 60% occurs," should say "just over 60%."
- In a section headed "Perceived benefits," Rucker et al. invoke Occam's razor in support of regression over median splits, but their statement is incorrect in saying that a median split requires more parameters than regression. In addition, statisticians and modelers assess parsimony and robustness on several dimensions, one being the number of assumptions required by the model, and obviously regression is more demanding on that score.
- In the section on graphs, we too are fans of data visualization and welcome its return with "big data," but plots will not likely be replacing summary statistics in journals in the near term. (Even so, we liked Rucker et al.'s Figs. 1 through 3 and believe that they convey essentially the same theoretical information.)
- It might also be mentioned that Rucker et al. are somewhat overstating the certainty that researchers may attach to measured variables. For example, they say that combining an observation that is very high on some variable with one that is somewhat high on the variable is "simply incorrect"; and three times they use an example of saying that attitude scores of 6.5 and 9.5 are different. Perhaps they are, but perhaps they are not; the distinction overstates the accuracy and precision with which most unobservable traits and constructs are measured. It ignores a vast psychometric literature (Allen & Yen, 2001; Anastasi & Urbina, 1997; Carmines & Zeller, 1979; DeVellis, 2011; Kerlinger, 1999; Nunnally, 1978; Pedhazur, 1991), e.g., beginning with Classical Test Theory positing an observed data point as a function of a person's true score plus error: $X = \tau + \epsilon$ (with ϵ contributing variability over trials within a person, or over persons within a sample). Even if a researcher is not familiar with measurement theory, note that the model specification is much like that in classical statistics: $X = \mu + \epsilon$ (where ϵ is most frequently considered to be contributing variability over persons within a sample).
- The Gelman and Park (2009) idea of splitting a variable into three parts is fine, but note the relative efficiencies being discussed compare their dropping the middle part of a distribution of data, which reduces the effective sample size. Nowhere have we advocated deleting observations. Related to this point is the concern that Rucker et al. use the term "efficient" in quite a number of locations throughout their commentary. Yet a proper assessment of statistical efficiency (related to an asymptotic standard error (ASE), or a uniformly minimum variance unbiased estimator (UMVUE), or even the efficiency of an experimental design) requires more information than the commentators have provided, e.g., regarding distributions and sufficient statistics (cf., Shao, 2003).
- In the section labeled "Multiple measured variables," we actually agree that sometimes ignoring a little bit of multicollinearity can be acceptable. Even in the case of median splits, the damage that multicollinearity creates is minor. But we were a bit bemused that Rucker et al. were fairly cavalier about multicollinearity but seem to take a hard stand on median splits.
- The title to their Table 1 claims numerous statements, none of which we made. We have said that if predictor variables are uncorrelated, the use of a median split is valid.

In sum, Iacobucci et al. (2015) delineated methodological contexts in which there are no concerns of Type I error risk being inflated by median splits, and in other contexts in which median splits can increase Type I error risk, the threat is minute. Accordingly, the Rucker et al. (2015) commentary left the grounds concerning Type I errors for an altogether different direction of concern over Type II errors. Even students in introductory statistics classes are taught that on occasion, researchers might weight the penalty of a Type II error more heavily than a Type I error, and naturally we agree. Yet the majority of our collective attention is in the journals and it is easy to argue that in the falsification philosophical orientation of theory building and testing, scientific progress and publishing, a nearly exclusive focus on Type I error dominates, a perspective with which we suspect Rucker et al. (2015) would agree, even if perhaps ruefully. Whether one's research focuses on Type I or Type II errors, we trust our colleagues to exercise common sense.

Commentary by McClelland, Lynch, Irwin, Spiller, and Fitzsimons

McClelland et al.'s commentary covers a lot of ground, and, as with Rucker et al., we hope that their thoughts, as well as our response, will help the field move toward a more accurate understanding of the properties of performing median splits. We very much appreciate the work of Irwin and McClelland (2001) in clearly explicating the notions of interaction tests in regression, in the vein of work by Jaccard, Turrisi, and Wan (1990) and Aiken and West (1991). Similarly, Spiller, Fitzsimons, Lynch, and McClelland (2013) showed how to use the broader logic of confidence regions (geometric-based confidence intervals, also see Hayes & Matthes, 2009), and Fitzsimons (2008) performed a valuable service by reminding researchers of the utility of regression when considering their analytical choices. Given these researchers' work advancing and communicating the benefits of regression-based analyses, it is not surprising that they would advocate regression approaches in lieu of median splits. By comparison, our research came from a point of empirical inquiry, not a position of advocacy, pro or con, with respect to continuous or median split variables. Our findings indicate that although using regression as prescribed by the commentary team is fine, the empirical evidence we found simply does not support their prescriptive advice to unconditionally reject median splits. Our research provides the remedy to swing the pendulum back to center; a moderate position from which a researcher is free to choose to analyze a continuous variable or a median split variable.

In their efforts to sway researchers from considering median splits, McClelland et al. offer a variety of arguments in a shotgun approach, hoping some criticism will hit the target. Although some of their exposition is pretty obviously not useful (e.g., their fallacious opening with reference to ESP, or claiming that they are "quite sure" that a departed statistician would agree with them), many of their points are worth discussing. Their issues come in one of three forms: 1) statements about statistical relationships that are interesting but irrelevant, and because they have no bearing on the median split issue, we will call these distractors or red herrings, 2) statements about statistical relationships or our research that are incorrect and that we will clarify, and 3) several statements are comments with which we agree, and we will happily point out our concurrence. The McClelland et al. commentary offers the peripheral cue of sheer number of arguments, yet, so that readers are not misled, we will patiently trod through them to explain their sources of confusion or error. Their commentary frequently uses expressions such as "We strongly disagree," and we will not because we need not: we are not offering opinions in matters of preference, we are presenting facts and mathematical and statistical findings, thereby elevating the discussion to matters of science rather than the politics or sociology of science. For example, in several places, McClelland et al. refer to "technical errors," "discrepancies," or "serious problems" with our simulations, and their criticisms are incorrect. We will show where, how, and why their criticisms are wrong.

One size does not fit all—regression is not an omnibus cure

Before addressing McClelland et al.'s specific points, we begin with a general consideration of regression. It is important to recognize at the outset that there is nothing magical about the regression model, and that, in and of itself, using regression does not make results better. Regression is a wonderful technique for many applications, but is not appropriate for all types of research questions. Accordingly, it would be problematic for our field to develop a norm in which researchers feel compelled to use regression when other procedures could potentially yield more meaningful results.

