(Cheung, S. F., Cheung, S.-H., Lau, E. Y. Y., Hui, C. H., & Vong, W. N. (2022). Improving an old way to measure moderation effect in standardized units. *Health Psychology, 41*(7), 502–505. https://doi.org/10.1037/hea0001188. This is the version accepted, before final typesetting.)

Improving an Old Way to Measure Moderation Effect in Standardized Units

Shu Fai Cheung[1]

Sing-Hang Cheung[2]

Esther Yuet Ying Lau[3,4,5,6]

C. Harry Hui[2]

Weng Ngai Vong[1]

[1] Department of Psychology, University of Macau, Macao SAR, China

[2] Department of Psychology, The University of Hong Kong, Hong Kong SAR, China

[3] Department of Psychology, The Education University, Hong Kong SAR, China

[4] Sleep Laboratory, Department of Psychology, The Education University of Hong Kong, Hong Kong SAR, China

[5] Centre for Psychosocial Health, The Education University of Hong Kong, Hong Kong, Hong Kong SAR, China

[6] Centre for Religious and Spirituality Education, The Education University of Hong Kong, Hong Kong SAR, China

Author Note

This study was supported by the Multi-Year Research Grant (MYRG2019-00072-FSS) from University of Macau, Macao SAR, China. The dataset used was supported by the General Research Fund (EdUHK#18611419), Research Grants Council, Hong Kong SAR, China. We would like to thank Mike W.-L. Cheung for his valuable comments on an earlier draft of this manuscript. Correspondence concerning this article should be addressed to Shu Fai Cheung, Department of Psychology, Faculty of Social Sciences, University of Macau, Avenida da Universidade, Taipa, Macao SAR, China or by email (sfcheung@um.edu.mo).

Abstract

Moderation effects in multiple regression, tested usually by the inclusion of a product term, are frequently investigated in health psychology. However, several issues in presenting the moderation effects in standardized units and their associated confidence intervals are commonly observed. While an old method had been proposed to standardize variables in moderated regression before fitting a moderated regression model, this method was rarely used due to inconvenience and even when used, the confidence intervals derived were biased. Here, we attempt to solve these two problems by providing a tool to conveniently conduct standardization in moderated regression without the step of standardizing the variables beforehand and to accurately form the nonparametric bootstrapping confidence intervals for this standardized measure of moderation effects. Health psychology researchers are now equipped with a tool that can be used to report and interpret standardized moderation effects correctly.

Keywords: moderation, effect size, standardized solution, confidence interval

Improving an Old Way to Measure Moderation Effect in Standardized Units

Despite its long history and popularity since its inception as early as late 1960s (Hobert & Dunnette, 1967), the moderation effect is still usually presented in a metric that can be misinterpreted, such as the standardized coefficient of the product term. We introduce below an old but rarely used way to estimate the *standardized moderation effect*, the moderation effect in standardized units. We then discuss the problem with forming the confidence interval of this standardized effect and introduce a tool to assist forming the confidence interval of this effect size measure using nonparametric bootstrapping. We also illustrate analytically how it can complement existing standardized measures moderation effect. Lastly, we make recommendations for reporting standardized effect size measures of moderation effects correctly.

A Brief Overview of Moderated Regression

We denote the independent variable by X, the outcome variable by Y, and the moderator variable by W. The moderated regression model is:

$$Y_{Predicted} = B_0 + B_X X + B_W W + B_{WX} WX$$

The coefficient of B_{WX} represents how the effect of X changes with W. When fitting this model to a sample of data, B_{WX} measures the moderation effect on the scales of X, W, and Y. For each one unit increase in W, the regression coefficient of X changes by B_{WX}. In practice, it is common to mean center either only the moderator or both the moderator and the independent variable (Aiken & West, 1991; Edwards, 2009; see Appendix A, available at the OSF page of this paper, for the details[1]). Mean-centering has the advantage that all three coefficients are directly interpretable without the need for further calculation. Note that B_{WX} is not affected by centering either or both the independent variable and the moderator. This

[1] All appendices and other supplementary materials are available from the OSF page for this manuscript: https://osf.io/ac8de/)

illustrates the well-known phenomenon that any kind of centering, mean-centering included, does not affect the estimation and test of moderation effect (Aiken & West, 1991; Brambor et al., 2016; Edwards, 2009). As argued by Hayes (2018) and others, mean-centering has nothing to do with multicollinearity, as the standard error of the conditional effect is not affected.

