AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology of Addictive Behaviors

ISSN: 0893-164X

2022, Vol. 36, No. 3, 254–259
https://doi.org/10.1037/adb0000832

INTRODUCTION

Harmonization of Theory, Methods, and Inferences in Addictions Research

Kevin M. King and Jonas Dora
Department of Psychology, University of Washington

The articles in the present special section highlight four ways in which our applications of methods, and their harmonization with theory, can hold us back, and each offers an avenue for improvement that brings us closer to our goal of building a cumulative scientific record of the study of addiction. It brings together four articles that are intended to provide new ideas and directions for research on addictive behaviors. It is important for researchers to consider how their study designs, measurements, and statistical tests are specific expressions of the theories they wish to test. Each article illustrates a dimension of the gaps between theory and methods, provides an illustrated example of how to bridge those gaps, and provides easy to follow advice for how to apply these ideas in our own work. By designing for replication (Pearson et al., 2021), considering model-theory harmonization (Littlefield et al., 2021), moving toward plain language interpretation of effects (Halvorson et al., 2021), and thinking of models across levels of analysis (Soyster et al., 2021), we can move toward a more robust, replicable, and impactful science of addictive behaviors.

Keywords: research methods, theory, statistical analysis

For scientific progress in addictions research to help the communities we aim to serve, we need to be able to trust, understand, and evaluate the body of work that we build our future studies on. Although psychologists have made tremendous progress, there is still a long way to go. Methodological failures in psychological sciences have received a great deal of attention, focus, and efforts at reform. Prominent theories have either failed to replicate or seen their replication effect sizes dwindle (e.g., ego depletion; Hagger et al., 2016; Vohs et al., 2021), entire lines of inquiry have been shown to be misleading or downright false (e.g., candidate gene studies; Border et al., 2019), measurement modalities have been shown to capture little more than noise (e.g., inhibition tasks; Enkavi et al., 2019), and some literatures have been shown to be constructed largely of false positives (Open Science Collaboration, 2015). Much of the effort and attention paid toward rectifying these problems has been directed at improving the transparency of the research process, such as implementing preregistration, registered reports, or open sharing of data (Tackett et al., 2017, 2019). Other reform efforts have turned attention to methodological improvements, such as how to improve the implementation of quantitative methods in applied settings (King et al., 2019), providing tutorials on specific methods (McCabe et al., 2018, 2021a), or providing improved tools to use existing methods (McCabe et al., 2021b). Some methodological reviews have focused on the importance of harmonizing theory and methods (Collins, 2006). Even more recently, some have argued that improvements in theories themselves are critical to improving progress in science. For example, Devezer et al. (2021) have argued that, in the same way that highly reliable measures can still be poor measures of a latent construct, highly replicable findings may nonetheless be poor tests of a proposed theory. Many authors have noted that most theories in psychological science are vague, verbal theories that rarely translate into specific, testable hypotheses (Robinaugh et al., 2021). Each of these methodological, statistical, and theoretical issues are critical to address if we hope to build a reliable and cumulative science of addictive behaviors.

The articles in the present special section highlight four ways in which our applications of methods, and their harmonization with theory, can hold us back, and each offers an avenue for improvement that brings us closer to our goal of building a cumulative scientific record of the study of addiction.

Pearson et al. (2021). Examining Replicability in Addictions Research: How to Assess and Ways Forward

Replicability is a core feature of scientific progress—barring some exceptions (e.g., studying the effects of a unique event on a psychological outcome of interest), replication should considerably increase our confidence in the original findings. Given that research is incremental, without an understanding of the replicability of key findings it is impossible to know whether we are building our theoretical castles on a solid foundation, or merely on a pile of low powered p-hacked effects. Recently, several articles have been published that should make us optimistic that high replicability is achievable if we improve the rigor of our methods (Protzko et al., 2020; Soderberg et al., 2021).