Consistent with the notion of the limitations of regression, Rosenthal and Rosnow (1991) maintain that the problems associated with the use of multiple regression are "too infrequently recognized" (p. 558). For example, the "magnitude, sign, and statistical significance of each regression coefficient *depend entirely on exactly which other predictor variables are in the regression equation* (italics in the original) …. The p values printed for the overall R and for the regression coefficients of each predictor variable are the same whether the particular battery of predictors was planned as the only battery of predictors to be employed (almost never the case) or whether some algorithm was used to pick out the best set of k predictors from a larger set of possible predictors (almost always the case). The printed p values are accurate only in the former (unlikely) case; they are not accurate in the latter (common) case" (Rosenthal & Rosnow, 1991, pp. 558–559). It is unclear how to compute p values in the latter case, and shrinkage, or weaker effects in replication studies are common (Moses, 1986).

Collinearity or multicollinearity—high correlations among predictor variables—also produces interpretational problems (Moses, 1986). Rosenthal and Rosnow (1991, p. 559) say that "multiple regression approaches to inferring causality can yield results that are very misleading" (see also Cook & Campbell, 1979). Other scholars concur: the "Interpretation of the [regression coefficients] from the results of a simultaneous regression of [highly correlated independent variables] that ignores their multicollinearity will necessarily be misleading" (Cohen, Cohen, West, & Aiken, 2003, p. 98).

Any research approach or analytical model has its strengths and weaknesses—this truism also holds for regression. Allison (1999, p. 63) says, "Multiple regression is designed precisely for separating the effects of two or more independent variables on a dependent variable when the independent variables are correlated with one another, but there's a limit to what regression can do …. Multicollinearity doesn't have to be so extreme to cause problems, and unfortunately, those problems often go undetected." For example, it is known that when there is multicollinearity, "small differences in [the predictors'] bivariate relationships with the dependent variable get magnified into large differences in the regression coefficients" (Allison, 1999, p. 144). Multicollinearity renders regression coefficients unstable, and less robust to minor errors or departures from the assumptions of the regression model.

Interpretation of regression results in the face of multicollinearity is ambiguous at best. Allison (1999, p. 63) states that if X_1 and X_2 are correlated, "it's unlikely that either coefficient would be statistically significant. On the other hand, it's entirely possible that if we ran the regression with only the [X_1] variable or, alternatively, only the [X_2] variable, we might find large and highly significant coefficients for both variables. This would tell us that one or both of these variables had a substantial effect …, but we couldn't say which one."

Berry and Feldman (1985, p. 40) explain: It is known that multicollinearity makes it "impossible to separate out the effect" of one predictor. This effect is due to the increase in the size of the covariances between the parameter estimators. In a simple model with just two predictors, "the correlation between estimators b_1 and b_2 is $-r_{X_1X_2}$, the inverse of the correlation between the two independent variables …. This, of course, implies that conclusions drawn about the relative impacts of the two independent variables on the dependent variable … are very shaky" (Berry & Feldman, 1985, p. 42).

Not only is it the case that multicollinearity can create ambiguity in interpretation, it is also well-known that multicollinearity diminishes power. To see how, recall the standard error for a regression coefficient for predictor X_i is defined as follows (e.g., see Berry & Feldman, 1985, p. 13):

$$s_{b_i} = \sqrt{\frac{\sum_{j=1}^{n}\left(Y_j - \hat{Y}_j\right)^2}{\sum_{j=1}^{n}\left(X_{ij} - \bar{x}\right)^2\left(1 - R_i^2\right)(n-k-1)}}$$

where R_i^2 is the squared multiple correlation when predicting variable X_i from all the other predictors in the model. To the extent that a variable X_i is multicollinear with one or more of the other predictors or any linear combination thereof, then R_i^2 will be large, and $(1 - R_i^2)$ will be small. A small $(1 - R_i^2)$ in the denominator of the equation above ensures that the standard error for variable X_i, s_{b_i} will be large. Given that the standard error serves in the denominator of the t-statistic that tests whether the contribution of variable X_i to the prediction and understanding of the dependent variable is different from zero, the size of the t-test is diminished. The bottom line here is that multicollinearity increases a standard error and thereby decreases a t-test, thus decreasing the power of the t-test and the likelihood of a conclusion that the predictor is significant.

Iacobucci et al. (2015) had demonstrated that a median split coupled with multicollinearity similarly can be problematic. It is easy to understand that multicollinearity is the main culprit, given that it causes analytical problems, per above, even without a median split. Multicollinearity was also shown to be the main culprit in Iacobucci et al. in that in its absence, there were no problems with median splits regarding Type I errors.

McClelland et al.'s specific concerns

We turn now to the specific concerns raised in McClelland et al.'s (2015) comments. They first offered four statistical claims, they say, in a "nutshell," then they considered non-statistical issues, next they discussed additional statistical considerations relating to Type I and Type II errors, and then they raised questions regarding our simulations. To assist the reader, we proceed by mapping our answers sequentially to their questions.

McClelland et al.'s statistical argument a

McClelland et al.'s first concern regards the measurement error that results when a median split is used. Regardless of whether one analyzes a continuous or median split variable, there is of course measurement error in presumably any variable, any operationalization of a latent construct, whether that error is recognized and separated out, as in a structural equations model, or ignored and left confounded, as is done in a regression model's lack of fit error term. The important point that our simulations make is that no bias toward Type I error occurs when a median split is conducted in the absence of multicollinearity.

McClelland et al.'s statistical argument b

McClelland et al. (2015) refer to a thought experiment of "scrambling" data. We are not certain about this point being made regarding random in, random out. However, we certainly agree that if a researcher were to scramble data, then it would not be surprising if the results were random and less than optimally interpretable. We have demonstrated empirically when median splits do and do not lead to increased Type I error risk. The important point, as we discuss in more detail later, is that when multicollinearity is absent, McClelland et al.'s concerns about unscrupulous researchers picking the technique more complementary of their theories cannot apply because the continuous and median split results converge.

McClelland et al.'s statistical argument c

McClelland et al. next raise an issue regarding nonlinearity. Most researchers, behavioral and quantitative, users of median splits or regressions, do not tend to track down functional forms, be they step functions, quadratic or other power terms, or the like. Almost never do researchers show interest in, or more importantly, theory for, points along curves. And almost never do researchers using regressions report that they tested the underlying assumptions, such as homoscedasticity, linearity, and normality (of ϵs for the statistical tests of βs to be valid). Nor are we proposing that these basic tests be required, primarily because we can grant that regression is a robust technique. Of course, a median split will be all the more robust (because it doesn't require an assumption of linearity, and the homoscedasticity assumption in regression simplifies to homogeneity of variances in just two groups rather than at all points along a continuous predictor variable). A median split focuses on two points in space; a regression focuses on a slope, which is defined by two points in space. A geometric entity of n-dimensionality may be defined on n + 1 (noncollinear) points in space. Thus, a one-dimensional line needs two points for definition. Those points may be contained among others, as in a regression line, defined as $b_0 + b_1X_1$, e.g., one point having coordinates $[(X_1 = 0), b_0]$ and another point having coordinates $[(X_1 = c),(b_0 + b_1c)]$, or they may be points that represent two means, one in a median split low group and the other in a median split high group.

