

## Peer Review Comments

**Article:** Karch, J. (2020). Improving on Adjusted R-Squared. *Collabra: Psychology*, 6(1): 45. DOI: <https://doi.org/10.1525/collabra.343>

**Article type:** Original research report

**Editor:** Don van Ravenzwaaij

**Article submitted:** 12 March 2020

**Editor decision:** Revisions required

**Revision submitted:** 1 May 2020

**Article accepted:** 1 July 2020

**Article published:** 29 September 2020

---

### Ported Reviews [remove heading if none]

Comments by the editor and the reviewers are black, author's answers in blue, and references to the manuscript text in red.

MET-2019-0215

Improving on Adjusted R-Squared

*Psychological Methods*

Dear Mr. Karch,

I am writing in regard to manuscript MET-2018-0215: "Improving on Adjusted R-Squared" that you submitted to *Psychological Methods*. Thank you for submitting your work. I was fortunate to receive reviews from three experts in regression and R-squared estimation. The reviewers offer a range of criticisms and suggestions for improvement, but ultimately, two recommended rejection and one recommended a major revision. I read the paper carefully myself before reading the reviews. Based on the reviewers' comments, their ratings, and my own evaluation of the manuscript, I am regretfully rejecting the manuscript for publication in *Psychological Methods*.

This paper has much going for it: It tackles an interesting topic, it is written clearly, the simulation study was competently conducted, and R software is provided. However, the reviewers elaborate on several concerns. The major ones follow:

1. Lack of detail in simulation results. Reviewers 1 and 3 emphasize that the simulation results do not provide detail sufficient to give researchers enough information to judge which R-squared is best-suited for their own particular situations.

I had provided what the reviewers asked for (bias + MSE for all conditions) in the supplement of the original submission as csv files, as I thought readers might

benefit most from a digital version of the results. In the revised version, I now include those values (BIAS + MSE for all conditions) in the form of a pdf document as well. I now refer to those two versions of the full results at the beginning of the results section (p. 11):

In the supplementary material, I present bias and MSE for each design cell.

2. Redundancy with past research. While not completely overlapping, as Reviewer 3 notes, much of the simulation was redundant with prior studies by Shieh (2008), Yin and Fan (2001), etc.

I agree that the raw data produced in the simulation study overlaps with the two mentioned papers. However, how I analyze those data completely differently. As I note in the introduction (p. 4):

Second, similarly to Raju et al. (1997), Shieh (2008), I evaluate the estimators using both MSE and bias as both are essential. However, in contrast to Raju et al. (1997), Shieh (2008), I do not combine MSE and bias in an informal way but, instead, as suggested by concepts established within theoretical statistics (Lehmann & Casella, 2003). This strategy avoids non-transparent, conflicting recommendations such as Raju et al. (1997) recommending to use the standard adjusted R<sup>2</sup> estimator and Shieh (2008) the positive-part Pratt estimator despite both basing their conclusion on bias and MSE. It also leads to considering the optimality of estimators through different perspectives, which weight the relative importance of bias and MSE differently.

Note also that I come to completely different conclusions. Claiming too much redundancy with previous research seems akin to claiming a completely novel analysis of existing empirical data, which leads to contrary conclusions, is too redundant for publication. Thus, I do not agree with this.

3. Lack of detail in derivation of the exact R-squared estimator. The paper's major theoretical contribution is arguably the improved algorithm for computing Olkin and Pratt's exact R-squared estimator. Yet, as Reviewer 3 emphasizes, not enough detail was provided to inform readers exactly how it was done in either the main text or the appendix.

I extended the description of the algorithm by providing the R function that implements it in the appendix. The R function should make it absolutely clear how the algorithm works. Furthermore, I now discuss the numerical problems that I encountered and how they were solved (see Appendix A).

4. Size of contribution. The paper has two primary contributions, one theoretical and one practical: An improved algorithm for estimating the exact R-squared, and an R program to facilitate its estimation. The reviewers and I agree that there is definitely a contribution here, but the level is too small to warrant publication in *Psychological Methods*. One reviewer noted privately that, although there is clearly a contribution, it

would be better suited for publication as a technical note for the development of statistical software.

