REFERENCES

Abadie, A., Athey, S., Imbens, G. & Wooldridge, J. (2017). When should you adjust standard errors for clustering? National Bureau of Economic Research, Working Paper 24003. http://www.nber.org/papers/w24003

Academy of Medical Sciences (2015). Stratified, personalised or P4 medicine: A new direction for placing the patient at the centre of healthcare and health education. Downloaded at https://acmedsci.ac.uk/download?f= file&i=32644

Achen, C. (1987). *Interpreting and using regression*. SAGE.

Achen, C. (1990). What does "explained variance" explain?: Reply. *Political Analysis*, 2, 173-184.

Achen, C. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, 22, 327–339.

Acock, A. C. (2014). *A gentle introduction to Stata*. Stata Press.

Adedia, D., Adebanji, A., Okyere, E., Agyen, J. K. (2016). Leverages, outliers and the performance of robust regression estimators. *British Journal of Mathematics and Computer Science*, 15,1 -14.

Agresti, A. (1996). *An introduction to categorical data analysis*. Wiley.

Agresti, A. (2010). *Analysis of ordinal categorical data*. Wiley.

Agresti, A. & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*, 54, 280-288.

Agresti, A. & Tarantola, C. (2018). Simple ways to interpret effects in modeling ordinal categorical data. *Statistica Neerlandica,* 72, 1-14.

Aguinis, H., Werner, S., Abbott, L. et al. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, 13, 515–539.

Allen, M., Iliescu, D. & Greiff, S. (2022). Single item measures in psychological science: A call to action. *European Journal of Psychological Assessment*, 38, 1–5

Algina, J., Moulder, B. & Moser, B. (2002). Sample size requirements for accurate estimation of squared semipartial correlation coefficients. *Multivariate Behavioral Research*, 37, 37–57.

Algina, J. & Olejnik, S. (2003). Sample size tables for correlation analysis with applications in partial correlation and multiple regression analysis. *Multivariate Behavioral Research*, 38, 309–323.

Algina, J. & Swaminathan, H. (2011). Centering in two-level nested designs. In J. Hox and J. Roberts (Eds.) *Handbook of advanced multilevel analysis*. pp. 285–312. Taylor & Francis.

Allison, P. (2002). *Missing data*. SAGE.

Allison, P. D. (2005). Imputation of categorical variables with PROC MI. Paper presented at the 2005 SAS Users Group International conference. Retrieved January 30, 2009, from www2.sas.com/proceedings/sugi30/toc.html.

Allison, P. (2012a). Logistic regression for rare events. Downloaded April 2020 from https://statisticalhorizons.com/logistic-regression-for-rare-events.

Allison, P. (2012b). When can you safely ignore multicollinearity? Downloaded at https://statisticalhorizons.com/multicollinearity.

Allison, P. (2012c). Why maximum likelihood is better than multiple imputation. Downloaded from https://statisticalhorizons.com/ml-better-than-mi/

Allison, P. (2013). Why I don't trust the Hosmer-Lemeshow test for logistic regression. Downloaded from https://statisticalhorizons.com/hosmer-lemeshow.

Allison, P. (2014). Listwise deletion: It's NOT evil. Downloaded from https://statisticalhorizons.com/listwise-deletion-its-not-evil/.

Allison (2015). Imputation by predictive mean matching: Promise and peril. Statistical Horizons. Downloaded from https://statisticalhorizons.com/predictive-mean-matching/.

Allison, P. (2017). The peculiarities of missing at random. Downloaded from https://statisticalhorizons.com/missing-at-random/.

Allison, P. (2019). How many imputations do you need? Downloaded from https://statisticalhorizons.com/how-many-imputations/

Allison, P. (2020). Better predicted probabilities from linear probability models. Downloaded from https://statisticalhorizons.com/better-predicted-probabilities.

Altman, A.R. (1985). Comparability of randomised groups. *Journal of the Royal Statistical Society. Series D*, 34, 125–136.

Altman, D. G. & Andersen, P. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, 8, 771–783.

Amrhein, V., Greenland, S. & McShane B. (2019). Scientists rise up against statistical significance. *Nature*, 567, 305–307.

Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology*, 26, 1323-1333.

Ando, T. (2010). *Bayesian model selection and statistical modeling*. CRC Press.

Andridge, R. & Little, R. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78, 40–64.

Angrist, J., & Imbens, G.W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90, 431-442.

Angrist, J., Imbens, G. & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–472.

Angrist, J. & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Annemarie N. E., Winckers, J., Mackenbach, S., et al. (2015). Educational differences in the validity of self-reported physical activity, *BMC Public Health*, 15, 1299-1307.

Antonakis, J., Bastardoz, N. & Rönkkö, M. (2021). Ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, 24, 443-483.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly*, 21, 1086–1120.

Appiah, S. & Adedia, D. (2016). Leverage effects on robust regression estimators. *Journal of Mathematical Theory and Modeling*, 6, 53-73.

Ardito, R.B. & Rabellino, D. (2011).Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology*, 2, 1-11.

Arnold, A. & Kronmal, R. (2003). Multiple imputation of baseline data in the cardiovascular health study. *American Journal of Epidemiology*, 157, 74-84.

Asendorpf, J. B., Rindermann, H., Woodley, M. A., Stratford, J., Rabaglia, C., Marcus, G. and Lane, S. (2012). Bias due to controlling a collider: A potentially important issue for personality research. *European Journal of Personality*, 26, 391–413.

Ashworth, E., Panayiotou, M., Humphrey, N. & Hennessey, A. (2020). Game on—Complier average causal effect estimation reveals sleeper effects on academic attainment in a randomized trial of the good behavior game. *Prevention Science*, 21, 222–233.

Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural Equation Modeling, 12, 411-434.

Asparouhov, T. (2006). General multilevel modeling with sampling weights. *Communications in Statistics: Theory and Methods*, 35, 439-460.

Asparouhov, T. (2018). Blog post on Monday, March 05, 2018 at http://www.statmodel. com/discussion/messages/23/4554.html?1522263321

Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

Asparouhov, T. & Muthén, B. (2010a). Plausible values for latent variables using Mplus. Technical report. Los Angeles: Muthén & Muthén. www.statmodel.com.

Asparouhov, T. & Muthén, B. (2010b). Bayesian analysis using Mplus, version 4. Technical report. Los Angeles: Muthén and Muthén. www.statmodel.com.

Asparouhov, T., & Muthén, B. (2010c). Multiple imputation with Mplus (Technical report). Downloaded at www.statmodel.com.

Asparouhov, T. & Muthén, B. (2010d). Chi-square statistics with multiple imputation. Downloaded from http://statmodel.com/download/MI7.pdf

Asparouhov, T., & Muthén, B. (2010e). Resampling methods in Mplus for complex survey data. *Structural Equation Modeling*, 14, 535–569.

Asparouhov, T., & Muthén, B. (2012). Multiple group multilevel analysis (mplus web notes no. 16). Retrieved from http://statmodel.com/exam ples/webnotes/webnote16.pdf

Asparouhov, T., & Muthén, B. (2017). Prior-posterior predictive p-values. Mplus Web Notes: No. 22. Downloaded from https://www.statmodel.com/download/PPPP.pdf

Asparouhov, T. & Muthén, B. (2018a). SRMR in Mplus. Technical report downloaded from https://www.statmodel.com/download/SRMR2.pdf.

Asparouhov, T. & Muthén, B. (2018b): Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling*, 26, 119–142.

Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling*, 26, 119–142.

Asparouhov, T. Muthén, B. (2019b) Nesting and equivalence testing for structural equation models, *Structural Equation Modeling*, 2, 302-309.

Asparouhov, T. & Muthén, B. (2020). Bayesian estimation of single and multilevel models with latent variable interactions, *Structural Equation Modeling*, 28, 314–328.

Asparouhov, T. & Muthén, B. (2021). Multiple imputation with Mplus. Technical report. Los Angeles: Muthén and Muthén. www.statmodel.com

Asparouhov, T., Muthén, B. & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, 41, 1561-1577.

Assman, S.F., Pocock, S.J., Enos, L.E., & Kasten, L.E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet,* 355, 1064–1069.