McClelland et al.'s statistical argument d

McClelland et al.'s final argument in this section is that median splits are sample dependent, that is, samples will be different from each other with regard to their distributions. We see this argument as less than compelling as a criticism, indeed we would take the claim farther than they and state that every statistic calculated based on observations of any sample will be sample dependent, including, for example, the summary statistics used in regression. It is true that if two different samples are split into "low" and "high" then the "lows" in one group might be nearly as "high" as the highs in another group. For example, if a company tracked customer satisfaction scores in the U.S., they would have some range of values, from low to high. If the company tracked satisfaction scores in a country known for its polite acquiescence, those scores would also have some range, from low to high, where the high scores in the U.S. might be among the lower scores in the second country. This issue has nothing to do with categorization. If the satisfaction scores were not dichotomized, then the two continua might overlap, but the U.S. continuum would fall lower than (or to the left of) the second country's continuum. The bottom line is that this argument is a bit of a red herring—the median in a sample is exactly that; a statistic that describes a given sample. Similarly, a regression analysis of a sample of data will not characterize another sample of data. As with other statistics, it is important to be mindful when interpreting data and not to claim unwarranted external validity. This admonition is appropriate in any context in which data is collected from any sample.

McClelland et al.'s nonstatistical issues

McClelland et al.'s nonstatistical issue 1

McClelland et al.'s first nonstatistical issue is predicated on their claim that we argued that median splits should be considered as a viable option because they are popular. This is not quite what we said. In Iacobucci et al. (2015), we *reviewed the literature* on why median splits are often attractive to researchers, and wanted to catalog their popularity to frame the importance of furthering our knowledge and sophistication regarding when median splits are, and are not, appropriate. Our contribution is delineating the answer to this question.

McClelland et al. then forward the case of a *Science* paper investigating the relationship between income and cognitive functions, a case with which we see at least three concerns. First, from our research, we know that empirically the only way that the median split results could have yielded significance when the continuous variables did not would have been if there had been a great deal of multicollinearity, thus, we may infer from those findings that there had been multicollinearity in their data. Second, we do not know exactly how the analyses were conducted in those studies (nor do McClelland et al.). The articles state that Mani, Mullainathan, Shafir, and Zhao (2013) wished to dichotomize their income variable since, unlike attitudes and other constructs that form the mainstay in consumer psychology that are very often at least roughly normally distributed, economists will verify that income is consistently skewed. As the commentators in the *Science* debate, Wicherts and Scholten (2013) treated income as continuous, they indicated an extreme unbalancedness in the design which is known to be detrimental to statistical power (one sample size was approximately only 40% of the size of the other sample), and then they used an unspecified transformation on the dependent variable, which they said created a platykurtic distribution (like a normal, but flatter); i.e., a distribution shape closer to uniform that might very well nullify regression results. All of these issues have nothing to do with the dichotomization. Finally, of course, any of the Mani et al. or Wicherts and Scholten analyses may have been done improperly. Given the ambiguity in their data and analyses, we find the example to be not particularly enlightening for informing our findings. We ran simulations precisely to study the phenomena fully, across a 375-cell immense factorial, rather than focusing on only one anecdotal data set.

McClelland et al.'s nonstatistical issue 2

As with the first issue, much of McClelland et al.'s critique is of ideas that were not present in Iacobucci et al. Our point here is that when researchers are interested in variables that are typically thought of as categorical, median split analyses are a good match for researchers' conceptualizations of their research problem. For example, if a paper compares behavioral differences between liberals and conservatives as defined by a median split, those authors are not claiming that there are not gradations of liberalism and conservatism, or that the numbers of liberals and conservatives in the broader environment are equal, but simply that individuals categorized as liberals versus conservative behave differently. That is, we agree with McClelland et al. that a continuous measure of a person's degree of leaning toward liberalism or conservativism could be very useful, but we can also state unequivocally that a binary conceptualization is obviously very useful too, e.g., in predicting election outcomes, when a perhaps continuous political opinion will manifest itself in a binary behavioral outcome, namely a vote for a Democratic or Republican candidate (cf., the political science literature).

McClelland et al. argued that most researchers consider the need for cognitive closure construct to be continuous and not categorical, and they cited two articles by Arie Kruglanski. However, they failed to cite a representative sample of Kruglanski's path-breaking work on the need for cognitive closure (e.g., Kruglanski, Dechesne, Orehek, & Pierro, 2009; Kruglanski et al., 1997). Kruglanski is an extremely productive and influential researcher in psychology (e.g., he has 23,000 citations in Google Scholar), and he routinely uses median splits in over 95% of his articles on the need for cognitive closure. A careful reading of Kruglanski's work reveals yet another benefit of median splits: Performing median splits on the measured need for cognitive closure permits comparison to studies using situationally manipulated need for cognitive closure. Kruglanski frequently finds a similar pattern of results for both types of variables.

Finally, it is unclear why McClelland et al. feel that presenting data as a median split would preclude researchers from examining nonlinear relationships. Descriptive statistics may be computed for continuous or median split variables, plots may be examined, and model assumptions tested. Indeed, should there be obvious nonlinearities, researchers would be better served by a median split analysis than a linear regression conducted on the continuous variable (the regression requires the assumption of linearity to hold, and while the regression model is rather robust, the assumption is not required of the median split analysis).

McClelland et al.'s nonstatistical issues 3 and 4

Third, McClelland et al. (2015) state that regression is easy to use and that creating interaction plots is easy. We agree. Regression models are fit, and parameter estimates used to compare slopes and to generate points to plot, and multiple primers abound (cf., Aiken & West, 1991; Bauer & Curran, 2005; Berry & Feldman, 1985; Dawson, 2014; Dawson & Richter, 2006; Fraas & Newman, 2005; Hayes & Matthes, 2009; Irwin & McClelland, 2001; Jaccard et al., 1990; Johnson & Fay, 1950; Potthoff, 1964; Spiller et al., 2013). Yet by comparison, the steps for a median split analysis are simply: 1) find the median, 2) split the data at the median and 3) run the ANOVA (or regression).

In Iacobucci et al., we had also pointed out that a plot probing an interaction in regression yields very similar information to a plot of means after a median split variable is created. In Fig. 3, we depict the relationship: regression interactions are often plotted at $+1.0$ and -1.0 standard deviation. In a normal distribution, the quartiles fall at $z = -0.67$, $z = 0.00$, and $z = 0.67$ standard deviations, and the means above and below $z = 0.00$ fall at ± 0.8 standard deviation units, incidentally close to $+1$. Researchers are comparing essentially the same information.

McClelland et al. seem to imply that there exist researchers without the sophistication to conduct regressions or regression plots. We give our colleagues more credit than that—researchers know ANOVA and regression, and they continue to learn how to analyze data using optimal methods for their projects. The point of our paper had been to document that recent criticisms of median splits are overstated, and that median splits are acceptable when predictors are uncorrelated. Accordingly, our position is that researchers can decide for themselves which technique is more preferred in such a situation.