The paper has **three** primary contributions. I clearly mentioned them in the original manuscript (see quote below). I am surprised that the two of the reviewers and the editor believe that the contribution is too little. Essentially, I provide a better alternative to adjusted R-squared, arguably one of the most used statistical quantities in psychology. Admittedly, the improvement is minor but nevertheless noticeable.

In this study, I extend these previous comparisons in three aspects. First, I add a new estimator. This estimator is an implementation of the Olkin-Pratt estimator (Olkin & Pratt, 1958). The Olkin-Pratt estimator has been shown to be optimal under the most prevalent optimality criterion used in statistics, which considered an estimator optimal if it always has least mean squared error (MSE) among all unbiased estimators (Olkin & Pratt, 1958). Despite this favorable property, it is not used and has not been included in any of the previous comparisons due to the difficulties associated with computing it. Here, I show that the Olkin-Pratt estimator can be computed relatively straightforwardly, which enables using it.

Second, similarly to Raju et al. (1997), Shieh (2008), I evaluate the estimators using both MSE and bias as both are essential. However, in contrast to Raju et al. (1997), Shieh (2008), I do not combine MSE and bias in an informal way but, instead, as suggested by concepts established within theoretical statistics (Lehmann & Casella, 2003). This strategy avoids non-transparent, conflicting recommendations such as Raju et al. (1997) recommending to use the standard adjusted R<sup>2</sup> estimator and Shieh (2008) the positive-part Pratt estimator despite both basing their conclusion on bias and MSE. It also leads to considering the optimality of estimators through different perspectives, which weight the relative importance of bias and MSE differently. To obtain bias and MSE of the estimators, I perform a Monte-Carlo simulation study.

Third, to enable usage of the alternative estimators, I provide an R package ("altR2" on CRAN), which contains all estimators compared here. This includes all estimators compared previously.

I agree with all reviewer comments. All three reviewers raise smaller issues that will be important to address at some point.

I realize this news is unwelcome. The topic of explained variance is important and of interest to many researchers. If you decide to update the manuscript for submission to another journal, I strongly encourage you to pay close attention to the reviewers' comments. I wish you luck in submitting your work elsewhere.

For your guidance, I append the reviewers' comments below and hope they will be useful to you as you prepare this work for another outlet.

Thank you for giving us the opportunity to consider your submission.

Sincerely,

XXXXXXXXXX

Associate Editor

*Psychological Methods*

Reviewers' comments:

Reviewer #1: Thank you for the opportunity to review. I found the paper conceptually well-grounded and interesting, and hope the author finds these comments helpful.

My first recommendation would be to present the empirical bias and MSE results for each estimator by rho-squared,  $p$  (# of explanatory variables), and sample size, either in tables or figures. Figure 2 reflects the right idea, but for a limited number of estimators and a single value of  $p$  and  $n$ . More complete results will allow readers to assess whether differences in bias and MSE are large enough to matter under various conditions. The number of values of  $p$  and  $n$  considered could be cut back a bit if need be, but I would retain values of rho-squared across its range. (Aside: it may be the ratio of  $n$  to  $p$  that accounts for much of the difference in performance across conditions.) Simply knowing which estimator performed best by some criterion under a particular condition is less informative.

[See answer to issue 1 of the editor.](#)

A second recommendation is to omit the demonstration that the OP estimator is the unique UMVUE, which has already been established theoretically. We know the OP is unbiased and has lower variance than other unbiased estimators; I think the question is how the other estimators compare, and under what conditions the differences are large enough to matter.

[I partly agree with the reviewer's comment. The methods section should take into account better that the OP estimator has theoretically been shown to be the](#)

UMVUE. To address this, I changed the "Uniformly Minimum Mean Squared Error Unbiased." subsection slightly.

However, I think it is still important to demonstrate that the implemented exact OP estimator is unbiased. For example, reviewer 3 was skeptical whether this is the case, and I also think it is important to demonstrate empirically that the implemented exact OP estimator is unbiased.

To address the second question (how do the other estimators compare and under what condition are the differences large enough), I now discuss the magnitude of the bias values in the results section. For the details, see the difference document.

P3, sentence beginning "While standard ..." should be re-worded to clarify that these software packages do not \*only\* report adjusted R-squared (they also report the unadjusted R-squared).