Austin P. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine*, 32, 2837–2849.

Austin, P. & Stuart, E. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34, 3661-3679.

Azen, R. & Budescu, D. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8, 129–148.

Azen, R., & Budescu, D. V. (2006). Comparing predictors in multivariate regression models: An extension of dominance analysis. *Journal of Educational and Behavioral Statistics*, 31, 157–180

Azen, R. & Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, 34, 3-19-347.

Bacon, D. W., & Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58, 525-534.

Baiocchi, M., Cheng, J., & Small, D. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine*, 33, 2297–2340.

Baldwin, S., Bauer, D., Stice, E. & Rohde, P.(2011). Evaluating models for partially clustered designs. *Psychological Methods*, 16, 149–165.

Balzer, L., Zheng, W., van der Laan, M. & Petersen, M. (2019). A new approach to hierarchical data analysis: Targeted maximum likelihood estimation for the causal effect of a cluster-level exposure. *Statistical Methods in Medical Research*, 28, 1761-1780

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, 86, 948 –955

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.

Barr, D., Levy, R., Scheepers, C. & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences,* 42, 815–824.

Barrett, T (2019). MarginalMediation: v 0.7.2. https://tysonbarrett.com/MarginalMediation/

Barrett, T., Lockhart, G. & R. Cruz (2022). Marginal mediation analysis: A practical statistical framework for interpretable mediation effects. Downloaded at https://osf.io/preprints/psyarxiv/fgm8t

Bartlett, J. (2021): Reference-based multiple imputation — What is the right variance and how to estimate it. *Statistics in Biopharmaceutical Research*. On-line pre-print.

Bartlett, J. & Hughes, R. (2020). Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Statistical Methods in Medical Research*, 29, 3533–3546.

Barzi, F. & Woodward, M. (2004). Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology*, 160, 34-45.

Bauer, D., Preacher, K. & Gil, K. (2006) Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11, 142-163.

Bauer, D., Sterba, S. & Hallfors, D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research*, 42, 210-236.

Baumbartner, H. & Steenkamp, J. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.

Beaton, D. (2003). Simple as possible? Or too simple?: Possible limits to the universality of the one half standard deviation. *Medical Care*, 41, 593-596.

Beatty, P. & Willis, G. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71, 287–311.

Becker, L. & Coolidge, F. (1991). Interaction means in analysis of variance: A reply to Rosnow and Rosenthal, *Psychological Reports*, 68, 483-490.

Beckstead, J. (2014). On measurements and their quality. Paper 4: Verbal anchors and the number of response options in rating scales. *International Journal of Nursing Studies*, 51, 807–814.

Bell, B., Morgan, G., Schoeneberger, J., Kromrey, J. & Ferron, J. (2014). How low can you go? An investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodology*, 10, 1–11.

Bell, A., Fairbrother, M. & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality and Quantity*, 53, 1051-1074.

Bell, R. & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28, 169–181.

Bentler, P. & Yuan, K. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34, 181-197.

Beribisky, N., Mara, C. & Cribbie, R. (2020). An equivalence testing approach for evaluating substantial mediation. *Quantiative Methods for Psychology*. 16,424-441.

Beresniaket, A., Medina-Lara, A., Paul, J., De Wever, A., Praet, J., Tarricone, R., Torbica, A., Dupont, D., Lamure, M. & Duru, G. (2015). Validation of the underlying assumptions of the Quality-Adjusted Life-Years Outcome: Results from the ECHOUTCOME European Project. *PharmacoEconomics*, 33, 61–69.

Berger, V. (2005). Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biomedical Journal, 47*, 119-127.

Berk, R., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2013). Valid post-selection inference. *Annals of Statistics, 41,* 802–837.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, 2, 47–53.

Bernaards, C., Belin, T. & Schafer, J. 2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine*, 26, 1368–1382.

Biesanz, J., Falk, C. & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, 45, 661-701.

Black, N., Murphy, M., Lamping, D., McKee, M., Sanderson, C., Askham, J., et al. (1999). Consensus development methods: A review of best practice in creating clinical guidelines. *Journal of Health Service Research and Policy*, 4, 236–248.

