International Journal of Behavioral Medicine (2019) 26:577–579
https://doi.org/10.1007/s12529-019-09811-5

INTEGRATIVE REVIEW

Statistical Guideline #3: Designate and Justify Covariates A Priori, and Report Results With and Without Covariates

Suzanne C. Segerstrom [1]

Published online: 14 August 2019

Abstract
From the Editors: This is one in a series of statistical guidelines designed to highlight common statistical considerations in behavioral medicine research. The goal was to briefly discuss appropriate ways to analyze and present data in the *International Journal of Behavioral Medicine* (IJBM). Collectively, the series will culminate in a set of basic statistical guidelines to be adopted by IJBM and integrated into the journal's official Instructions for Authors and also to serve as an independent resource. If you have ideas for a future topic, please email the Statistical Editor, Suzanne Segerstrom at segerstrom@uky.edu.

Keywords Statistical guidelines · Covariates · Statistical control · Statistical adjustment

The Statistics Guru

It is probably fair to say that *what* covariates to include in a model, *how*, and *when* are among the least standardized decisions in the behavioral sciences. The third statistical guideline for IJBM is a recommendation for authors to (1) report findings without covariates, (2) justify a set of a priori covariates, and (3) stick to that analysis plan. Covariates can have different purposes. Covariates can reduce "error" variance and increase the power and precision of estimates, and they can indicate the amount of overlap between two explanatory or predictor variables. However, they can also decrease power, precision, and construct validity, and be used post hoc in disreputable ways.

A covariate (or set of covariates) can reduce error variance in the outcome (or dependent) variable and allow the explanatory (or independent) variable(s) to account for more of the remaining variance. For example, in a study of body weight and blood pressure, three automated blood pressure cuffs might yield slightly different results. If which cuff was used was not associated with explanatory variables, covarying cuff removes error variance (see Fig. 1). As a consequence, the ratio of reliable:unreliable variance increases (reliable variance stays the same and unreliable variance decreases), and the explanatory variable can therefore account for a larger proportion of the remaining variance (e.g., in a regression model, R^2 for the explanatory variable will be higher).

More controversially, a covariate might be selected to show that the effect of the explanatory variable is not due to the "confounding" effect of the covariate. For example, in the same study, adjusting for age might be intended to show that the effects of body weight are independent of age. This use can be more problematic, and therefore, covariates should be used sparingly and thoughtfully.

One potential problem in both usages is that covariates can bias estimation. Covariates might not be measured with the same precision or attention as primary study variables. When a covariate is measured with error, the relationship between the explanatory and outcome variables can be overestimated, yielding Type I error, or underestimated, yielding Type II error. In either case, the result is misleading [1]. This is one reason why reliabilities of all variables in a model, including covariates, should be reported for the analytic sample (see Statistical Guideline #2, [2]).

The zero-order relationship between the explanatory/independent variable and the outcome/dependent variable should always be reported in addition to the adjusted relationship, because covariates can interfere with the reliability and construct validity of the explanatory and outcome variables. Lynam and colleagues [3] and Miller and Chapman [4] give extended examples of this phenomenon in psychopathology. In general, the more highly a covariate correlates with the

✉ Suzanne C. Segerstrom
segerstrom@uky.edu

[1] Department of Psychology, University of Kentucky, 125 Kastle Hall, Lexington, KY 40506-0044, USA

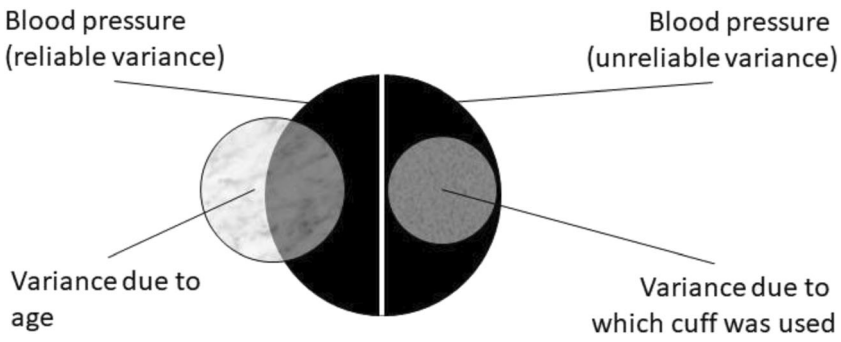

Fig. 1 Venn diagram in which blood pressure hypothetically has 50% reliable variance (on the left) and 50% unreliable or error variance (on the right), and overlap with age and method variance is illustrated

variable(s) of interest, the more it reduces that variable's reliability and changes its meaning. In some cases, "the relations may vary to such a degree that results from analyses that involve partialling cannot be applied to the original constructs that inspired the measures of the variables" [3, p. 330]. That is, the ability to invoke the original construct and its nomological net [5] may not survive the inclusion of covariates. In Fig. 1, adjusting for age removes reliable variance. As a consequence, the ratio of reliable:unreliable variance *decreases*, and the nature of "blood pressure" has changed: both effects are "perils of partialling" [3].