I agree. I changed the relevant sentence (p. 3) to:

While standard linear regression software, such as SPSS and the "lm" function in R (R Core Team, 2018), and consequently the vast majority of psychological publications report only R<sup>2</sup> and adjusted R<sup>2</sup>

P9, I would recommend clarifying that a uniformly minimum variance unbiased estimator only needs to have variance less than or equal to that of all unbiased estimators (i.e., there can be more than one UMVUE), which isn't clear from the statement that UMVUE must have "smallest" variance.

To address this comment, I now use the word minimum instead of least in the relevant sentence.

P9, test for empirical bias. Note that failure to reject the null is not evidence for the null, as p-values can only quantify evidence \*against\* the null. Also, failure to reject may be due to inadequate power to detect the difference. Rather than assess bias in binary terms (yes/no), it seems better simply to \*quantify\* bias (e.g., by its mean or median across simulated data sets).

In principle, I agree that it would be better to quantify bias. To this end, I provided the bias value of all conditions and estimators in the supplement, which I now extended and referred to more prominently (see answer to issue 1 of the editor). However, summarizing all 276 bias values in a meaningful way seems impossible. I do not think that using any kind of average, such as the mean or median, across all conditions is a good approach, as the conditions vary greatly, and thus the mean or

median will not be representative of any particular condition a researcher encounters. Instead, I prefer to summarize the behavior of the estimator across conditions by indicating which estimator performed more or less unbiased in which condition. Instead of summarizing the bias values across conditions, I now discuss the bias values for two selected conditions in the text and added two new figures displaying them.

While I agree with the reviewer's comment regarding the logic of p-value, note that the sample size was 100,000. Thus, the power to detect even small differences is high. Nevertheless, to make the reader aware of this problem, I added the following (p. 9-10)

Note that an estimator might still be unbiased despite being empirically unbiased. Indeed, all estimators besides the Olin-Pratt estimator are biased. However, if they are empirically unbiased, the bias is likely negligible.

P11, I'm not sure averaging MSE over all the rho-squared values considered is analogous to placing a uniform prior on rho-squared, as the rho-squared values here aren't evenly distributed across the  $[0, 1]$  range. This also gives more weight to smaller rho-squared values when averaging MSEs, which will benefit R-squared estimators that perform better in the  $[0, 0.1]$  range for rho-squared.

I thank the reviewer for spotting this inconsistency. I now average on evenly distributed values across the  $[0, 1]$  range. This changed some results. However, it did not change any conclusions.

P20, paragraph beginning "In case ...," does the author mean the \*unbiased\* uniformly minimum MSE is the standard in psychology? Note that this is arguable, as ML methods are also prominent, and ML estimators are not always unbiased (as they are with the slope and intercept estimates in Gaussian OLS).

I thank the reviewer for spotting this crucial left-out word.

With regard to the argument that maximum likelihood methods are also prominent, which seems to be in conflict with claiming that uniformly minimum MSE being the standard in psychology, I want to note that in my experience, maximum likelihood is mostly used because it is easier. If a maximum likelihood estimator and a uniformly minimum MSE exist, the uniformly minimum MSE is generally preferred. However, I see the reviewers point that some readers might disagree with this and also such a strong statement is not needed for the discussion. I thus changed the relevant sentence to (p. 21):

The rationale for this is that the uniformly lowest MSE unbiased perspective is the standard in regression analysis.

P25, it may be better to avoid references to complex numbers here, as technically these are numbers with an imaginary component, which we don't encounter in data analysis. E.g., we can simply say the Gamma function generalizes the factorial function to the positive real line.

I simplified and extended the description of the hypergeometric function in the appendix. It now does not refer to complex numbers anymore.

Thank you for allowing me to review your paper. I enjoyed it.

Reviewer #2: This article addresses the estimation of coefficient of determination or squared multiple correlation coefficient for the multiple linear regression under the assumption that the criterion and predictor variables have a joint multi-normal distribution. Although the unique minimum variance unbiased estimator (MVUE) has been proposed in Olkin and Pratt (1958), the computational complexity has limited its practical use. Accordingly, numerous alternative and simplified estimators have been described in the literature. The major goal of this article is to propose an efficient algorithm for calculating the optimal MVUE by using some simplified properties of the hypergeometric function. Extensive numerical investigations were conducted to examine and compare MVUE with other notable estimators in terms of bias and mean square error over a wide range of model configurations. Due to the selected model settings and performance criteria, the findings do not completely agree with the existing results. Overall, the results are useful in some sense but the exact computation of MUVE is not really a significant contribution.