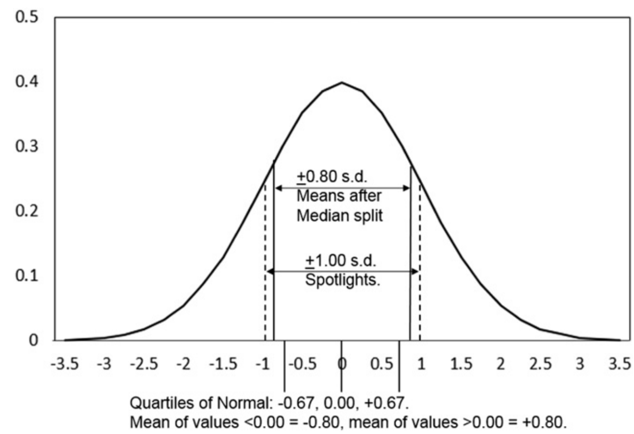

Fig. 3. Information contrasted via two means after a median split compared with regression lines drawn to probe interactions.

McClelland et al.'s (2015) fourth point, about parsimony, is essentially the same as their third, although they try to strengthen their case by name-dropping famous scientists. We find their invoking Galileo particularly interesting considering that Galileo was trying to get people to understand that the earth revolved around the sun, using facts (much like trying to convince people that median splits may be used, using facts) and he faced opposition based on polemics and opinion.

Regarding parsimony, we suggest, as a thought experiment, adopting the perspective of a statistical modeler for a moment. A modeler would see regression as more complex and less parsimonious than a median split in part because regression carries more assumptions. Yet we can grant that this complexity does not constitute a reason to impose a blanket restriction on conducting regression analyses. Finally, contrary to McClelland et al.'s definition, most researchers and the literature would characterize median splits as more parsimonious (cf., Fitzsimons, 2008).

McClelland et al.'s second section on statistical issues

McClelland et al. deny the fact that median splits are more conservative

McClelland et al. open this section by arguing that median splits reduce power, a fact known in the literature for decades, and something we discussed both in our original paper and in our response to Rucker et al. We stand by the statement, per the literature, that using a median split is conservative. McClelland et al. then mention Type II errors, and do not feel that trading off a reduction in power is worthwhile without a commensurate advantage provided by median splits. This is a perfectly acceptable preference, but it is pseudodiagnostic regarding our findings that median splits may lead to Type I errors when multicollinearity is present, but not when absent.

The notion that median splits provide another way to do an analysis, which would allow researchers to choose whichever technique is most favorable to their theories and thereby increase the field wide incidence of Type I errors, is not a strong argument because it is not unique to median splits; there are always multiple ways to analyze any given data set. The focus in our original paper is on the statistical properties of median splits, and the findings we delineated are completely independent of concerns with researchers choosing analyses that best support their theories, nor did we ever endorse, state or imply such a practice. We believe that in practice, researchers seem to have a natural affinity for working with median splits or continuous variables and rarely bother to check the alternative analysis because they are choosing to work with the variable operationalization that best fits their substantive theoretical conceptualization. In fact, we all know some researchers who like median splits so much that they produce regressions and regression plots only if they must, at the request of reviewers and editors. We all also know some researchers who like regressions and regression plots so much that they would prefer to never conduct a median split.

McClelland et al. then offer a small, one-cell simulation to show that “picking and choosing” inflates a nominal Type I error rate of 0.05 to 0.08. No doubt Type I errors were inflated. However, this issue is another distractor. First, we have never endorsed or even mentioned engaging in their indefensible practice of running two correlations and choosing whichever was larger. Second, if either a continuous *or* median split variable (but not both) were tested on its own, the Type I error rate would approximate the nominal rate of 0.05. Let us make it *explicit* that McClelland et al. were *implicitly* testing two correlation hypotheses, not one. Thus it is not in the least bit surprising that they obtained inflated Type I errors. It is well-known that in such circumstances, researchers may wish to use a Bonferroni adjustment on the nominal $\alpha = 0.05$ by dividing it by the number of tests (in this case, 2), to obtain the proper significance level for each test so that the family-wise Type I rate is not increased. Third, their correlations were based on the same data, hence involving a covariance across the two statistical tests violating an assumption of independence in conducting the tests. Thus their Type I errors were boosted in part because the tests were conducted on redundant information, and boosted further still because of their “picking and choosing” procedure, which is supported by no statistical principle, theory, or practice. For all these reasons, their finding is simply an error and off point. Our simulations reported effects (e.g., βs) and significance levels (p values) for the continuous treatment of a variable in its entirety across the full design, and also reported those results for the median split treatment of a variable, again in its entirety across the full design. We did not make comparisons to select and go forward to report only one finding that exceeded another. It had not occurred to us that researchers might engage in the opportunistic analyses that McClelland et al. conducted.

The next topic offered in the commentary involves Bayes. It seems unlikely that McClelland et al. (2015) are suggesting that researchers begin using Bayes analyses, either fully Bayes or contemporary empirical Bayes (Berger, 1993; Casella, 1985; Kass & Raftery, 1995; Scott & Berger, 2010; Zellner, 1981). In any event, a Bayesian analysis was not shown, only the Bayes' theorem. Presumably the point of invoking Bayes was to demonstrate that large effect sizes can seem compelling. Similarly, McClelland et al.'s power analyses are fine, but again, their comments are not relevant to our core findings because they are not specific to the statistical properties of median splits. One way to increase power is to increase the strength of an experimental manipulation. The logical implication of McClelland et al.'s argument would seem to be that every experiment should be conducted with maximally strong manipulations, because otherwise power will decrease and the field will suffer. Yet researchers often choose to try to demonstrate effects in “minimal conditions” using non-intrusive manipulations, in part because reviewers often object to “heavy-handed” manipulations, and a more subtle manipulation speaks to the robustness of the effects and the sophistication of the theory. We prefer subtlety as more elegant, compared to an approach of “hitting them over the head.” Using McClelland et al.'s analyses to prescribe against the use of minimally intrusive manipulations would be on par with their arguments against median splits, and in both cases we feel differently. Let researchers decide—diversity in such advocacies should be valued.

Moreover, with respect to issues about effect sizes in general, it is important to note that it is typical in the maturation of any area within a discipline to see large effect sizes characteristic of main effects being established early on, and as researchers progress toward more refined studies to examine subtle interactions, by definition, such conditional effects will be smaller (Chow, 1988; O'Grady, 1982). Popper (2002b) refers to the calculus of probability, that in studying factors A and B, p(A) and p(B) will equal or exceed p(A&B) (cf., rules of conjunctive probabilities of Tversky & Kahneman, 1983), and says that with the growth of scientific knowledge, scholars want to work with theories of increasing content, thereby implying decreasing probabilities or effect sizes. Hence ironically, relatively small effect sizes can be indicative of mature theories and mature literatures. In any event, these issues are not germane to our demonstration that median splits do not lead to increased Type I error when independent variables are uncorrelated.