Standardization in Moderated Regression: An Old but Right Way

To measure the moderation effect on standardized metric, a simple way has been proposed nearly four decades ago by Friedrich (1982), who demonstrated that the proper way to present the moderation effect in standardized metric is to standardize the independent variable and moderator *before* computing the product term, and then interpret the *unstandardized regression coefficient*. Suppose we standardize W, then $Z_W = (W - C_W)/S_W$ is used instead of W in the regression model. If we standardize all variables, then, as shown in Appendix A:

$$Z_{Y_{Predicted}} = \frac{(B_0 + B_W C_W + B_X C_X)}{S_Y} + \left[\frac{(B_X + B_{WX} C_W) S_X}{S_Y}\right] Z_X + [\frac{(B_W + B_{WX} C_X) S_W}{S_Y}] Z_W + [\frac{S_W S_X}{S_Y}] B_{WX} Z_W Z_X$$

Following Friedrich's procedure, the unstandardized regression coefficient of the product term in this model, $\frac{S_W S_X}{S_Y} B_{WX}$, represents the change of the *standardized* effect of the independent variable on the outcome for each one *standard deviation* increase of the moderator. This coefficient tells us the nature of the moderation effect in the same way the unstandardized coefficient does, but in the standardized metric. Naturally, we call this term the *standardized moderation effect*[2]. The equation also shows that, like the standardized

[2] We do not claim credit for proposing this term, as it sounds natural to name the effect this way. This term has been used in a few published studies (e.g., Markwart, Bomba, Menrath, Brenk-Franz, Ernst, Thyen, et al., 2020). However, this term was not clearly defined or was used to denote the standardized coefficient of the product term.

regression coefficients for a model without moderators, the standardized moderation effect can be computed from the unstandardized regression coefficient and the standard deviations of *X*, *W*, and *Y*, which are readily available in the regression output of all common statistical software packages.

Despite its long history and recent discussion (e.g., Hayes, 2018), we found this method rarely used in published studies in health psychology when standardized effects are presented or interpreted. In a review of recent papers on health psychology (details at the OSF page), only a minority of the studies adopted this approach, even though many studies reported some standardized measures of effect sizes. Therefore, it is important to revive this rarely used but correct method for presenting a moderation effect on the metric of standardized score, when this metric is appropriate in a study.

Forming the Confidence Interval for the Standardized Moderation Effect

Although Friedrich's approach is easy to implement, it has one problem. The standard error, and hence the confidence interval, of the standardized moderation effect can be biased. For example, in a regression model without moderators, Yuan and Chan (2011; see also Jones & Waller, 2013) showed that the standard errors and the confidence intervals of the standardized regression coefficients are biased because they ignore the sampling variability of the standard deviations.[3] Therefore, even though some researchers may have done standardization correctly, the standard errors and the confidence intervals reported may still be wrong. To our knowledge, no analytical method has been proposed for forming a confidence interval of the standardized moderation effect that addresses this problem, and we hereby propose using the nonparametric bootstrapping technique (Efron & Tibshirani, 1993).

Despite the simplicity of bootstrapping, the case is more complicated for the

[3] It is not easy to derive analytically the confidence interval of the standardized coefficient of a predictor because the sampling distribution is complicated even under multivariate normality (see Yuan & Chan, 2011). It is even more complicated in moderated regression because the product term, which is a product of two variables, is not normally distributed (Craig, 1936).

standardized moderation effect. This is because bootstrapping resampling is required *before* standardizing the independent variable and the moderator to form the product term. Doing standardization once and then using bootstrapping to form the confidence interval on the parameters is incorrect, because it does not take into account the sampling variance in the standard deviations of the independent variable, moderator, and outcome variable (Cheung, 2009). Therefore, using existing tools that allow obtaining the bootstrapping confidence intervals for regression coefficients can yield incorrect confidence intervals for the standardized moderation effect estimate. To implement nonparametric bootstrapping in standardized moderation effect correctly and easily, we developed an R package, `stdmod`, that does not require writing additional custom functions. Researchers who use R (R Core Team, 2020) can use the package after they do moderated regression as usual, without the need to do the standardization themselves. The main function is `std_selected_boot`. Users can do regression by `lm` as usual, pass the output to `std_selected_boot`, and specify which variables need to be standardized. It will automatically draw the requested number of bootstrap samples and do all the steps (standardization, forming the product term, estimating the regression coefficients) in each bootstrap sample, and then return the confidence intervals. Details of the package, as well as annotated examples, can be downloaded from the OSF page of this manuscript (in the folders *stdmod_package*). The examples can be viewed directly on the site without the need to download them.