See answer to issue 4 of the editor.

Additional comments:

1. The notation for MLE and unbiased estimator of  $\sigma^2 Y$  and  $\sigma^2 \varepsilon$  (p. 5) should be different.

I now use different symbols for the unbiased and maximum likelihood estimators.

2. The probability function of  $R^2$  is quite complicated. The generation of described random variables  $R^2$  in the simulation requires more detailed explanation.

I do not understand this request.

3. It is described (p. 9) that the Welch t-test to test whether an estimator is empirically unbiased. This procedure is commonly applied to test mean difference of two treatment groups under variance heterogeneity. It is not clear why it is necessary and applicable here. Also, the number of replication 100, 000 (p. 9) is substantially larger than the common sample sizes for hypothesis testing and tend to have high power to detect any minor effect.

This was a typo. Instead of the Welch t-test the one-sample t-test was used.

With regard to the sample size, this was intentionally high to also be able to detect small biases. To investigate whether the effects are minor are not, I added a discussion of the obtained bias values for some selected conditions.

4. Because the simulation study has a large number of replications, it is argued that the estimation findings should be nearly identical to those obtained by the exact numerical integration. But, this is not consistent with the emphasis of exact calculation for the MVUE.

In general, I agree with this comment and would have also preferred to perform exact numerical integration. However, I did not do that because as I mention on p. 22

I did not use more precise alternatives such as analytical derivations or numerical methods (Shieh, 2008) because both approaches would not allow a direct assessment of the provided R package.

Reviewer #3: The purposes of the article were to propose a method of computing the Olkin & Pratt (1958) bias-adjusted  $R^2$  estimator, and compare the sampling properties of the estimator to other existing measures. It is well-known that the sample analog estimator of the population  $R^2$  is positively biased. Many  $R^2$  estimators have been proposed to adjust for this bias, with the Ezekiel adjusted  $R^2$  often the default in statistical software. The proliferation of  $R^2$  adjustments results from different approaches for approximating the bias of the sample  $R^2$ . Systematic reviews have offered conflicting recommendations to researchers regarding which estimator should be used. An unbiased  $R^2$  estimator derived by Olkin & Pratt (1958) has strong theoretical justification, but its computational difficulty has impeded its widespread use in applied research. The author shows that the computationally difficult component of the Olkin & Pratt estimator (i.e., hypergeometric function) can be expressed in a more tractable closed-form solution. The bias and MSE of the exact Olkin & Pratt estimator were then examined and compared to other  $R^2$  estimators. The exact Olkin & Pratt estimator was deemed



to be empirically unbiased, but had higher MSE than several other estimators. An R package was created to compute the exact  $R^2$  estimator, and recommendations for using  $R^2$  estimators under various conditions are provided.

The article was clear and well-written. The motivating issue of conflicting recommendations for adjusted  $R^2$  is certainly worth resolving, and the addition of a practical form of an unbiased  $R^2$  estimator would also be beneficial to researchers. The proposed comparisons of  $R^2$  estimators using multiple criteria for bias and accuracy provides a more complete picture of the estimators' properties, and the results would likely be useful in resolving discrepancies from prior literature. In addition, the MC simulation conditions were thoughtfully chosen to be representative of applied psychological studies.

#### Primary concerns

\* The simulation study results present an incomplete picture of the properties of the estimators. First, magnitudes of bias were not provided, and statistical evaluations of bias were only provided for a subset of the estimators. One can't evaluate the bias/variance tradeoff of using certain shrinkage estimators without this information. For example, simulation results show the positive part Claudy estimator has the lowest MSE for several conditions, and would be recommended for those study designs assuming the researcher knows the population  $R^2$ . However, Yin and Fan (2001) and Shieh (2008) show this is also the most biased estimator. It is reasonable to assume a researcher may then want to select a less biased estimator with a marginally higher MSE. Second, only summaries of MSE are provided, not the specific MSE values per condition. It would be important to know if there are any large discrepancies in MSE among the estimators. If there are only small discrepancies in MSE, bias estimates would then likely be more useful for choosing an estimator.