McClelland et al. deny how easily loss of power is offset

McClelland's second issue in this section also revolves around power, and the possibility of increasing sample size to increase power. First, a point of clarification: the reduction in power (already discussed in Iacobucci et al., 2015) on a correlation coefficient involving a median split is on the order of 0.80 times the size of the correlation on the continuous form of the variable (per Cohen, 1983), so when McClelland et al. (2015) use the figure 0.64, they are either referring to r^2 or an r in which median splits have been taken on both variables simultaneously. Next, it is well-known and quite standard to suggest increasing sample size to enhance the power of most statistical tests. Nowhere have we advocated that researchers collect some data, see if the desired results are significant, and if not, collect more data. Nor does this issue have any remote connection to median splits—researchers could do that using regressions if they wanted to. Further, as we have previously discussed, other factors such as the strength of a manipulation, as well as variability, also affect power. McClelland et al. seem to argue that any methodological decision that reduces power should be rejected. However, researchers often try to demonstrate effects with minimalist manipulations (as we noted earlier), as well as within increasingly diverse samples to assess generalizability (i.e., samples characterized by heterogeneity and high inter-subject variability). Both of these methodological choices serve to decrease power, and in our opinion neither are, in McClelland et al.'s emphatic wording, "unconscionable." We stand by our argument that sometimes researchers may wish to make the trade-off in power to conduct a median split.

Although their point is tangential, it is important to note that McClelland et al.'s notion that the median splits used by Kastrati et al. (2011) were dangerous is clearly incorrect. This study involved patients in cardiac distress receiving one of two often used treatments, and thus issues of experimental power in the study have no bearing at all on patient mortality. As we have stated multiple times, we believe that researchers should have the option of conducting median splits, but be under no compulsion to do so.

McClelland et al.'s questions regarding the simulations

McClelland et al. (2015) raise several questions about our simulations. They refer to "technical errors," "discrepancies," or "serious problems," but it is their criticisms that are incorrect. We need not simply protest, rather, we will show where, how, and why their criticisms are wrong.

The first set of concerns surround interactions. First, they criticize that we have generated data with a null effect for an interaction. That is true. We also generated the data to have a null effect for variable X_2. Recall that the question in the literature revolved around whether a median split on X_1 could impact Type I error rates on X_2 (or on their interaction). If one is to study Type I error rates, one must have null effects in the population. We will say more on this matter shortly (they return to the issue to reiterate the criticism). They also criticize the use of a multivariate normal to generate data for X_1, X_2, and Y, yet that is not quite what we did (in part to avoid this and related issues), which McClelland et al. seem to know (because in a moment we address that they criticized what we actually did, as well). Perhaps we had not explained our methodology clearly enough, however, if we used, call it method "A" to generate data, then McClelland et al. might find grounds to criticize method "A" but they cannot also criticize method "B" which was not used. Generally, a critique is more persuasive when it is internally consistent.

Next, they criticize that we offer two simulations, where we could have simply offered one, to obtain the results presented in our paper's first three figures. That's true. We explain how that came to be and readers can judge for themselves whether this issue is problematic.

Recall that the small data set that Maxwell and Delaney (1993) constructed had three variables: Y, X_1, and X_2. The concern in the literature had been that a median split on X_1 would not just diminish power on a test about X_1 but that it might create a Type I error in testing for an effect on X_2. Accordingly, in our original paper, the simulations examined the effects of a median split on X_1 on the tests for both X_1 and X_2. Those results came from our first simulation code. The reviewers, and the Area Editor and Editor of the original submission desired to see the effect on an interaction, and we were happy to oblige—great idea. To do so, we had to modify the first code to yield the second program. Why? The first code covering the simpler scenario of just two predictors, was programmed to obtain estimates of β_1 and β_2 as a function of the zero-order correlation coefficients: $r_{X_1 \cdot X_2}$, $r_{Y \cdot X_1}$, and $r_{Y \cdot X_2}$. Specifically, for standardized scores, the regression coefficients may be written (and were encoded in our program) as: $\beta_1 = \frac{(r_{Y \cdot X_1} - r_{Y \cdot X_2} r_{X_1 \cdot X_2})}{(1 - r^2_{X_1 \cdot X_2})}$ and $\beta_2 = \frac{(r_{Y \cdot X_2} - r_{Y \cdot X_1} r_{X_1 \cdot X_2})}{(1 - r^2_{X_1 \cdot X_2})}$.

When a third term is added to a regression, those equations are no longer elegant and it is standard to revert to matrix algebra to obtain the estimates (recall, $\hat{\beta} = (X'X)^{-1}(X'y)$), hence, that is what we did. The second program derives results on X_1 and X_2 as had the first program, and in addition, derives the results on the interactions. The first program is therefore indeed redundant, and its entirety (data generation and analysis) is subsumed in the second program. The reader can see this in our code in the second to last column; in the groups of commands that appear after the phrase, "After the …" and "Relabel the …" there are listed variable names *mbeta1* (mean for beta1), *mbeta2*, *mbetaint* (*int* for interaction), along with the median split forms (abbreviated *mdn*), thus all terms are present. Researchers can also note our header toward the bottom of the second column of code that says, "Next, we were asked to add the effect of interaction terms." Given that the first program is fully contained within the second, we could have offered the second program alone. However, for purposes of completeness, we included both programs in our Appendix.

Next, contained in that discussion, McClelland et al. raised a question about how the data were generated, demonstrating that they were confusing the method of testing for an interaction by creating a product term (e.g., $X_1 \times X_2$) with a probability-generating function from which the data were derived. We are happy to clarify this point as well. In statistical theory for this domain, the sources of variability in an ANOVA or regression for a model containing two main effect terms (X_1 and X_2) and their interaction are stochastically independent—each main effect might be significant or not, and the interaction term might be significant or not, in any given data set. For "significant or not" being two states of reality, and three predictor terms such as two main effects and an interaction, the number of combinations of significant patterns is 2^3. If a sample data set were instead generated by creating two main effects and multiplying those together, the population source for the interaction would be a derivative function of the other two effects, which is not what statistical theory and the model requires, nor does it reflect what is possible in real data sets. In addition, in our simulation, we had imposed an effect on X_1 to study the reduction in power of a median split on X_1, and we had imposed null effects in the population on the other main effect and the interaction because we were investigating, per the concern in the literature, whether a median split on one variable could create spurious effects on other terms in the model. Our approach was thus in keeping with statistical theory, and allowed us to demonstrate the lack of ill effects from a median split over a broader range of values on the accompanying second main effect and interaction terms. At the same time, we can certainly accommodate this concern of McClelland et al.—not to the data generation stage of the program but to the data analysis stage of the program—and here we agree that it would be more conventional to, and we should have, simply taken the product the way they suggest. To do so, the interested reader would make the following changes. In the third column of our code, after the "Replace" header, the command "interact = x[,4];" would be replaced with "interact = x[,2]#x[,3];" and further down that column, in the line that begins, "betatint;" the command "interact = x[,4]#x[,5];" would be replaced with a "interact = x[,3]#x[,5];" term. As it happens, the difference is immaterial, since both X_2 and the interaction term $X_1 * X_2$ were built to have null effects in the population (we reran the code with these changes and the results were identical, as McClelland et al. (2015) no doubt noticed when running our programs).