Editor's Note. Katie Witkiewitz served as the action editor for this article.—KW

Kevin M. King https://orcid.org/0000-0001-8358-9946

Kevin M. King played lead role in conceptualization and writing of original draft and supporting role in writing of review and editing. Jonas Dora played lead role in writing of review and editing and supporting role in writing of original draft.

Correspondence concerning this article should be addressed to Kevin M. King, Department of Psychology, University of Washington, P.O. Box 351525, Seattle, WA 91895, United States. Email: kingkm@uw.edu

In their thought-provoking article, Pearson and colleagues take a critical look at our field on a meta-level, noting that addiction researchers have not attempted to estimate the replicability of seminal findings in our field to the same extent as other fields of psychology, such as social, experimental, and personality psychology (Camerer et al., 2018; Soto, 2019). Some evidence may suggest that at least some addictions research may have poor replicability. For example, reviews of clinical psychological science have noted that samples are often underpowered to detect small effects (Reardon et al., 2019), which in all likelihood translates to many of the seminal findings in addiction research based on smaller, select samples of individuals with addictive behaviors. Other reviews have suggested that the evidence supporting even "evidence-based treatments" is often very weak and based on small samples with low replicability (Sakaluk et al., 2019). On the other hand, the replicability in psychological science has been shown to vary broadly across many subdisciplines that addictions research draws upon (Fraley & Vazire, 2014). Regardless, it is hard to gage to what extent influential findings in addiction research are replicable because of a lack of attention to the matter (Heirene, 2021).

While it may be easy to decide that replicating important findings in the field is a good idea, it is less straightforward to conduct high-quality replication research and to evaluate whether or not a finding replicated (Nosek et al., 2022). Pearson and colleagues provide researchers with an excellent overview of different replicability metrics, as well as their respective advantages and disadvantages. Often, replications may be difficult or expensive because of the unique nature of a sample. In their article, they illustrate one way of testing the replicability of a finding: by treating data collection sites of a large-scale collaborative study as replications of one another, the authors illustrate that meta-analyzing and focusing on raw effect sizes give us a better picture of replicability compared to a focus on statistical significance. Their findings also show that replication studies are unlikely to be informative unless they are higher-powered (and thus more precise) than the original. Other approaches to understanding the replicability of a finding include individual participant data meta-analysis, which allow the application of a harmonized data analysis protocol (e.g., Dora et al., 2022). Another recent development is the mini meta-analysis (Goh et al., 2016), which is possible when researchers incorporate similar instruments across multiple studies, essentially building replication of multiple findings into the regular research process (e.g., King et al., 2018, 2022). Hopefully, reading this article will (a) motivate you to value and incorporate replication studies in your research agenda and (b) provide you with hands-on advice with regard to maximizing the informativeness of your replication attempts. Given the comparatively high cost of replication in our field, work like this is much needed to provide us with the tools of assessing the progress (i.e., replication rate) of our efforts to increase our confidence in findings that we build our lines of research on.

Littlefield et al. (2021). Limitations of Cross-Lagged Panel Models in Addiction Research and Alternative Models: An Empirical Example Using Project MATCH

Littlefield and colleagues' work illustrates the importance of tying our statistical models to specific versions of our verbal theories. There is often a broad "inference gap" in psychological studies, which reflects the difference between statistical models and the theories they test (Yarkoni, 2022). Although statistical models are very precise in how they describe relations among variables, they typically are used to draw inferences about much less precise psychological theories. For example, the affect regulation theory of alcohol use, which posits that alcohol use is regulated by emotional experiences (e.g., Baker et al., 2004; Cox & Klinger, 1988), has been tested by examining the associations between internalizing symptoms and alcohol use and disorder symptoms across young adulthood (King et al., 2020), the associations between coping motives and alcohol use both between- (Littlefield et al., 2012) and within-persons (Stevenson et al., 2019), and the within-day associations between affect and alcohol use (Dora et al., 2022). Each of these, however, tests a *specific different version* of affect regulation. For example, King et al. (2020) tested a *between-persons* (or a "people who") *hypothesis*: Do *people who* reported more symptoms of internalizing disorders also report higher levels and larger changes in alcohol use and alcohol use disorder (AUD). On the other hand, Dora et al. (2022) tested a *within-persons* (or "when people") hypothesis: *When people* experiences higher negative or positive affect than usual, were they more likely to drink or to drink more?