With regard to the bias / variance tradeoff. Note that MSE already captures the bias-variance tradeoff as it can be expressed as  $\text{bias}^2 + \text{variance}$ . In the example, the reviewer talks about a bias / **MSE** tradeoff. In particular, he claims that researchers want to select a less biased estimator with a marginally higher MSE. I would, in general, argue against doing that. As I have said, the MSE already quantifies the bias / variance tradeoff. Thus, by selecting a marginally higher MSE estimator in favor of less bias, the researcher selects an estimator with a worse bias / variance tradeoff. Instead of doing that, I recommend, as I argued in the paper, to first decide whether unbiasedness is important or not and based on that use either the unbiased minimum MSE exact Olkin-Pratt estimator or carefully select the estimator with the smallest MSE, which again already quantifies the bias variance tradeoff. Of course, a researcher might want to use a different weighting of bias and variance than the one used by MSE. However, I am not aware of any other generally accepted approach.

Regarding the second point, I now provide mean and bias for all conditions as supplementary material. See also my answer to issue 1 by the editor.

In addition, although some additional measures were used to summarize results, much of the simulation was redundant with prior studies of adjusted  $R^2$  from Shieh (2008), Yin and Fan (2001), etc.

See answer to issue 2 by the editor

\* I do not follow the logic of the statistical evaluations for bias (e.g., Table 1). The exact Olkin & Pratt estimator is by definition unbiased because it is a function of the complete sufficient statistic. The other Olkin & Pratt estimators are approximations so must be biased regardless of sample size. The bias may be vanishingly small and difficult to distinguish statistically from zero with increasing sample size, but they shouldn't be considered in the same class of unbiased estimators as the exact estimator.

I changed the statistical evaluation for bias such that they now are in line with this comment. See my answer to the second recommendation of reviewer 1 for the details.

\* The pattern of results from the simulation study in Table B1 do not appear readily generalizable to study designs beyond those considered in the article. Is there a sample size at which the pattern of MSE results would stabilize? Would results converge on a single best estimator for large samples?

With regard to the convergence to a single best estimator, note that all estimators considered here are consistent. Thus, for large sample sizes, they all converge to the same true value. I thought this would be obvious but the reviewer's comment made me aware that this is not the case. I added a new subsection to make the readers aware of this, which is called "A note on consistency".

\* Lack of detail provided for the derivation of the exact  $R^2$  estimator. The key identity used to re-express the infinite series as a closed-form solution is given a citation, but no proof is provided in either the appendix or citation. Although it appears from the simulation results that bias is reduced compared to the other  $R^2$  estimators, it is not clear whether the proposed estimator is in fact unbiased or a better approximation. In addition, details of the closed form solution may be generalizable to other difficult estimators that include the hypergeometric function (e.g., Bessel function)

Note that Olkin-Pratt (1959) showed that their Olkin-Pratt estimator is unbiased. Assuming that the sources I cite are not erroneous, I derive an approach to calculate the Olkin-Pratt estimator exactly. I also demonstrate that no bias can be detected with 100,000 samples in any condition. Thus, I do not agree with the statement "it is not clear whether the proposed estimator is in fact unbiased or a better approximation." It is extremely probable that the proposed estimator is unbiased.

I understand the motivation of the reviewer to request proof is 1) to increase the credibility of the closed-form solutions claimed in the cited literature, and 2) to see whether some steps taken can be used to calculate other difficult estimators such as the Bessel function. In principle, I agree that being able to point to proofs would be beneficial, especially to see whether some steps taken can be used to calculate other difficult estimators. However, I am not aware of sources that prove the specific statements. Additionally, I doubt that many readers of Collabra would be able to follow such a proof. Thus, I would prefer not to include proofs in the paper. With regard to credibility of the cited literature, I want to note that the sources I cite are a book from the National Institute of Standards and Technology, and the software Mathematica. The probability that such authoritative documents contain erroneous statements seems negligible, especially since my simulation results are in line with their claims.

\* An empirical example would be useful for the reader, otherwise it is hard to appreciate the differences in magnitudes among the estimators.

I agree that an empirical example might be instructive. Despite this, I would prefer not to include an empirical example as it might create the impression that it is meaningful to compare the estimators based on one sample only, which is not the case. The estimators can only be compared with regard to their performance when applying them repeatedly, which I investigate in the simulation study.