McClelland et al. (2015) return to their previous error in logic when they express concern that the data sets we studied were drawn from populations with no interaction effects. It is true that our population was defined to have null interaction effects, and again, here's why that was important. We agree that real data typically look different in this regard, at least we hope that most researchers are able to achieve significant (non-null) moderating relationships to illustrate boundary conditions. However, our research was not about finding some interaction between two theoretically relevant predictors and a consumer behavior criterion of interest. Our research was pointed to address the then-concern in the literature that a median split might create spurious results. A spurious result is and had been defined in this literature as a Type I error. Type I errors can only occur when the null hypothesis is true. Therefore, if we wished to test whether a median split on X_1 might create a Type I error on X_2 or an interaction term, then it needed to be the case that the null hypothesis was true for X_2 and the interaction term. Thus, to make progress in a literature concerned with Type I errors, we had to create conditions in which they might arise.

In their section, "Simulations versus derivations," McClelland et al. (2015) present the equations for estimating β_1 and β_2 as a function of the zero-order correlation coefficients as we have done above and as had been in our code (interestingly, given their concerns with our paper, their equations are based on a model that does not include an interaction term). It is certainly true that some of our results may be proven analytically. They used Mathematica in their online technical Appendix to manipulate some equations, but these derivations are actually easy to do by hand. For example, if we begin with the equations for β_1 and β_2:

$$\beta_1 = \frac{(r_{Y\cdot X_1} - r_{Y\cdot X_2} r_{X_1\cdot X_2})}{(1 - r^2_{X_1\cdot X_2})} \text{ and } \beta_2 = \frac{(r_{Y\cdot X_2} - r_{Y\cdot X_1} r_{X_1\cdot X_2})}{(1 - r^2_{X_1\cdot X_2})}$$

and focus, per our paper's demonstrations, on the fact that median splits create zero problems when there is zero multicollinearity (and minimal problems when there is), we can substitute 0 for $r_{X_1\cdot X_2}$ to derive the adjusted equations:

$$\beta_1 = \frac{(r_{Y\cdot X_1} - r_{Y\cdot X_2} \times 0)}{(1 - 0^2)} = r_{Y\cdot X_1} \text{ and } \beta_2 = \frac{(r_{Y\cdot X_2} - r_{Y\cdot X_1} \times 0)}{(1 - 0^2)} = r_{Y\cdot X_2}.$$

We know that if we take a median split on X_1, we can expect the reduction in effect size of about 0.80, so the new $\beta_1 = (0.80) r_{Y\cdot X_1}$ and because there is no X_1 term in the equation, $\beta_2 = r_{Y\cdot X_2}$ remains unaffected. These results are precisely what are shown in Iacobucci et al. (2015)—a median split on X_1 reduces its own effect size and does not affect results on the other variables whatsoever.

Next, regarding "Unrepresentative sampling" and "Estimates of β_1 in Study 1," McClelland et al. complain that we showed only part of our results. This concern is a little odd. As we explained in our paper, our full design yields results requiring a large table with 375 cells. Its inclusion would be cumbersome, so, as we stated in the paper, we presented a plot of a subset of results. The full design is analyzed later in the paper, and we have provided our code for interested readers if they wish to see the results in the full 375-cell table. Certainly in an extensive factorial, some combinations may be encountered less frequently than others, but the principle of a full simulation is to examine the phenomena over expansive ranges on orthogonal bases (per linear algebra). Furthermore, the commentators try to counter our vast coverage with the use of a single anecdote, one cell in the factorial, from which they espouse more general concerns. Note also that their choice involves high multicollinearity ($r_{X_1 X_2} = 0.7$ in their example), which is 0.7 greater than the level of multicollinearity wherein we mention that median splits perform fine. Similarly, the commentators believe we should not have aggregated the results, to which we might respond by saying that in regression, a plot of N pairs of X_i and Y_i data is aggregated to an intercept and a slope parameter. We reiterate that readers can easily regenerate the results with the code.

In the section, "Estimates of β_2 in Study 1," McClelland et al. then explore some relationships, which is in the spirit of our goal for a dialog about median splits. However, a number of questions may be posed about their assumptions. For example, they show relationships that occur when the extent of multicollinearity, $r_{X_1\cdot X_2}$ is less than the ratio of correlations of the predictors with the dependent variable, $r_{Y\cdot X_1}$ and $r_{Y\cdot X_2}$. That condition can certainly happen, but it does not have to, therefore their statements characterize an unknown and limited principle. Next, they explored results under even more restrictive conditions, namely when $r_{Y\cdot X_1} = r_{Y\cdot X_2}$. That condition can also certainly happen, but it tends not to, therefore the conclusions are of little utility because of the extreme restrictions used in producing the analyses. Finally, they make up an Eq. (4) to combine βs on the continuous variables to represent the β on the median split variable. This is creative math. The equation should state that the estimate is also a function of the correlation between X_1 and X_2 (as they do in their Appendix). Doing so is obviously important, because they use the equation to demonstrate some property when there is multicollinearity, to which we would say yet again: we are not advocating the use of median splits in the presence of multicollinearity. If instead, the correlations between the predictors is zero, $r_{X_1 \cdot X_2} = 0$, then their weight term, "w," reduces to zero, such that their Eq. (4) reduces to: $\beta^*_{Y_{2.1}} = \beta_{Y_{2.1}}$; that is, the median split estimate β^* is equal to the continuous estimate β, as we have maintained.

In the section on "Study 2," McClelland et al. criticize our study as covering "a very narrow set of conditions" even though our factorial covers three levels on sample sizes and five levels of effect sizes, which is 14 more conditions than the anecdotal observations offered by these and other authors. Then, in a footnote, McClelland et al. state that they could not reproduce our translation to mean differences, and we're happy to clarify—based on rudimentary regression relationships, we know $\hat{y} = \beta x = r_{Y,X} x$ for standardized variables. Plugging in some value for x, essentially z_x, is some number of X standard deviation units, say ± 1.0, ± 1.5 and ± 2.0, and using some value for $r_{Y,X}$ in our simulation, say 0.5, would yield predicted values of ± 0.5, ± 0.75, and ± 1.0 standard deviations on $\hat{y}$, for a total spread of 2(0.5) = 1.0, 2(0.75) = 1.5, and 2(1.0) = 2 standard deviation unit effects for predicted mean differences. In their next footnote, McClelland et al. identify that we had a typo in our Fig. 4. That's true, and we are grateful that they found it—in Fig. 4 in Iacobucci et al. (2015), the 8th bar from the left (in the Factor A results) should be at height 0.08 not at 0.11. Again, we thank them for finding that error. McClelland et al. then reiterate again their desire to see non-null effects (here, on factor B and the interaction A × B), once again demonstrating confusion about conditions under which Type I errors may be studied.