Littlefield et al. (2021) focus on the inference gap in the cross-lagged panel model (CLPM), which has long been a workhorse model in developmental research. In short, CLPMs offer to test the plausibility of hypotheses about reciprocal causation between constructs over time, such as "Does depression lead to AUD, or does AUD lead to depression, or both?" However, there is a fundamental mismatch between what the CLPM can test, and the theories that researchers wish to test. In short, researchers frequently wish to use CLPMs to test "*when people*" theories, like "when people are depressed, they are likely to later develop an AUD, but not vice versa." Almost by definition, researchers who use CLPMs are not interested in testing "people who" hypotheses, like "people who tend to be depressed also tend to meet criteria for AUD."

Littlefield and colleagues offer a clear demonstration of the kinds of restrictive assumptions made by CLPM that essentially make them implausible in most cases, especially when there is a "*when people*" hypothesis. In short, for variables that are measured repeatedly over time, variance at any one time point is likely attributable to both between- and within-person variance. Although sometimes described as "traits" and "states," these simply describe a disaggregation of variance to two levels of time: that which is stable across the time observed, and that which isn't. For example, even stressful life events that are uncontrollable and external to the individual can be shown to occur at different average frequencies for different people over time (King et al., 2008). Recently, several methodological critiques of the CLPM have been forwarded, arguing that CLPMs conflate within- and between-person variance, which makes it impossible to attribute cross-lagged effects in a standard CLPM to within-person processes (Curran et al., 2014; Hamaker et al., 2015). Littlefield extends this research and provides illustrated examples of the inferences that can be drawn from a traditional CLPM versus more recent alternatives which are designed to explicitly separate within- and between-person variance. Littlefield and colleagues also highlight fundamental problems with the logic of the CLPM. First, they highlight the restrictive assumptions of the CLPM, which by definition is identical to a random-intercept CLPM that forces between-person variance to be zero. Although researchers frequently believe that controlling for an outcome at a prior time point provides a test of "stability," of the outcome, Littlefield demonstrates how this is not the case. For example, by doing the path

tracing, they demonstrate that the standard CLPM is identical to both a series of two predictor regressions. In other words, CLPMs are no more than a series of regressions in a trench coat. Although researchers frequently wish to interpret predictors in CLPMs as though they were residualized on prior time points (e.g., controlling for earlier measurements of the predictor), without explicitly including those lags as predictors, this is not the case. The issues highlighted by Littlefield's work on the CLPM reflect a broader problem that the field must grapple with: we must do a better job of being clear exactly what our statistical models can and cannot test about a theory, and what specific form of a theory our models are testing. In the same way that multiple well-fitting alternative structural models can equally explain data well (Tomarken & Waller, 2003), researchers should keep in mind that the same theory might be tested in many different ways, and to keep in mind the importance of the scientific goal of modeling rather than just the statistical goal of fitting the data well (Navarro, 2019). As Littlefield and colleagues demonstrate, many of these ways may provide a different, nuanced information about a theory, while others (such as the CLPM) may provide muddled, incomplete, and misleading information that does little to advance our knowledge about a theory. This article teaches us about the importance of translating our verbal hypotheses into the specific "language" of the statistical tests we use in any given study, which places those tests in a larger body of evidence about a hypotheses in a more precise way.