The justification and selection of covariates a priori should be described in the "Methods" section and, ideally, preregistered (e.g., at osf.io or clinicaltrials.gov). Covariates are part of investigator degrees of freedom, that is, those design and analysis decisions that affect the results. When investigator degrees of freedom are used post hoc to decrease p values (e.g., by selectively including covariates to achieve a $p < 0.05$), inflation bias or "p-hacking" occurs [6]. In a comparison of studies that reported results only with a covariate versus studies that reported results without the covariates (whether or not an adjusted result was also reported), the set of studies that reported results only with covariates lacked evidential value [7]. That is, the pattern of results was consistent with the presence of selective reporting (reporting only some of the analyses that were performed) and "p-hacking" (using investigator degrees of freedom to achieve $p < 0.05$). This is another reason to report the zero-order relationship between the explanatory/independent variable and the outcome/dependent variable in addition to the adjusted relationship.

Covariates are sometimes selected empirically, based on their statistical associations with the substantive variables (typically the explanatory/independent variable(s) and/or outcome/dependent variable). This approach is not recommended. When different investigators use different covariates, the results cannot be compared directly because the meaning of the substantive variables is different (see effects on construct validity, above). This approach is contrary to cumulative science, in which knowledge of a domain increases continually as a function of accumulating evidence. For this to happen, study constructs need to be comparable. Second, selecting model variables based on their p values can result in the "wrong" variables being included, particularly when the sample size is small [8]. (This is also true when one uses automated stepwise selection in multiple regression analysis.) Quite simply, do not do this. If including all covariates would use an unacceptable number of degrees of freedom, it is preferable to use a data reduction technique (e.g., principal components analysis of the covariates and inclusion of component scores in the model) than to use an idiosyncratic selection method.

Some people might argue that if they use component scores to adjust for covariates, they will not know which covariate was most important. This argument suggests that the covariates might have a substantive role in the model and not merely a "controlling" role. In the example above, perhaps a model in which age is an "upstream" predictor of body weight and blood pressure is reasonable. In other cases, covariates might be mediators. Perhaps the relationship between body weight and blood pressure is mediated by physical activity? Before including covariates to rule out other sources of variance, think theoretically about whether their variance should be "erased" [9]. Although covariates are used frequently in epidemiology and public health as well as behavioral medicine, they may be critical variables for understanding health behavior processes.

Finally, sometimes, covariates are used in moderated models, as "control" for the effect of an interaction. This approach requires careful thought, as a covariate might control for error variance in the outcome, in which case its main effect belongs in the model. However, sometimes, the model calls for an "alternative moderator" [10]. For example, an interaction between optimism and task effort predicted subsequent cellular immune response, measured by injecting a small amount of antigen into the skin [11]. If different batches of antigen were used to test immune response, then including a covariate indicating which batch was used could reduce error variance in this measure [12]. However, to test whether the optimism effect overlapped with that of neuroticism or conscientiousness, controlling for the main effects of these personality dimensions would not suffice. The *interaction* between neuroticism or conscientiousness and task effort

belonged in the model. (Spoiler alert: the overlap with conscientiousness accounted for the optimism interaction.)

The thoughtful use of covariates can help us make more precise estimates, when they are measured reliably and are used to reduce *error* variance. They can help us understand what part of a construct (e.g., the part of age that overlaps with body weight) accounts for variance in another (blood pressure). They also can, however, involve pitfalls and undermine cumulative science. This guideline is intended to maximize their advantages and minimize their disadvantages.

Compliance with Ethical Standards

Conflict of Interest The author declares that she has no conflicts of interest.

Human and Animal Rights This article does not contain any studies with human participants or animals performed by the author, and so there was no requirement for informed consent.

References

1. Liu K. Measurement error and its impact on partial correlation and multiple linear regression analyses. Am J Epidemiol. 1988;127:864–74.
2. Segerstrom SC. Statistical Guideline #2. Report appropriate reliability for your sample, measure, and design. Int J Behav Med. In press.
3. Lynam DR, Hoyle RH, Newman JP. The perils of partialling: cautionary tales from aggression and psychopathy. Assessment. 2006;13:328–41.
4. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52:281–302.
5. Miller GA, Chapman JP. Misunderstanding analysis of covariance. J Abnormal Psychol. 2001;110:40–8.
6. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biol. 2015;13:e1002106.
7. Simonsohn U, Nelson LD, Simmons JP. *P*-curve: a key to the file-drawer. J Exp Psychol Gen. 2014;143:534–47.
8. Flack VF, Chang PC. Frequency of selecting noise variables in subset regression analysis: a simulation study. Am Stat. 1987;41:84–6.
9. Revenson TA. All other things are *not* equal: an ecological perspective on the relation between personality and disease. In: Friedman HS, editor. Personality and disease. New York: Wiley; 1990. p. 65–94.
10. Hull JG, Tedlie JC, Lehn DA. Moderator variables in personality research: the problem of controlling for plausible alternatives. Personal Soc Psychol Bull. 1992;18:115–7.
11. Segerstrom SC, Castañeda JO, Spencer TE. Optimism effects on cellular immunity: testing the affective and persistence models. Pers Ind Diff. 2003;35:1615–24.
12. Segerstrom SC, Smith GT. Methods, variance, and error in psychoneuroimmunology research: the good, the bad, and the ugly. In: Segerstrom SC, editor. Oxford handbook of psychoneuroimmunology. New York: Oxford; 2012. p. 421–32.