Proposed Convention to Interpret the Level of Standardized Moderation Effect

We propose to interpret the standardized moderation effect on the basis (a) of the common rule of thumb for bivariate correlation, (b) that a standardized mean difference of 0.5, corresponding to a difference of 0.5 standard deviation (SD) between two population means, is usually considered a medium effect, and (c) that the standardized coefficients of .10, .30, and .50 are small, medium, and large standardized independent variable moderation effects respectively. We interpret changes of. 05, .10, and .15 in the standardized independent variable effect as small, medium, and large in size, respectively. Thus, a standardized moderation effect of .20 is medium in effect size, given that a medium increase (0.5 SD) of the moderator would lead to a medium change in the standardized independent variable effect. Correspondingly, we regard standardized moderation effects of .10 and .30 as small and large, respectively. We ought to emphasize that these are just rules of thumb, based on existing conventions for other effect size measures. Other criteria that researchers consider appropriate and justifiable should also be considered for adoption.

Comparing Standardized Moderation Effect to R^2 Increase, Standardized Coefficient of the Product Term, and Two Variance Proportion Measures

We are not aware of any discussion on how standardized moderation effect is related to the other measures of moderation effect. We will compare the standardized moderation effect with two commonly reported measures and two recently proposed measures (Liu & Yuan, 2020) analytically in the following sections.

The R^2 Increase When Adding the Product Term

The R^2 increase when adding the product term to a model without the product term is a useful measure because it frames the assessment of the moderation effect from a model comparison perspective. However, the R^2 increase is influenced by both the standardized moderation effect and correlation between the independent variable and the moderator. As shown in Appendix B (on the OSF page), two moderated regression models with the same standardized moderation effect can have different R^2 increases, and two moderated regression models with the same R^2 increase can also have different standardized moderated effects. Considering the simple case of having only one independent variable and one moderator, assuming that their joint distribution is multivariate normal, and holding the R^2 without the moderation effect constant, the population R^2 increase is $\dot{\beta}^2_{Z_W Z_X}(1 + \rho^2_{WX})$, where $\dot{\beta}_{Z_W Z_X}$ is the population standardized moderation effect, and ρ_{WX} is the population correlation between the independent variable and the moderator. Therefore, R^2 increase cannot replace standardized moderation effect because the former also depends on the correlation between the independent variable and the moderator.

Standardized Regression Coefficient of the Product Term

As demonstrated in Appendix C (on the OSF page), if we assume that the population means of the independent variable and the moderator are zero and their joint distribution is multivariate normal, then the population standardized coefficient of the product, denoted as $\dot{\beta}_{WX}$, is equal to $\sqrt{1+\rho_{WX}^2}\dot{\beta}_{Z_X Z_X}$, where ρ_{WX} and $\dot{\beta}_{Z_X Z_X}$ are defined as above. In this case, $\dot{\beta}_{WX}$ is equal to the standardized moderation effect *only if* the independent variable and the moderator are uncorrelated. As this correlation increases, the standardized coefficient of the product term will increase and will be larger than the standardized moderation effect by a factor of $\sqrt{1+\rho_{WX}^2}$. That is, $\dot{\beta}_{WX}$ always overestimates this moderation effect except when the independent variable and the moderator are uncorrelated. As argued before (e.g., Hayes, 2018), this coefficient should not be reported, or at least should not be interpreted.

Two Recently Proposed Measures Applicable to Moderated Multiple Regression

Liu and Yuan (2020) proposed several measures of moderation effect based on the proportion of variance explained. Two of them are applicable to moderated multiple regression: ΔR_{mo}^2 and VR_2^2, the latter reduced to $VR_{2(N)}^2$ when X and W are multivariate normally distributed. Liu and Yuan argued that these measures are more appropriate than R^2 increase because they use the appropriate denominator to assess the variance explained attributable to the moderator. Appendix C shows that, if X and W are multivariate normally distributed as in $VR_{2(N)}^2$, ΔR_{mo}^2 and $VR_{2(N)}^2$ are functions of $\dot{\beta}_{WX}$ and other parameters: ρ_{XW}, ρ_{WY}, and $\dot{\beta}_{WY(X,W)}$ (the standardized regression coefficient of W when using X and W in prediction) in ΔR^2_{mo}; ρ_{XW} and $\dot{\beta}_X$ in $VR^2_{2(N)}$. Unlike R^2 increase, even if $\rho_{XW} = 0$, ΔR^2_{mo} and VR^2_2 are not monotonic functions of the standardized moderation effect. In sum, ΔR^2_{mo} and VR^2_2 quantify moderation effect in terms of proportion or ratio of variances but cannot replace the standardized moderation effect because they are also affected by other parameters. On the other hand, the standardized moderation effect informs researchers of the form of the moderation effect but does not reflect the magnitude of the moderation effect in terms of variance explained. Therefore, the standardized moderation effect and the pair of measures, namely ΔR^2_{mo} and VR^2_2 complement each other, addressing different research questions.