In their final point, McClelland et al. illustrate in their Fig. 3 an abecedarian relationship between the sampling distribution of a correlation and the impact that sample size typically has on test statistics—that is, with increased sample sizes, standard errors are smaller (that's why the distributions look tighter). In other words, with increased sample sizes, power is enhanced. This is a perfectly fine point to make, indeed one we have been making. They claim that larger effects are problematic, referring back to their Eq. (4), but recall that we showed how their equation supports our research that when there is no multicollinearity, there is no increase in Type I errors, hence researchers may use a median split.

There were other errors in McClelland et al. (2015), such as several instances of referring to a result as being "more significant" or "less significant." Presumably these comments were meant to say "more likely to be significant" or "larger effects."

We opened our rebuttal to McClelland et al. with comments on regression, and we close with a few more. We note that generally, performing regression analyses on continuous scores rather than blocking subjects into conditions based on a median split can mask important differences that can be better observed by comparing means across conditions. Consider the case of Hirt, Kardes, and Markman (2004). This study investigated effects of reasoning and "debiasing." When people are asked to provide explanations as to why a plausible event should occur, doing so tends to make them form even more extreme judgments of the likelihood of the event's occurrence. Debiasing techniques are cognitive methods for reducing that exaggerated enhancement. Table 1 presents mean judgments (on a scale from 0 to 100) of how likely it was that Portland (the league's then favorite) would win the NBA championship, as a function of different explanation tasks and a median split performed on scores on the need for cognitive closure scale. Need for cognitive closure assesses the degree to which people prefer a definitive answer to a judgment or decision problem, even if this answer requires oversimplification or the sacrifice of accuracy. As the first two rows in Table 1 show, the explanation effect was more pronounced for individuals high (vs. low) in the need for closure. Specifically, when the need for closure was high, likelihood judgments were greater after subjects explained why a particular outcome should occur, namely that the NBA favorite Portland team should win the NBA title, compared to a no explanation control condition (*M*s = 79.55 vs. 64.37). The explanation effect was also significant but less pronounced when the need for closure was low (*M*s = 69.56 vs. 65.95).

Several debiasing conditions were investigated. Explaining how a plausible alternative could win the NBA title significantly reduced likelihood judgments, but explaining how an implausible alternative could win did not. This study also examined the effect of explaining alternatives in unrelated domains (i.e., NFL teams and sitcoms) on likelihood judgments of the favorite NBA team winning the NBA title. When the need for closure was low, debiasing was effective when the debiasing task was easy to perform (i.e., explain two alternatives). However, debiasing backfired when the debiasing task was difficult to perform (i.e., explain eight alternatives). Debiasing backfired regardless of whether the unrelated domain was the NFL (*M*s = 79.30 vs. 69.56) or television sitcoms (*M*s = 76.29 vs. 69.56). When the need for closure was high, little evidence of debiasing was found across conditions.

Table 1
Likelihood judgments as a function of a median split performed on Need for Cognitive Closure (NFCC) scores and explanation task.

Explanation task	Low NFCC	High NFCC
1) Control	65.95	64.37
2) Explain favorite	69.56	79.55
3) Explain favorite and plausible alternative	52.64	66.55
4) Explain favorite and implausible alternative	73.36	82.58
5) Explain favorite and 2 NFL alternatives	59.55	72.11
6) Explain favorite and 8 NFL alternatives	79.30	70.89
7) Explain favorite and 2 sitcom alternatives	53.74	68.35
8) Explain favorite and 8 sitcom alternatives	76.29	71.24

Unfortunately, that manuscript's editor insisted on replacing median split analyses with regression analyses performed on continuous need for closure scores. These analyses showed that the effectiveness of easy-to-perform debiasing tasks increased as the need for closure decreased, and that the effectiveness of difficult-to-perform debiasing tasks decreased as the need for closure decreased. However, regression analyses and plots masked the conditions under which debiasing backfired, an effect that can be seen clearly by comparing means across the debiasing conditions and the control condition. It would be theoretically and practically useful to know when a debiasing technique is likely to be effective, and when a debiasing technique backfires by actually increasing bias. This example from Hirt et al. (2004) shows that important information can be lost by performing regression analyses on continuous scores.

Furthermore, in general, blocking or dividing subjects into subgroups increases power and precision by decreasing "the size of the within-group or error variation" (Rosenthal & Rosnow, 1991, p. 457), is "equally as efficient for curvilinear as for linear relationships between independent and dependent variables" (Rosenthal & Rosnow, 1991, p. 462), and "can be employed even when the blocks differ in qualitative rather than quantitative ways" (Rosenthal & Rosnow, 1991, p. 462). Rosenthal and Rosnow (1991, p. 462) acknowledge that, "blocking always imposes some cost in terms of loss of *df* for error, but that cost is usually small in relation to decreased *MS* error." Finally, some statisticians recommend tertiary or quartiary splits over median splits (Gelman & Park, 2009), which is also fine and would also permit the comparison of mean differences.

Discussion

The core take-away from Iacobucci et al. (2015) is that when independent variables are uncorrelated, conducting a median split will not increase the likelihood of a Type I error being made. This finding is important, because many researchers developed an opposition to median splits based on the single, simple, odd hypothetical data set that Maxwell and Delaney (1993) created. Our study explains why, inconsistent with the claims made in numerous subsequent papers critical of median splits, Maxwell et al.'s data set does not imply that median splits lead to the universal incidence of increased Type I error. Instead, when independent variables are uncorrelated, performing a median split does not pose a threat.

As we make clear in Iacobucci et al. (2015), median splits may reduce power, which was the launching point for Rucker et al. (2015). In our opinion and as witnessed by its use in vast literatures, the benefits that researchers obtain from having the option to report a median split analysis outweigh concerns with reduced power, and accordingly we believe that researchers should feel free to use median splits when, in their judgment, such an analysis would add value. Rucker et al. feel differently, and believe that concerns with Type II error are sufficient to conclude that median splits should never be conducted. We hope that readers will find our exchange useful as they form their own opinions.

McClelland et al. (2015) took a different approach, and offered a series of claims in their doubling down on the position that median splits are always inadvisable. Although individuals sometimes are indeed persuaded by the peripheral cue of the number of arguments in a persuasive message (Petty & Cacioppo, 1984), we addressed each of McClelland's et al.'s objections to facilitate readers as they form their conclusions about this debate.