Halvorson et al. (2021). Making Sense of Some Odd Ratios: A Tutorial and Improvements to Present Practices in Reporting and Visualizing Quantities of Interest for Binary and Count Outcome Models

Even once we've settled on a model as a specific test of some theory, we need to provide an interpretation of that test in a way that effectively communicates our findings to our target audiences. Although there is a strong case to be made that psychological science undervalues prediction (e.g., explaining variation in some target outcome; Yarkoni & Westfall, 2017), most research focuses on explanation. In other words, researchers wish to come to some mechanistic understanding of the world, and hypothesis tests are represented by estimating specific parameters of interest. For example, if two people differ in their depression symptoms by one standard deviation, how does their likelihood of developing an AUD differ? If a person has a day where their negative emotions are one standard deviation higher than what is normal for them, how much more likely are they to drink? The importance of focusing on effect sizes and uncertainty (e.g., confidence or credible intervals) has long been noted (Cohen, 1992; Cumming, 2014; Pek & Flora, 2018), and despite APA guidelines to report effect sizes and confidence intervals, many empirical articles manuscripts fall short of these ideals. This is especially true when researchers interpret their results in a discussion section, where we commonly see descriptions of which effects were or were not significant, with no attention to the magnitude of effects or the uncertainty around them. Translating effects to plain, understandable language, rather than expecting readers to refer to tables, is important for effectively communicating our results to other researchers and to key stakeholders (Flora, 2020). For example, a recent study reported that longer item batteries were related to lower compliance in an EMA study (Eisele et al., 2022). Does that information alone help researchers make a decision to use a longer versus a shorter item battery, or would it help to know that doubling the number of items from 30 to 60 was associated in a 5% (±2%) decrease in compliance, from 89% to 84%?

Reporting and interpreting effect sizes is even more challenging when outcomes are nonlinear, such as models predicting the presence or absence of diagnosis, or the number of drinks a participant consumed in an evening. Halvorson et al. (2021) offer a tutorial on how to interpret and understand the odds and risk ratios that arise from these nonlinear models. Building on work by King et al. (2000), the authors expand on the importance of thinking of effect sizes in terms of *quantities of interest* (QoIs) in reporting findings from general linear models (GLMs): What are the empirical quantities that map directly on to the research question at hand? For linear GLMs and their extensions (such as an ordinary least squares regression or a factor loading), the QoIs are straightforward and constant across the whole range of a predictor. However, when predicting nonlinear outcomes, such as binary diagnoses or count outcomes (e.g., number of drinks), model coefficients reflect the association of a predictor and a transformed version of the QoI (such as a logit). Even common transformations of these coefficients, such as an odds or risk ratio do not reflect quantities that are readily interpretable, even by experts. Moreover, because of the nonlinearity of the QoI, coefficients from nonlinear GLMs are conditional on the level of other covariates, *even when a product term (a common test of moderation) is not included.* Unfortunately, Halvorson and colleagues show in a random sample of 52 articles reporting results from nonlinear GLMs that 95% fail to interpret results in terms of a QoI, and the vast majority (69%) only report coefficients in terms of significance and direction. This represents a clear failure of researchers to interpret effects in ways that researchers and other stakeholders can understand and act upon. Halvorson and colleagues provide clear, concrete, and actionable recommendations for how to interpret and present results from these models which are very common in addictions research. They recommend including graphical presentation of key models in terms of QoIs, in a way that readers can intuitively understand, and to plot those models across key values of predictors to illustrate how key associations vary as a function of meaningful covariates. Moreover, they provide examples of how to present tables of predicted results (such as predicted probabilities and counts) to provide numerical references to supplemental figures, as well as code examples to facilitate adoption of these practices. Halvorson et al. (2021) provide a way to facilitate comparison across studies and to increase the impact of research articles without increasing burden on readers. This work and its applications have immediate and important implications for research as well as intervention. For example, imagine finding that a one standard deviation within-person change in positive mood doubles the odds of engaging in binge drinking. Should a just-in-time adaptive intervention be delivered when a one standard deviation change in positive mood is detected? Halvorson's work highlights that without knowing the impact of other covariates, we might run the risk of delivering interventions too often in situations (such as weekday mornings) when they are not needed, spoiling our chances to deliver effective interventions when they are (such as weekend afternoons). Given the complex outcomes that we often deal with in addictions research, the recommendations of Halvorson and colleagues will help us to more clearly and precisely communicate our

findings, which will in turn help us to clearly understand and build on each other's findings.