Should Standardized Moderation Effect Always be Reported? It Depends

Standardization is just a change of units (Hayes, 2018). Therefore, our recommendation is to compute and report the standardized moderation effect when (a) the original units of all variables involved in the moderation are arbitrary, or (b) the researchers want to compare the magnitude of a moderation effect with those in other studies using different samples or even different measuring instruments of the same constructs. If only some, but not all, variables involved have meaningful units, or if either the independent variable or the moderator is a categorical variable, then standardization can be done for only those variables that can benefit from the change in units before computing the product term, as demonstrated by Hayes (2018). The R package we developed can do this easily. The function `stdmod_selected_boot` allows users to specify which variables will be standardized. The bootstrapping confidence interval computation will take this into account. If only some of the variables are standardized, we can call the moderation effect a *partially standardized moderation effect*. The R functions `stdmod_selected_boot` will also skip those categorical variables in standardization automatically because their coefficients are not interpretable if standardized (Hayes, 2018).

Application of the standardized moderation effect assumes that the standard deviation of *X* does not depend on the level of *W*. As shown before (e.g., Smithson & Shou, 2017), if the standard deviation of *X* also changes with the level of *W*, then the standardized moderation effect is not an appropriate measure of moderation effect in standardized metric. If *W* is categorical and standardized measure of moderation effect is desired, multisample technique is needed to compute the partially standardized moderation effect within each sample. If *W* is continuous, then it may not be possible to meaningfully compute the standardized moderation effect, because it is not easy to compute the standard deviation of *X* and *Y* conditioned on *W*.

Future Directions

Four issues warrant further exploration. First, it is not yet clear how to analytically form the confidence intervals of the standardized moderation effect. Yuan and Chan (2010) demonstrated that merely rescaling the confidence interval would not yield the correct confidence interval. They and others (e.g., Dudgeon, 2017; Jones & Waller, 2013) proposed methods to obtain the standard error and the confidence interval for standardized regression coefficients. To extend these methods to standardized moderation effects involving product variables, one possibility is nonparametric bootstrapping, which is implemented in the package `stdmod`. Interval estimation of standardized coefficients is a new topic, calling for more efforts to extend previous work to standardized moderation effects. Second, the convention we proposed for categorizing standardized moderation effects was derived from another commonly used convention. More work can be conducted to review previous studies with moderation effects, with a view to provide an empirical basis for this way of categorizing standardized moderation effects. Third, work needs to be done for interval estimation of standardized moderation effect when the moderator is categorical. Shieh (2019) presented an innovative way to quantify the effect of a categorical moderator on the standardized metric. Procedures need to be developed to form the confidence intervals of the standardized moderation effect between specific pairs of categories. Lastly, for cases where the independent variable is categorical and so only the moderator and/or outcome variable can be meaningfully standardized, novel methods to generate confidence intervals are also to be devised.

References

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

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, *14*(1), 63–82. https://doi.org/10.1093/pan/mpi014

Cheung, M. W.-L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods, 41*(2), 425–438. https://doi.org/10.3758/BRM.41.2.425

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Lawrence Erlbaum Associates.

Craig, C. C. (1936). On the frequency function of xy. *The Annals of Mathematical Statistics, 7*(1), 1–15. https://doi.org/10.1214/aoms/1177732541

Edwards, J. R. (2009). Seven deadly myths of testing moderation in organizational research. In *C. E. Lance (Ed.) & R. J. Vandenberg (Ed.), Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 143–164). Routledge.

Efron, B. & Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall.

Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science, 26(4)*, 797. https://doi.org/10.2307/2110973

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd Ed.). The Guilford Press.

Hobert, R., & Dunnette, M. D. (1967). Development of moderator variables to enhance the prediction of managerial effectiveness. *Journal of Applied Psychology*, *51*(1), 50–64. https://doi.org/10.1037/h0024239

Jones, J. A., & Waller, N. G. (2013). Computing confidence intervals for standardized regression coefficients. *Psychological Methods, 18*(4), 435–453. https://doi.org/10.1037/a0033269

Liu, H., & Yuan, K.-H. (2020). New measures of effect size in moderation analysis. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000371

Markwart, H., Bomba, F., Menrath, I., Brenk-Franz, K., Ernst, G., Thyen, U., Hildebrandt, A., & Schmidt, S. (2020). Assessing empowerment as multidimensional outcome of a patient education program for adolescents with chronic conditions: A latent difference score model. *PLoS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.0230659

R Core Team. (2020). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Smithson, M., & Shou, Y. (2017). Moderator effects differ on alternative effect-size measures. *Behavior Research Methods, 49(2),* 747–757. https://doi.org/10.3758/s13428-016-0735-z

Yuan, K.-H., & Chan, W. (2011). Biases and standard errors of standardized regression coefficients. *Psychometrika*, *76*(4), 670–690. https://doi.org/10.1007/s11336-011-9224-6