Despite taking a different approach in their commentaries, both Rucker et al. and McClelland et al. take the extreme position that median splits should not be an option to researchers when data are reported. Given the absolutism of the admonitions contained in the commentaries, one would assume that given the traditional popularity of median splits, and the large number of papers that have used them over the decades, that both commentaries would provide numerous instances in the consumer and psychological literatures of the untoward effects of median splits. Indeed, whether the conclusions of individual papers, or the erroneous evolution of theoretical perspectives, if conducting median splits poses as ample and nefarious a threat to science as claimed by Rucker et al. (2015) and McClelland et al. (2015), it should be straightforward to craft a historical retrospective on the use of median splits that documents myriad false findings and missed opportunities.

Of course, neither set of commentators was able to match their exposition with a delineation of the scientific blunders that should be easy to find if the gravity of their concerns was in step with analytic reality. The conclusions drawn in the commentaries have the feel of "arguments from a vacuum" (Dawes, 1979). If one takes the stance of a blanket prohibition against median splits, supporting cases from the consumer and psychological literatures should be numerous if the commentators' concerns were warranted. As we discussed, the one example from the medical literature that was offered (Mani et al. versus Wicherts) was less than compelling. The absence of problematic cases should loom large for readers weighing whether such a prohibition is justified by the (lack of) evidence. Our simulations and analyses are more consistent with the scientific case history and vast median split usage in the literature in supporting the conclusion that conducting a median split is totally acceptable as long as independent variables are uncorrelated.

In Iacobucci et al. (2015), we summarized reasons, as stated in numerous sources in the literature, why median splits are popular with researchers, and delineated the conditions under which it is appropriate to conduct median splits. The traditional popularity of median splits is consistent with the experience of Hirt et al. (2004) and attests to the idea that median splits can be a very valuable analytic technique. The Hirt et al. example also illustrated how extreme prohibitions against conducting median splits undermined the scientific process.

In the final analysis, working with a median split when predictors are uncorrelated is completely valid. Equally valid is the choice to analyze a continuous variable via regression. We believe scientific progress is surer and swifter when researchers have a full repertoire of appropriate statistical tools available. Making demands that researchers pull median splits from their toolboxes is both unwarranted analytically and evidentially, and poses an undue constraint on researchers' investigation and communication of important findings.

It is important to not ignore the data. Our simulations clearly show that median splits do not increase Type I error when the independent variables are uncorrelated. When the independent variables are correlated, regression is not in and of itself the answer. We agree with Kruglanski et al. (1997) that the appropriate statistical analyses depend on theory and the specific research question that is being addressed, not on dogma.

We recognize that our work stands in opposition to what has become a mainstay of recent methodological teaching, but that teaching is incorrect, and of course that is why this topic makes for an interesting research dialog. At the end of the day we have shown that median splits are *robust* in that performing them does not lead to Type I errors when independent variables are uncorrelated, and we, along with the dutiful efforts of the commentators, have *refined* our collective understanding of the properties of median splits by delineating when they will and will not be appropriate. Accordingly, we have *revived* the median split as a viable analytic option for researchers in the contexts we have specified.

Going forward from this debate, we will live in a more enlightened time. We will know that to state or imply that using a median split is somehow inferior to using a continuous variable is to be median-splitist, which should not be acceptable in polite, scientific society. Researchers may once again use median splits.

Viva the median split!

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jcps.2015.06.014.

References

Aaker, D. A., & Bagozzi, R. P. (1979). Unobservable variables in structural equation models with an application in industrial selling. *Journal of Marketing Research*, *16*(May), 147–158.

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

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory.* Waveland Pr Inc.

Allison, P. D. (1999). *Multiple regression: A primer.* Thousand Oaks, CA: Pine Forge Press.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Pearson.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*(3), 373–400.

Berger, J. O. (1993). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer.

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice.* Newbury Park, CA: Sage.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, *8*(2), 197–207.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Sage.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, *39*(2), 83–87.

Chow, S. L. (1988). Significance test or effect size? *Psychological Bulletin*, *103*(1), 105–110.

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, NJ: Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*, 1–19.

Dawson, J. F., & Richter, A. W. (2006). Probing three-way interactions in moderated multiple regression: Development and application of a slope difference test. *Journal of Applied Psychology*, *91*(4), 917–926.

DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Sage.

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

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, *18*(3), 382–388.

Fraas, J. W., & Newman, I. (2005). The use of the Johnson–Neyman confidence bands and multiple regression models to investigate interaction effects: Important tools for educational researchers and program evaluators. *Multiple Linear Regression Viewpoints*, *24*, 14–24.

Gelman, A., & Park, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *American Statistician*, *62*, 1–8.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, *41*(3), 924–936.

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology*, *40*, 374–383.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask. *Journal of Consumer Psychology*, *19*(4), 673–680.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90–98.

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, Matthew J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, *25*(4), 652–665.

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.

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

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Johnson, P. O., & Fay, L. C. (1950). The Johnson–Neyman technique, its theory and application. *Psychometrika*, *15*(4), 349–367.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

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.

Kerlinger, F. N. (1999). *Foundations of behavioral research* (4th ed.). Cengage.

Kruglanski, A. W., Atash, M. N., DeGrada, E., Mannetti, L., Pierro, A., & Webster, D. M. (1997). Psychological theory testing versus psychometric nay-saying: Comment on Neuberg et al'.s (1997) critique of the need for closure scale. *Journal of Personality and Social Psychology*, *73*(5), 1005–1016.

Kruglanski, A. W., Dechesne, M., Orehek, E., & Pierro, A. (2009). Three decades of lay epistemics: The why, how, and who of knowledge formation. *European Review of Social Psychology*, *20*, 146–191.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*(30, August), 976–980.

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

McClelland, G. H., Lynch, J. G., Jr., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, type II errors, and false positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*, *25*(4), 679–689.

Moses, L. E. (1986). *Think and explain with statistics.* Reading, MA: Addison-Wesley.

Nunnally, J. C. (1978). *Psychometric theory.* McGraw-Hill.

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin*, *92*(3), 766–777.

Pedhazur, E. J. (1991). *Measurement, design, and analysis: An integrated approach.* Psychology Press.

Petty, R. C., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*, 69–81.

Popper, K. R. (2002a). *The logic of scientific discovery* (2nd ed.). New York: Routledge Classics.

Popper, K. R. (2002b). *Conjectures and refutations: The growth of scientific knowledge.* New York: Routledge Classics.

Potthoff, R. F. (1964). On the Johnson–Neyman technique and some extensions thereof. *Psychometrika*, *29*(3), 241–256.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rucker, D. D., McShane, B. B., & Preacher, K. J. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology*, *25*(4), 666–678.

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, *38*(5), 2587–2619.

Shao, J. (2003). *Mathematical statistics.* New York: Springer.

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*(April), 277–288.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315.

Wicherts, J. M., & Scholten, A. Z. (2013). Comment on 'Poverty impedes cognitive function'. *Science*, *342*(6 December), 1169-d.

Zellner, A. (1981). Posterior odds ratios for regression hypotheses: General considerations and some specific results. *Journal of Econometrics*, *16*, 151–167.