Behaviour Research and Therapy 98 (2017) 1–3

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Best practice guidelines for modern statistical methods in applied clinical research: Introduction to the Special Section

Modern statistical methods offer a variety of exciting analytic possibilities and statistical advantages over traditional methods. For example, state-of-the-art missing data methods (e.g., multiple imputation, direct maximum likelihood) are associated with higher statistical power and less parameter estimate bias compared to traditional ways of dealing with missingness (e.g., listwise or pairwise deletion of cases with missing observations) (cf. Enders, 2010, 2017). In addition to the ability to incorporate extant knowledge within the analytic model (informative priors), Bayesian estimation outperforms traditional (frequentist) statistical methods in many ways including better performance in small samples, lack of a reliance on asymptotic theory (does not assume the distribution of parameter estimates are normal on the basis of large-sample theory), missing data handling, higher statistical power in many data scenarios (e.g., non-normal sample distributions), and a more veridical estimation of the precision of parameter estimates (credibility intervals as opposed to confidence intervals in traditional statistical approaches) (cf. Baldwin & Larson, 2017; Kaplan, 2014).

Although innovations in various data analytic techniques have occurred rapidly (e.g., propensity score analysis), many of these developments have been overlooked or have been slow to find their way into applied mental health research. Alternatively, whereas some procedures such as growth and factor mixture modeling have begun to receive widespread application, key aspects of these approaches are often neglected or misapplied (e.g., proper consideration of the full range of possible models in the class enumeration phase of mixture modeling; adjustment for classification uncertainty when examining predictors or distal outcomes of the latent classes). In some instances, the adoption and proper use of these methods has been impeded by the lack of user-friendly sources prepared with the applied researcher in mind. Methodological developments are usually disseminated in journals or edited volumes that target a quantitatively oriented readership. It is often the case that the fundamental concepts and procedural aspects of these methods are fairly straightforward, yet it can be difficult to discern the practical application of the analytic approaches to clinical research problems when they are presented in formats steeped with quantitative terminology, equations, and symbols.

The goal of this Special Section is to provide practical overviews of selected topics in modern statistical methods and their best practice applications to clinical research. Although the potential range of topics that could be included in this Special Section is far-reaching, we aimed to recruit papers that addressed data analytic problems that are frequently encountered in the peer review process of *Behaviour Research and Therapy* (e.g., mismanagement of missing data, failure to evaluate and adjust for violations of model assumptions; improper testing of indirect and interaction effects), as well as papers that will hopefully propel the journal's readership to adopt more powerful, but oft-neglected, analytic techniques in their applied clinical research (e.g., conditional process analysis, propensity score analysis, Bayesian estimation). To foster the impact of the Special Section (i.e., maximize the chances that readers will correctly use the techniques in their research), the invited articles have an applied rather than a quantitative focus. Contributors were asked to keep the discussion of the mathematical underpinnings and equations to a minimum and to do so only when such presentation would foster the conceptual understanding of the method.

The papers in this Special Section illustrate the proper implementation of the selected methods in data sets and study designs that are germane to applied clinical psychology research (e.g., treatment outcome trials, observational studies of the temporal course of psychopathology), and discuss salient procedural, practical, and methodological issues, and advantages over traditional methods. Along these lines, the contributing authors were asked to present the best strategies for addressing pitfalls and methodological quandaries associated with the method (including any unresolved issues) that are often encountered in applied data sets (e.g., convergence issues, problems in model selection). Moreover, each paper contains one or more clinically relevant data-based example that is worked through fully and that is accompanied by supplementary material that includes input data, syntax, and output using readily accessible statistical software (e.g., Mplus, R, SPSS). These materials are available for download on the journal's companion website.