Blair, E. & Blair, J. (2014). *Applied survey sampling*. Sage.

Blalock, H. M. (1964). *Causal inferences in non-experimental research*. University of North Carolina Press.

Bland, J.M. & Altman, D.G. (2011). Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials*, 12, 264.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, 61, 27–41.

Bleich, J., Kapelner, A., Jensen, S. & George, E. (2013). Variable selection inference for Bayesian Additive Regression Trees. Downloaded from https://www.researchgate.net/publication/258105822_Variable_Selection_Inference_for_Bayesian_Additive_Regression_Trees.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8, 225–246.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, 10, 689–709.

Bodner, T. (2008). What improves with increased missing data imputations. *Structural Equation Modeling*, 15, 651–675.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. *Sociological Methodology*, 25, 223-251.

Bollen, K. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37-72.

Bollen, K. (2019). Model implied instrumental variables (MIIVs): An alternative orientation to structural equation modeling, *Multivariate Behavioral Research*, 54, 31-46.

Bollen, K. (2020) When good loadings go bad: Robustness in factor analysis. *Structural Equation Modeling*, 27, 515-524.

Bollen, K. (2021). An introduction to model implied instrumental variables using two stage least squares (MIIV-2SLS) in structural equation models (SEMs). *Psychological Methods*, In press.

Bollen, K. & Barb, K. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review*, 46, 232–239.

Bollen, K. & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, 16, 265-284.

Bollen, K. & Diamantopoulos, A. (2015). In defense of causal-formative indicators: A minority report. http:// www.researchgate.net/profile/Kenneth_Bollen/publication/ 282046424_In_Defense_of_Causal-Formative_Indicators_A_Minority_Report/links/ 5669f0d608ae1a797e378258/In-Defense-of-Causal-Formative-Indicators-A-Minority-Report.pdf.

Bollen, K., Gates, K. & Fisher, Z. (2018) Robustness conditions for MIIV-2SLS when the latent variable or measurement model is structurally misspecified. *Structural Equation Modeling*, 25, 848-859.

Bollen, K., Harden, J., Ray, S. & Zavisca, J. (2014). BIC and alternative Bayesian Information Criteria in the selection of structural equation models. *Structural Equation Modeling*, 21, 1-19.

Bollen, K. & Hoyle, R. (2012). Latent variables in structural equation modeling. In R. Hoyle (Ed.). *Handbook of structural equation modeling*. Guilford

Bollen, K., Kolenikov, S. & Bauldry, S. (2014). Model-implied instrumental variable—Generalized method of moments (MIIV-GMM) estimators for latent variable models. *Psychometrika*, 79, 20–50.

Bollen, K. & Lennox, R (1991). Conventional wisdom in measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305-314.

Bollen, K. A., Lilly, A. G., & Luo, L. (2022). Selecting scaling indicators in structural equation models (sems). *Psychological Methods*. Published online.

Bollen, K. A. & Maydeu-Olivares, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, 72, 309-319.

Bollen, K. & Pearl, J. (2013). Eight myths about causality and structural equation modeling. In S.L. Morgan (Ed.), *Handbook of causal analysis for social research*, pp. 301-328, Springer 2013.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 115-140.

Bolzern, J.E., Mitchell, A. & Torgerson, D.J. (2019). Baseline testing in cluster randomised controlled trials: Should this be done? *BMC Medical Research Methodology*, 19, 106-115.

Bon, J. (2025). Using the tidytreatment package with BART. At https://cran.r-project.org/web/packages/tidytreatment/vignettes/use-tidytreatment-BART.html

Bonifay, W., Lane, S. & Reise, S. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5, 184–186.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, 50, 229–242

Boomsma, A. & Herzog, W. (2013). Correcting structural equation model t statistics and indexes under small-sample and/or large-model conditions Version 1.2, August 20, 2013. Downloaded from https://www.researchgate.net/publication/338544950_R_function_ swain_Correcting_structural_equation_model_fit_statistics_and_indexes_under_small-sample_andor_large-model_conditions_Version_12

Borgen, N., Haupt, A. & Wiborg, 0. (2022). A new framework for estimation of unconditional quantile treatment effects: The residualized quantile regression (RQR) model. SocArXiv, doi: 10.31235/osf.io/42gcb.