\* Although the complete results for bias are not presented, the biases for many of the  $R^2$  approximations appear to be quite small and likely negligible for common study designs. In other words, the value these estimators would converge to across repeated studies would be very close to the truth, while also minimizing risk according to the MSE results. In addition, the  $R^2$  approximation formulae have the significant advantage of being calculable by hand, whereas the size of the exact formula appears to increase with sample size.

Note that since the estimators are consistent, the differences between the estimators not only in BIAS but also MSE become negligible. I communicate this now better to the reader, most prominently with the added subsection on consistency.

Thus, the hypothesis that the bias differences become negligible at some point, and then one can reside to only using MSE is wrong. Instead, both differences become negligible at some point.

With regard to the  $R^2$  approximations having the advantage of being calculable by hand: This does not seem to be a relevant advantage. Calculating the regression coefficients is also not feasible by hand. Thus, a computer has to be used anyway. If the reviewer is concerned about the time needed, calculating the exact estimator takes milliseconds on standard computers and is orders of magnitude faster than calculating the OLS estimates for the regression coefficients.

---

## Editor Decision for Version 1

**Editor:** Don van Ravenzwaaij

**Affiliation:** University of Groningen

**Editor decision:** Revisions required

**Decision date:** 23 April 2020

Dear Dr. Karch,

I have carefully read your manuscript, "Improving on Adjusted R-Squared" that you submitted for streamlined review. I have also read the original decision letter with your responses. Although I find myself in agreement with some of the issues that were raised by the original reviewers, I believe that your manuscript has important strengths and could, after suitable revisions, be suitable for publication. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

As this is a streamlined review, I will bring up the points that I believe are critical to address. Some of these have been brought up by the original reviewers, some are new and result from the fact that Collabra: Psychology has a different readership from Psychological Methods.

First and foremost, it is important to provide a better justification of your recommendation to go with the OP estimator. Specifically, why should we not have our decision be guided by MSE, which takes into account both bias and variance? Relatedly, I believe that this part "I do not combine MSE and bias in an informal way but, instead, as suggested by concepts established within theoretical statistics (Lehmann & Casella, 2003). This strategy avoids non-transparent, conflicting recommendations such as Raju et al. (1997) recommending to use the standard adjusted R<sup>2</sup> estimator and Shieh (2008) the positive-part Pratt estimator despite both basing their conclusion on bias and MSE." requires elaboration. You claim this strategy avoids non-transparent recommendations, but the strategy itself could be made more transparent.

Secondly, I agree with Reviewer 2 that the simulation specifics require elaboration (you state you do not understand this request). Please explain in the manuscript exactly what you have done. For instance, at what level(s) have you introduced sample variance (at the level of the squared multiple correlation, at the level of the individual predictors)? How large was the sample variance at each level? Please provide distributions you sampled from for each level where sample variance was introduced. An example of how one might go about this is in section 'Power simulations' in van Ravenzwaaij, Donkin & Vandekerckhove, 2017 (this is not meant as a covert suggestion to refer to this, just an example of how to go about this). The reference to Fisher, 1928 is too generic for the reader to ascertain whether there are choices you have made in the simulation set-up that could affect the

outcome. In addition to the in-paper description, please make available the R-code with your simulations.

Thirdly, the abstract gives the impression that the OP estimator is clearly the superior choice, whereas the manuscript gives a much more nuanced impression. I fear the typical Collabra: Psychology reader will be wondering how much this matters (and, as such, stick with what they know). It might be worth cherry-picking an example where E+ is hugely off and OPE+ is on the money, just so the reader gets a visualization of how wrong it can go.

Fourthly, the manuscript is quite technical. To an extent this cannot be avoided given the topic, but please go over the paper one more time and see if there are parts that are better relegated to an appendix.

Finally, small point, the second decimal in the tables in the supplementary material for rho squared are missing.

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at [editorialoffice@collabra.org](mailto:editorialoffice@collabra.org).

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Don van Ravenzwaaij Senior editor, Collabra: Psychology

---

### **Author's Response to Review Comments for Version 1**

**Author:** Julian Karch

**Affiliation:** Leiden University

**Revision submitted:** 1 May 2020

Dear Don,

Thanks a lot for your swift and detailed feedback. I agree with most points you raised and thus addressed them as thoroughly as possible.