We had the good fortune of having all of our first choices agree to contribute to this Special Section, each of whom is a leading expert in their topic area. In the first paper, Enders (2017) provides a thorough presentation of the proper application of multiple imputation (MI) to handle missing data in a variety of data situations that are encountered frequently by applied clinical researchers. Generally speaking, direct maximum likelihood (ML), also known as full information maximum likelihood, is usually regarded to be the preferred method for accommodating missing data because it performs as well as MI in most situations and it is easier to implement (in fact, direct ML is currently a default in widely used software programs such as Mplus). Nonetheless, Enders describes some data scenarios where direct ML does not perform as well as MI (e.g., different types of regression models that entail a combination of categorical and continuous predictor

http://dx.doi.org/10.1016/j.brat.2017.06.008

or outcome variables; analyses using composite scores where some constituent items/ratings are missing; multilevel data). In addition to presenting MI's theoretical framework and the best practice implementation of MI in several examples, Enders illustrates the identification and inclusion of auxiliary variables to reduce bias and foster statistical power, and discusses several other important topics (e.g., significance testing, the analysis of interaction effects). It should also be clear from reading Enders's paper that older, but well-entrenched, approaches to missing data handling (e.g., listwise and pairwise deletion, carrying the last observation forward) have no place in contemporary applied clinical research, except in rare circumstances (e.g., very large samples in which the proportion of missing data is negligible; cf. Schafer & Graham, 2002).

Field and Wilcox (2017) remind us of the perils of carrying out statistical analyses without considering whether the sample data support the model's assumptions, or under the (usually false) belief that the statistical test is robust to violations of these assumptions. Indeed, in applied data sets, it is more common than not that the data assumptions of the statistical procedure are violated. In the absence of remedial action, these violations (e.g., heteroscedasticity, non-normality) can have serious consequences (e.g., standard error bias resulting in incorrect significance test ratios, compromised statistical power, invalid goodness of model fit evaluation). In their paper, Field and Wilcox provide an overview of the critical assumptions of the general linear model (GLM), discuss various ways of dealing with assumption violations, and demonstrate the implementation of these robust statistical methods in context of some GLM-based analyses that are commonly used in applied clinical research. The reader will see how straightforward it is to conduct robust statistical methods, which, in some cases can be as simple as overriding a software program default to request a robust estimator (e.g., see Example 8 in Field & Wilcox, 2017). Another important "take home" message from this paper is that, because test assumptions are almost never met in clinical research data, researchers should always address and mitigate the potential deleterious impacts of these violations. This might entail relying on robust statistical methods from the outset (e.g., use robust ML instead of ML); or when conventional statistical tests are used, conducting a sensitivity analysis to ensure the conclusions from the model are equivalent to those produced by a robust alternative.

Although very readable sourcebooks on these topics exist (e.g., Darlington & Hayes, 2017; Hayes, 2013; MacKinnon, 2008), there continues to be many misunderstandings and misapplications of mediation and moderation analysis in the applied clinical research literature. For instance, in mediation analysis, researchers often test the significance of the total effect of the focal independent variable on the outcome as a prerequisite to evaluating of the posited indirect effect (cf. Baron & Kenny, 1986). In moderation analysis, it is common for investigators to unnecessarily test the statistical significance of arbitrary simple slopes (e.g., ± *SD*) as a strategy for characterizing the nature of an interaction. Hayes and Rockwood (2017) present the state-of-the-art methods for estimating, interpreting, and reporting mediated and moderated effects in applied clinical data. During the course of this presentation, the authors highlight the errors in mediation and moderation analyses that are often seen in the research papers published in *Behaviour Research and Therapy*, and explain the best strategies for dealing with these issues (e.g., calculating the statistical and clinical significance of indirect effects; depicting the nature of moderated effects). Another extremely valuable aspect of Hayes and Rockwood's paper is the illustration of conditional process analysis using the free PROCESS macro for SAS and SPSS. As noted by the authors, hypotheses bearing on moderated mediation are quite relevant to clinical research (e.g., are the treatment processes that lead to symptom improvement more salient for some patients than for others?). However, analyses of this nature are rarely seen in the literature perhaps because most investigators are unaware of their availability and implementation. Hayes and Rockwood show that the methods of conditional process modeling are readily within the grasp of the applied mental health researcher.