Borgers, N., de Leeuw, E. & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin of Sociological Methodology*, 66, 60-75.

Borkovec, T. D. & Whisman, M. A. (1996). Psychosocial treatment for generalised anxiety disorder. In M. Mavissakalian and R. Prien (Eds.) *Long-term treatment for the anxiety disorders*, pp. 171-199. American Psychiatric Press : Washington, DC.

Box, G. E. P. & Tiao, G. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley

Box, G. E. P. & Todwell, P. (1962). Transformation of the independent variables, *Technometrics*, 4, 531-550.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

Brandt, H., Umbach, N., Kelava, A., & Bollen, K. A. (2019). Comparing estimators for latent interaction models under structural and distributional misspecifications. *Psychological Methods*, 25, 321-345.

Braun, M., Converse, P. & Oswald, F. (2019). The accuracy of dominance analysis as a metric to assess relative importance: The joint impact of sampling error variance and measurement unreliability. *Journal of Applied Psychology*, 104, 593–602.

Brazier, J. & Tsuchiya, A. (2015). Improving cross-sector comparisons: Going beyond the health-related QALY. *Applied Health Economics and Health Policy*, 13, 557–565.

Breen, R., Karlson, K. & Holm, A. (2018). Understanding logits, probits, and other nonlinear probability models. *Annual Review of Sociology*, 44, 1–16.

Breiman, L., Friedman J.H., Olshen, R. A., Stone, C.J. (1984) *Classification and regression trees*. Brooks-Cole.

Brick, T. R., Koffer, R., Gerstorf, D., & Ram, N. (2018). Feature selection methods for optimal design of studies for developmental inquiry. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 73, 113–123.

Bring, J. (1994) How to standardize regression coefficients. *American Statistician*, 48, 209-213,

Bring, J. (1996). A geometric approach to compare variables in a regression model. *American Statistician*, 50, 57-62.

Bromaghin, J. (1993). Sample size determination for interval estimation of multinomial probabilities. *American Statistician*, 47, 203-206.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, Massachusetts: Harvard University Press.

Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44, 108–132.

Browne, M.W. & Cudek, R. (1993). Alternative ways of assessing model fit. pp. 136-162 in Bollen, K. and Long J.S (Eds.) *Testing structural equation models*. Newbury Park: . SAGE.

Browne, M., MacCallum, R., Kim, C., Andersen, B. and Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, 7, 403–421.

Brownson, R., Colditz, G. & Proctor, E. (2017). *Dissemination and implementation research in health: Translating science to practice*. Oxford University Press.

Bryant, F. & Satorra, A. (2012) Principles and practice of scaled difference chi-square testing, *Structural Equation Modeling*, 19, 372-398.

Buckland, S. T. (1984). Monte Carlo confidence intervals. *Biometrics*, 40, 811–817.

Budescu, D.V. (1993). Dominance analysis: A new approach to the problem of relative importance in multiple regression. *Psychological Bulletin*, 114, 542-551.

Budge, S. L., Adelson, J. L., & Howard, K. A. S. (2013). Anxiety and depression in transgender individuals: The roles of transition status, loss, social support, and coping. *Journal of Consulting and Clinical Psychology*, 81, 545–557.

Bunge, M. (1961). Causality, chance, and law. *American Scientist*, 69, 432–488.

Burnham, K. & Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33, 261-304.

Buzzoni, M. (2014). The agency theory of causality, anthropomorphism, and simultaneity. *International Studies in the Philosophy of Science*, 28, 375–395.

Cai, A., Tsay, R. & Chen, R. (2009). Variable selection in linear regression with many predictors. *Journal of Computational and Graphical Statistics*, 18, 573-591.

Cain, M. K., & Zhang, Z. (2019). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling*, 26, 39–50

Calinski, T. & Kageyama, S. (2003). *Block designs: A randomization approach: Volume II*. Springer.

Cameron, A. C., Gelbach, J. & Miller, D. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90, 414–327.