Soyster et al. (2021). Pooled and Person-Specific Machine Learning Models for Predicting Future Alcohol Consumption, Craving, and Wanting to Drink: A Demonstration of Parallel Utility

Finally, it is important to ensure that we draw conclusions from research at the appropriate level of analysis. In the same way that CLPMs conflate variance between people and variance within people, many of our models are group-level models (e.g., nomothetic) which make strong assumptions about individual-level variation. For example, growth mixture models are often (incorrectly) promoted as "person-centered" models (as opposed to "variable-centered" models like growth curve models) because they estimate classes of people with similar patterns of variation in within-person change over time. However, both types of models estimate a single (growth curve models) or multiple (growth mixture models) mean trajectory of change and variation around that mean assuming a normal distribution (King et al., 2018). In other words, these are "top-down" models which assume that within-person change in a population may be represented well by mean and standard deviation. Idiographic models, on the other hand, start from the "bottom-up," and model psychological processes as person-specific (Molenaar, 2004; Wright et al., 2019). Recently, novel methods have been introduced that attempt to integrate idiographic and nomothetic approaches (Foster & Beltz, 2018), which allows researchers to advance our understanding of problematic substance use and related harmful behaviors on a population level while simultaneously conducting research that is useful to the prevention and treatment on the level of the individual. Because of a tendency in the field to focus on nomothetic models, there is a tension between group-level findings and what researchers, providers, and patients experience. For example, our recent finding, using nomothetic methods, that negative affect is entirely unrelated to daily alcohol use conflicts with decades of theory and clinical intuition about the affect regulation of alcohol use (Dora et al., 2022). One explanation for this null finding is that our work ignored the idiographic processes that link negative affect and alcohol use. If this association is only true for some people, *and* if the shape of that association (such as the influence of contextual factors, timing of affect, etc.) varies among those people for whom it *is* true, a nomothetic approach like the one we used would not capture a true person-specific effect.

One fruitful solution to this dilemma might be to apply nomothetic and idiographic analyses in parallel to test to what extent our results generalize from the group-level to the individual-level and vice versa. Soyster and colleagues provide a nice example of what such a parallel approach could look like. First, it is important to note that such complimentary analyses require study concept and design to be appropriate for an idiographic analysis, which in simple terms requires more dense and frequent data from each individual participant. The authors use cross-validation to explore to what extent their nomothetic and idiographic prediction model of alcohol consumption generalize in a holdout sample. Their results indicated that both pooled and individual analyses were able to predict alcohol use, and they note important strengths of both approaches (i.e., the ability of nomothetic approaches to capture data generating processes that may generalize to individuals in the broader population and the ability of idiographic approaches to inform individualized just-in-time interventions and clinical applications more generally). Given that we often want to cumulatively build both our understanding of group-level and individual-level processes in our field, we believe that this article nicely illustrates the utility in pursuing both nomothetic and idiographic analyses, so that we can draw conclusions at the population and the individual level.

Conclusion

This special section brings together four articles that are intended to provide new ideas and directions for research on addictive behaviors. It is important for researchers to consider how their study designs, measurements, and statistical tests are specific expressions of the theories they wish to test. By designing for replication, considering model-theory harmonization, moving toward plain language interpretation of effects, and thinking of models across levels of analysis, we can move toward a more robust, replicable, and impactful science of addictive behaviors.

References

Baker, T. B., Piper, M. E., McCarthy, D. E., Majeskie, M. R., & Fiore, M. C. (2004). Addiction motivation reformulated: An affective processing model of negative reinforcement. *Psychological Review*, *111*(1), 33–51. https://doi.org/10.1037/0033-295X.111.1.33

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *The American Journal of Psychiatry*, *176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505–528. https://doi.org/10.1146/annurev.psych.57.102904.190146

Cox, W. M., & Klinger, E. (1988). A motivational model of alcohol use. *Journal of Abnormal Psychology*, *97*, 168–180. https://doi.org/10.1037/0021-843X.97.2.168