Comments by you are black, my answers in blue, and references to the manuscript text in red.

I have carefully read your manuscript, "Improving on Adjusted R-Squared" that you submitted for streamlined review. I have also read the original decision letter with your responses. Although I find myself in agreement with some of the issues that were raised by the original reviewers, I believe that your manuscript has important strengths and could, after suitable revisions, be suitable for publication. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

I would like to thank you for this generally positive assessment of my paper.

As this is a streamlined review, I will bring up the points that I believe are critical to address. Some of these have been brought up by the original reviewers, some are new and result from the fact that Collabra: Psychology has a different readership from Psychological Methods.

First and foremost, it is important to provide a better justification of your recommendation to go with the OP estimator. Specifically, why should we not have our decision be guided by MSE, which takes into account both bias and variance?

This and your third comment, which was about the discrepancy between the abstract and the main text, made me realize that I need to be more careful with how I phrase my recommendations. My main intended recommendation is indeed much more nuanced than just to go with the OP estimator, as I discuss in the discussion. Essentially, one should decide for each situation what is more important unbiasedness or lower MSE. If unbiasedness is more important, the estimator to pick is indeed the OP estimator; if not, the choice is much more complex. However, I am aware of the fact that for most psychologists choosing between unbiasedness and lower MSE is not feasible. For them, I also wanted to provide guidance and argued that the default position should be to value unbiasedness over lower MSE.

In response to your comment, I thus modified the second half of the abstract to be more in line with the more nuanced recommendations in the main text. It now reads:

The metrics I use for comparison closely resemble established theoretical optimality properties. Importantly, the exact Olkin-Pratt estimator is shown to be optimal under the standard metric, which considers an estimator optimal if it has the least mean squared error among all unbiased estimators. Under the important alternative metric, which aims for the estimator with the lowest mean squared error, no optimal estimator could be identified. Based on these results, I carefully provide recommendations on when to use which estimator, which first and foremost depends on the choice of which metric is deemed most appropriate. If such a choice is infeasible, I recommend using the exact Olkin-Pratt instead of the default adjusted

R-squared estimator. To facilitate this, I provide the R package `altR2`, which implements the Olkin-Pratt estimator as well as all other estimators.

Additionally, I extended my argument about why I think unbiasedness should be the default position in the discussion (p. 21):

In case it is not possible to determine what is more important unbiasedness or minimization of MSE, I recommend using the unbiased exact Olkin-Pratt estimator. There

are three reasons for this. The first one is consistency. The uniformly lowest MSE unbiased

perspective is the standard in regression analysis. Especially, the rationale given for using

the ordinary least squared regression coefficients is that they are uniformly lowest MSE

unbiased (Cohen, Cohen, West, & Aiken, 2003, p. 124). Second, psychology is traditionally

primarily concerned with explanatory modeling (Yarkoni & Westfall, 2017), and in

explanatory modeling, unbiasedness should be preferred over lower MSE (Shmueli, 2010).

Third and in relation thereto, the fact that an estimator is unbiased guarantees that when

estimates for the same property of interest from multiple studies are aggregated in a

meta-analysis, this aggregation eventually leads to the true value, which is not the case

when using a biased estimator, even if it has lower MSE.

I also rephrased the beginning of the Conclusion section to better reflect the nuances (p. 23):

In conclusion, I recommend using the exact Olkin-Pratt estimator by default.

However, if the researcher is confident that minimizing MSE is more critical than unbiasedness, then a different estimator should be used.

Relatedly, I believe that this part "I do not combine MSE and bias in an informal way but, instead, as suggested by concepts established within theoretical statistics (Lehmann & Casella, 2003). This strategy avoids non-transparent, conflicting recommendations such as Raju et al. (1997) recommending to use the standard adjusted R<sup>2</sup> estimator and Shieh (2008) the positive-part Pratt estimator

despite both basing their conclusion on bias and MSE." requires elaboration. You claim this strategy avoids non-transparent recommendations, but the strategy itself could be made more transparent.