Cameron, A. C. & Miller, D. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50, 317–372.

Cameron A.C. & Trivedi, P.K. (1998). *Regression analysis of count data*. Cambridge University Press.

Cameron, A.C. & Trivedi, P.K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Cameron, A. & Trivedi, (2010). *Microeconometrics using Stata*. College Station, Texas: Stata Press.

Campbell, M., Julious, S. & Altman, D. (1995). Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *British Medical Journal*, 311, 1145-1148

Campbell, D.T. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton-Mifflin.

Candlish, J., Teare, M., Dimairo, M. et al. (2018). Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. *BMC Medical Research Methodology*, 18, 105.

Carlson, K. & Wu, J. (2012). The illusion of statistical control: Control variable practice in management research. *Organizational Research Methods*, 15, 413–435.

Cartwright, N. (2007). *Hunting causes and using them*. Cambridge University Press.

Carter, A., Schnepel, K. & Steigerwald, A. (2012). Cluster robust inference for heterogeneous cluster samples. Technical Report, University of California, Santa Barbara. Downloaded from http://economics.ucdavis.edu/events/papers/ Steigerwald56.pdf

Catalano, R., Berglund, L., Ryan, J., Lonczak, H. & Hawkins, J.D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Annals of the American Academy of Political and Social Science*, *591*, 98-124.

Cesana, B. M., Reina, G. & Marubini, E. (2001). Sample size for testing a proportion in clinical trials: A "two- step" procedure combining power and confidence interval expected width. *American Statistician*, 55, 288-292.

Chalmers, T., Celano, P. Sacks, H. & Smith, H. (1983). Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine, 309*, 1358-1361.

Cham, H., Reshetnyak, E., Rosenfeld, B. & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research*, 29, 52, 12–30.

Chambers, C., Feredoes, E., Muthukumaraswamy, S., & Etchells, P. (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, 1, 4-17.

Charlton, B. (1994). Understanding randomized controlled trials: Explanatory or pragmatic? *Family Practice*, *11*, 243-244.

Chatla, S. & Shmueli, G. (2013) Linear probability models (LPM) and big data: The good, the bad, and the ugly. Working Paper. Indian School of Business, Hyderabad. Downloaded from http://eprints.exchange.isb.edu/84/

Chen, H. (2014). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective*. Thousand Oaks, CA: . SAGE.

Chen, F., Curran, P., Bollen, K., Kirby, J. & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods and Research*, 36, 462–494.

Chen, G. & Tsurumi, H. (2010) Probit and logit model selection. *Communications in Statistics - Theory and Methods*, 40, 159-175.

Chen, H. (2015). *Practical program evaluation*. SAGE.

Chen, K., Martin, R. & Wooldridge, J. (2023).Another look at the linear probability model and nonlinear index models. BLS Working Papers. Downloaded at https://www.bls.gov/osmr/research-papers/2023/pdf/ec230090.pdf

Chernick, M. & LaBudde, R. (2011). *An introduction to bootstrap methods with applications to R*. New York: Wiley.

Cheung, Y. B. (2007). A modified least squares regression approach to the estimation of risk difference. *American Journal of Epidemiology*, 166 ,1337-1344.

Cheung, M. W. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, 41, 425-438.

Chipman, H., George, E. & McCulloch, R. (2010). BART: Bayesian additive regression trees. *Annual Applied Statistics*, 4,266–298.

Cho, S., Preacher, K. & Bottge, B. (2015). Detecting intervention effects in a cluster-randomized design using multilevel structural equation modeling for binary responses. *Applied Psychological Measurement*, 39, 627–642.

Chudik, A., Kapetanios, G., & Pesaran, H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica*, 84, 1479–1512.

Chondros, P., Ukoumunne, O., Gunn, J. & Carlin, J. (2021). When should matching be used in the design of cluster randomized trials? *Statistics in Medicine*, 40, 5765-5778.