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*(5), 879–894. https://doi.org/10.1037/a0035297

Devezer, B., Navarro, D. J., Vandekerckhove, J., & Ozge Buzbas, E. (2021). The case for formal methodology in scientific reform. *Royal Society Open Science*, *8*(3), Article 200805. https://doi.org/10.1098/rsos.200805

Dora, J., Piccirillo, M., Foster, K. T., Arbeau, K., Armeli, S., Auriacombe, M., Bartholow, B. D., Beltz, A. M., Blumenstock, S. M., Bold, K., Bonar, E., Braitman, A., Carpenter, R. W., Creswell, K., DeHart, T., Dvorak, R., Emery, N. N., Enkema, M., Fairbairn, C., … King, K. M. (2022). *The daily association between affect and alcohol use: A meta-analysis of*

individual participant data [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/xevct

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5472–5477. https://doi.org/10.1073/pnas.1818430116

Flora, D. B. (2020). Thinking about effect sizes: From the replication crisis to a cumulative psychological science. *Canadian Psychology*, *61*(4), 318–330. https://doi.org/10.1037/cap0000218

Foster, K. T., & Beltz, A. M. (2018). Advancing statistical analysis of ambulatory assessment data in the study of addictive behavior: A primer on three person-oriented techniques. *Addictive Behaviors*, *83*, 25–34. https://doi.org/10.1016/j.addbeh.2017.12.018

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, *9*(10), Article e109019. https://doi.org/10.1371/journal.pone.0109019

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how: Mini meta-analysis. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., … Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Halvorson, M. A., McCabe, C. J., Kim, D. S., Cao, X., & King, K. M. (2021). Making sense of some odd ratios: A tutorial and improvements to present practices in reporting and visualizing quantities of interest for binary and count outcome models. *Psychology of Addictive Behaviors*, *36*(3), 284–295. https://doi.org/10.1037/adb0000669

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Heirene, R. M. (2021). A call for replications of addiction research: Which studies should we replicate and what constitutes a "successful" replication? *Addiction Research and Theory*, *29*(2), 89–97. https://doi.org/10.1080/16066359.2020.1751130

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*(2), Article 347. https://doi.org/10.2307/2669316

King, K. M., Feil, M. C., Halvorson, M. A., Kosterman, R., Bailey, J. A., & Hawkins, J. D. (2020). A trait-like propensity to experience internalizing symptoms is associated with problem alcohol involvement across adulthood, but not adolescence. *Psychology of Addictive Behaviors*, *34*(7), 756–771. https://doi.org/10.1037/adb0000589

King, K. M., Halvorson, M. A., Kuehn, K. S., Feil, M. C., & Lengua, L. J. (2022). Cross-Study, Cross-Method associations between negative urgency and internalizing symptoms. *Assessment*, *29*(3), 572–582.

King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., & Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. *Developmental Cognitive Neuroscience*, *33*, 54–72. https://doi.org/10.1016/j.dcn.2017.11.009

King, K. M., Molina, B. S. G., & Chassin, L. (2008). A state-trait model of negative life event occurrence in adolescence: Predictors of stability in the occurrence of stressors. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *37*(4), 848–859. https://doi.org/10.1080/15374410802359643

King, K. M., Pullmann, M. D., Lyon, A. R., Dorsey, S., & Lewis, C. C. (2019). Using implementation science to close the gap between the optimal and typical practice of quantitative methods in clinical science. *Journal of Abnormal Psychology*, *128*(6), 547–562. https://doi.org/10.1037/abn0000417

Littlefield, A. K., King, K. M., Acuff, S. F., Foster, K. T., Murphy, J. G., & Witkiewitz, K. (2021). Limitations of cross-lagged panel models in addiction research and alternative models: An empirical example using project MATCH. *Psychology of Addictive Behaviors*, *36*(3), 271–283. https://doi.org/10.1037/adb0000750