The remaining three papers in this Special Section cover powerful analytic methods that have yet to receive widespread application in clinical research. Baldwin and Larson (2017) provide a user-friendly overview of the theoretical framework and practical issues of Bayesian models, and illustrate each step of this approach using a linear regression example. In the first paragraph of this introduction to the Special Section, we alluded to some of the ways that Bayesian estimation outperforms traditional statistical methods. However, despite these many advantages, clinical researchers have been slow to adopt these methods because Bayesian statistics: (a) challenge many of the assumptions of traditional statistics (e.g., emphasize the uncertainty of estimates rather than *p* values and null hypothesis significance testing); and (b) have historically required the use of specialized statistical software. As shown by Baldwin and Larson (2017), software availability is no longer a barrier as there are now many programs that render statistical modeling from the Bayesian perspective quite feasible (e.g., R, Mplus). In addition to basic analyses (e.g., regression models exemplified in Baldwin & Larson, 2017), the benefits of Bayesian estimation extend to more complex analyses such as multilevel models and structural equation models (SEM). For example, one of the key strengths of Bayesian analysis in fitting SEM measurement models is the ability, through the use of informative (subjective) priors, to replace parameters that are usually fixed to zero (e.g., cross-loadings, indicator error covariances) with values that are close to zero, but not exactly zero (cf. Brown, 2015). The ability to specify measurement models in this fashion provides a more reasonable evaluation of model fit (e.g., cross-loadings of trivial size will not result in poor model fit) and can mitigate bias in the parameter estimates (e.g., in models where cross-loadings and error covariances are fixed to zero, factor correlations tend to be over-estimated). Thus, the Bayesian framework offers applied researchers a more realistic and flexible approach to modeling which is better aligned with substantive theory (Muthén & Asparouhov, 2012). Nevertheless, the capability to build previous knowledge into an analytic model of any type has the potential downside of researcher misuse or abuse (e.g., within a single sample, conducting a frequentist statistical analysis to determine the informative priors to be specified in the subsequent Bayesian analysis, a malpractice that will unduly foster the power and precision of the model). Given the substantial impact this can have on the results of Bayesian analysis, the use of informative priors should be backed by strong substantive justification and sensitivity analysis (i.e., evaluation of the stability of the results across a plausible range of priors).

Although propensity score analysis has been around for over 30 years, applications of this method in the clinical psychology literature are sparse. This is unfortunate because studies examining the efficacy of psychological treatments are often based on observational or quasi-experimental designs. When such designs are used, estimates of treatment efficacy are biased because patient characteristics (e.g., pre-treatment symptom severity) are related to the receipt of treatment. Lee and Little (2017) demonstrate how unbiased treatment effect estimates can be obtained by adjusting for treatment selection bias through the use of propensity scores, which represent the likelihood a patient is assigned to or selects treatment given his/her scores on multiple, observed covariates at baseline. Using the freeware software platform R, Lee and Little provide a detailed, step-by-step illustration of the propensity score methodology to estimate treatment effects after the groups

(e.g., treated vs. untreated patients) have been equated on the range of covariates analogous to what would have occurred in a randomized experiment. Evidence from the medical field has shown that propensity score analysis produces results that are indeed comparable to randomized clinical trials (Olmos & Govindasamy, 2014). Accordingly, it is hoped that Lee and Little's demonstration of propensity score analysis will inspire clinical researchers to adopt this methodology to produce better treatment effect estimates in studies where random assignment is not possible.