Since I explain my strategy and its links to theoretical statistics in detail in the "Evaluating the estimators" section, it seems the problem is that I do not explain the strategy immediately after the sentences you refer to. I considered explaining the gist of my strategy in the introduction after the sentences you mentioned. However, I realized that this will make the introduction rather technical and I thus prefer to explain the strategy in detail in the methods section. To make the reader aware that the strategy will be explained in detail in the methods section, I added the following sentence to the intro (p. 4):

**I will provide the detailed explanation of this strategy in the Method section.**

Secondly, I agree with Reviewer 2 that the simulation specifics require elaboration (you state you do not understand this request). Please explain in the manuscript exactly what you have done. For instance, at what level(s) have you introduced sample variance (at the level of the squared multiple correlation, at the level of the individual predictors)? How large was the sample variance at each level? Please provide distributions you sampled from for each level where sample variance was introduced. An example of how one might go about this is in section 'Power simulations' in van Ravenzwaaij, Donkin & Vandekerckhove, 2017 (this is not meant as a covert suggestion to refer to this, just an example of how to go about this). The reference to Fisher, 1928 is too generic for the reader to ascertain whether there are choices you have made in the simulation set-up that could affect the outcome.

I now understand what Reviewer 2 and you requested - thanks for clarifying - and added a new section, which explains in detail how I generated the data.

In particular, I added the following to the main text (p. 9):

### **Data Generation**

In Appendix B as well as in the code (<https://doi.org/10.24433/CO.8023088.v3.>), I describe how I simulated data for a given combination of sample size N, number of predictors p, and squared multiple correlation in detail. For generating the data, some



parameters, like the residual variance, had to be set to fixed values. Importantly, the

results are not influenced by those choices, as the sampling distribution of  $\hat{\beta}$  and thus the

bias and MSE of each estimator are independent of those parameter values (Fisher, 1928).

and added Appendix B (see next page)

I choose to relegate the details to the appendix because the simulation results are invariant to them and to not make the manuscript more technical.

In addition to the in-paper description, please make available the R-code with your simulations.

I had already made the code available for the original submission and referred to it in the text. I now added a reference in the newly added data generation section as well (p. 9):

“In Appendix B as well as in the code (<https://doi.org/10.24433/CO.8023088.v3>), I explain how I simulated data for a given combination of sample size  $N$ , number of predictors  $p$ , and squared multiple correlation  $R^2$  in detail.”

Thirdly, the abstract gives the impression that the OP estimator is clearly the superior choice, whereas the manuscript gives a much more nuanced impression.

As already mentioned, I modified the abstract to be more in line with the main text in this regard.

I fear the typical Collabra: Psychology reader will be wondering how much this matters (and, as such, stick with what they know). It might be worth cherry-picking an example where  $E+$  is hugely off and  $OPE+$  is on the money, just so the reader gets a visualization of how wrong it can go.

I had already considered including a cherry-picked example where the OPE estimator clearly outperforms due to the same considerations. I decided against it and would prefer to stick with this decision due to the following reasons. First and foremost, when evaluating estimator using bias or MSE, it is meaningless to compare the performance of two estimators on one particular data set. So, I do not want to create the impression that it was meaningful by performing such a comparison. The closest thing to what you suggest, which does not suffer from this limitation, is to describe how high the bias of the Ezekiel estimator can get in the worst case, which I already included in the results section. If you still feel that adding such an example would be beneficial, I am happy to do so.

Fourthly, the manuscript is quite technical. To an extent this cannot be avoided given the topic, but please go over the paper one more time and see if there are parts that are better relegated to an appendix.

I already had relegated the most technical content to the appendix because I had exactly the same concern. Since now I explain the data generation in detail in the Appendix, I was able to additionally remove a quite technical paragraph from the main text (used to be on page 5; starting with "It can be re-expressed as follows").

Finally, small point, the second decimal in the tables in the supplementary material for rho squared are missing.

I thank you for spotting this mistake, which I fixed.

---

## Editor Decision for Version 2

**Editor:** Don van Ravenzwaaij

**Affiliation:** University of Groningen

**Editor decision:** Accept submission

**Decision date:** 22 May 2020

Dear Julian D Karch,

I have now had a chance to read over your manuscript "Improving on Adjusted R-Squared", along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-production file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely, Don van Ravenzwaaij

Senior Editor, Collabra: Psychology