Littlefield, A. K., Talley, A. E., & Jackson, K. M. (2012). Coping motives, negative moods, and time-to-drink: Exploring alternative analytic models of coping motives as a moderator of daily mood-drinking covariation. *Addictive Behaviors*, *37*(12), 1371–1376. https://doi.org/10.1016/j.addbeh.2012.05.020

McCabe, C. J., Halvorson, M. A., King, K. M., Cao, X., & Kim, D. S. (2021a). Interpreting interaction effects in generalized linear models of nonlinear probabilities and counts. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2020.1868966

McCabe, C., Halvorson, M. A., King, K. M., Cao, X., & Kim, D. S. (2021b). *modglm: An R package for testing, interpreting, and displaying interactions in generalized linear models of discrete data* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6vmsa

McCabe, C. J., Kim, D. S., & King, K. M. (2018). Improving present practices in the visual display of interactions. *Advances in Methods and Practices in Psychological Science*, *1*(2), 147–165. https://doi.org/10.1177/2515245917746792

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Navarro, D. J. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain & Behavior*, *2*(1), 28–34. https://doi.org/10.1007/s42113-018-0019-z

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Pearson, M. R., Schwebel, F. J., Richards, D. K., & Witkiewitz, K. (2021). Examining replicability in addictions research: How to assess and ways forward. *Psychology of Addictive Behaviors*, *36*(3), 260–270. https://doi.org/10.1037/adb0000730

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208–225. https://doi.org/10.1037/met0000126

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., Buttrick, N., DeBell, M., Ebersole, C. R., Lundmark, S., MacInnis, B., O'Donnell, M., Perfecto, H., Pustejovsky, J. E., Roeder, S. S., Walleczek, J., & Schooler, J. (2020). *High replicability of newly-discovered social-behavioral findings is achievable* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/n2a9x

Reardon, K. W., Smack, A. J., Herzhoff, K., & Tackett, J. L. (2019). An N-pact factor for clinical psychological research. *Journal of Abnormal Psychology*, *128*(6), 493–499. https://doi.org/10.1037/abn0000435

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743. https://doi.org/10.1177/1745691620974697

Sakaluk, J. K., Williams, A. J., Kilshaw, R. E., & Rhyner, K. T. (2019). Evaluating the evidential value of empirically supported psychological treatments (ESTs): A meta-scientific review. *Journal of Abnormal Psychology*, *128*(6), 500–509. https://doi.org/10.1037/abn0000421

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), 990–997. https://doi.org/10.1038/s41562-021-01142-4

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, *30*(5), 711–727. https://doi.org/10.1177/0956797619831612

Soyster, P. D., Ashlock, L., & Fisher, A. J. (2021). Pooled and person-specific machine learning models for predicting future alcohol consumption, craving, and wanting to drink: A demonstration of parallel utility. *Psychology of Addictive Behaviors*, *36*(3), 296–306. https://doi.org/10.1037/adb0000666

Stevenson, B. L., Dvorak, R. D., Kramer, M. P., Peterson, R. S., Dunn, M. E., Leary, A. V., & Pinto, D. (2019). Within- and between-person associations from mood to alcohol consequences: The mediating role of enhancement and coping drinking motives. *Journal of Abnormal Psychology*, *128*(8), 813–822. https://doi.org/10.1037/abn0000472

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual Review of Clinical Psychology*, *15*, 579–604. https://doi.org/10.1146/annurev-clinpsy-050718-095710

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F., & Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, *12*(5), 742–756. https://doi.org/10.1177/1745691617690042

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578–598. https://doi.org/10.1037/0021-843X.112.4.578

Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., Alquist, J. L., Baker, M. D., Brizi, A., Bunyi, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J., … Albarracín, D. (2021). A multisite preregistered paradigmatic test of the ego-depletion effect. *Psychological Science*, *32*(10), 1566–1581. https://doi.org/10.1177/0956797621989733

Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*, *31*(4), 502–515. https://doi.org/10.1037/pas0000617

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, Article e1. https://doi.org/10.1017/S0140525X20001685

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Received March 8, 2022
Revision received March 9, 2022
Accepted March 11, 2022 ■