In the final paper of the Special Section, Lubke and Luningham (2017) present the current, best practice methods for estimating mixture models in applied data sets (e.g., factor mixture models, growth mixture models). Mixture models are quantitative methods used to explore population heterogeneity by explicating previously unobserved homogeneous groups of individuals in a sample (i.e., latent classes). For example, in cross-sectional data, factor mixture modeling might lead to the discovery of subgroups of patients possessing distinct dimensional risk profiles (e.g., a disorder can be subtyped based on different pathways of pathogenesis). In time-series data, a growth mixture model might reveal different patterns of adjustment to traumatic stress or response to psychological treatment, and the predictors of these distinct symptom trajectories. Mixture models are far more sophisticated and powerful than traditional approaches to identifying subgroups in heterogeneous populations (e.g., cluster analysis). Although these methods have begun to find their way into the clinical literature, the majority of these efforts has entailed rudimentary forms of mixture modeling such as latent class (profile) analysis. However, the assumptions of these models are almost always unrealistic. Latent class analysis has the assumption of local independence which means, for example, that patients residing within a given class do not differ on symptom severity. When this assumption does not hold, too many classes are usually extracted. Moreover, a perusal of the clinical literature indicates that studies utilizing more sophisticated types of mixture models are frequently hindered by procedural errors including the use of less than ideal methods of model selection, measurement invariance evaluation, and model and class validation. Lubke and Luningham's thorough coverage of the theoretical and practical aspects of latent variable mixture modeling should markedly improve the use of these methods in applied research. To augment the growth mixture modeling example provided in this paper, readers are referred to Clark et al. (2013) for a tutorial on factor mixture models using cross-sectional, clinically relevant data.

In addition to presenting applied illustrations of the methodologies at a level of detail that the reader can adopt, the papers in this Special Section convey the minimum practice guidelines that should be met for submission to this journal. For example, case deletion or carrying the last observation forward should not be used to manage missing data in studies submitted for publication consideration to *Behaviour Research and Therapy*; researchers should always consider the potential impact of violations in statistical test assumptions (e.g., sensitivity analyses). For authors planning to submit a research report to this journal: if the study used one or more of these statistical methods (or a data issue such as missingness or non-normality applied to your study), you are strongly encouraged to cite the relevant papers from this Special Section to document adherence to these guidelines. Our original intent of the Special Section was to improve the statistical acumen of the readership of *Behaviour Research and Therapy*, but we believe these invited papers promise to have a strong impact on the field as a whole. For that, we wish to extend our sincere gratitude to the contributing authors for their willingness to share their expertise for this Special Section. It is our hope that readers will find these papers invaluable for informing their applied research endeavors.

References

Baldwin, S. A., & Larson, M. J. (2017). An introduction to using Bayesian linear regression with clinical data. *Behaviour Research and Therapy, 98*, 38–55.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford Press.

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling, 20*, 681–703.

Darlington, R. B., & Hayes, A. F. (2017). *Regression analysis and linear models: Concepts, applications, and implementation*. New York: Guilford Press.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy, 98*, 4–18.

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy* (in press).

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy, 98*, 19–37.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York: Guilford Press.

Lee, J., & Little, T. D. (2017). A practical guide to propensity score analysis for applied clinical research,. *Behaviour Research and Therapy, 98*, 56–70.

Lubke, G., & Luningham, J. (2017). Fitting latent variable mixture models. *Behaviour Research and Therapy, 98*, 71–82.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Taylor & Francis.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313–335.

Olmos, A., & Govindasamy, P. (2014, October). *Randomized experiments vs. propensity scores matching: A meta-analysis* (Paper presented at the meeting of the American Evaluation Association, Denver, CO).

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Timothy A. Brown*
Center for Anxiety and Related Disorders, Boston University, United States

Andy P. Field**
School of Psychology, University of Sussex, United Kingdom

* Corresponding author.

** Corresponding author.
E-mail address: tabrown@bu.edu (T.A. Brown).
E-mail address: Andy.Field@sussex.ac.uk (A.P. Field).

Available online 29 June 2017