Cinelli, C., Forney, A. & Pearl, J. (2019). A crash course in good and bad control. Downloaded from http://causality.cs.ucla.edu/blog/index.php/2019/08/14/a-crash-course-in-good-and-bad-control/

Clarke, G., Eubanks, D., Reid, E., Kelleher, C., O'Connor, E., DeBar, L., Lynch, F., Nunley, S., Gullion, C. (2005). Overcoming depression on the internet (ODIN)(2): A randomized trial of a self-help depression skills program With reminders. *Journal of Medical Internet Research*, 7, e16

Clarke, K. (2005). The phantom menace: Omitted variable bias in econometric research. *Conflict Management and Peace Science*, 22: 341-352.

Clarke, K. (2009). Return of the phantom menace: Omitted variable bias in political research. *Conflict Management and Peace Science*, 26, 46–66.

Clarke, K., Kenkel, B., & Rueda, M. (2018). Omitted variables, countervailing effects, and the possibility of overadjustment. *Political Science Research and Methods*, 6, 343-354.

Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66, 27–44.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114, 494-509.

Cliff, N. (1996). *Ordinal methods for the behavioral sciences*. Erlbaum.

Clifton, J. & Depaoli, S. (2015). The impact of prior distributions in multilevel structural equation modeling. Downloaded from https://www.researchgate.net/profile/James-Clifton/publication/305851321_The_Impact_of_Prior_Distributions_in_Multilevel_Structural_Equation_Modeling/links/57a39ea308aefe6167a7310d/The-Impact-of-Prior-Distributions-in-Multilevel-Structural-Equation-Modeling.pdf

Cochrane Group (2019) RoB 2: A revised Cochrane risk-of-bias tool for randomized trials. Downloaded from https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, 49, 997-1003.

Cohen, J. & Cohen, P. (1984). *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Erlbaum.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*, 558–577.

Cole, D. & Preacher, K. (2014).Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, 19, 300–315.

Cole, S., Chu, H. & Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: A primer. *American Journal of Epidemiology*, 179, 252–260.

Cole, S. & Hernán, M (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168, 656–664.

Collins, L.M. (2018). *Optimization of behavioral, biobehavioral, and biomedical interventions: The multiphase optimization strategy (MOST)*. New York: Springer.

Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6, 330-351.

Courvoisier, D., Combescure, C. Agoritsas, T., Gayet-Ageron, A. & Perneger, T. (2011a) Performance of logistic regression modeling: Beyond the number of events per variable: The role of data structure. *Journal of Clinical Epidemiology*, 64, 993-1000.

Courvoisier, D., Combescure, C. Agoritsas, T., Gayet-Ageron, A. & Perneger, T. (2011b) Performance of logistic regression modeling: Beyond the number of events per variable: The role of data structure. *Journal of Clinical Epidemiology*, 64, 1463.

Creswell, J., & Plano Clark, V. (2017). *Designing and conducting mixed methods research*. Thousand Oaks, CA: . SAGE.

Cro, S., Morris, T., Kenward, M. & Carpenter, JR. (2016). Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation. *The Stata Journal*, 16, 443–463.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. Wiley.

Cronbach, L. J., Nageswari, R., & Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *British Journal of Statistical Psychology*, 16, 137–163.

Cuesta-Albertos, J., Gordaliza, A. & Matran, C, (1997). Trimmed k-Means: An attempt to robustify quantizers. *Annals of Statistics*, 25, 553–576.

Cuijpers, P., Reijnders, M. & Huibers, M. (2019). The role of common factors in psychotherapy outcomes. *Annual Review of Clinical Psychology*, 15, 207-231.

Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, 29, 89-93.

Cumming, G., & Fidler, F. (2009). Confidence intervals: Better answers to better questions. *Journal of Psychology*, 217, 15–26

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* Routledge.

Cumming, G. (2013). The new statistics: Estimation for better research. Retrieved from www.thenewstatistics.com

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7–29.

Cumming, G. & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60, 170–180.

Dallal, G. (2012). The little handbook of statistical practice. Available at https://www.amazon.com/dp/B00847SM6A.

Daniel, R., De Stavola, B., Cousens, S. & Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics*, 71, 1-14.

Daniels, M. & Hogan, J. (2008). *Missing data in longitudinal studies*. Chapman and Hall.

Darlington, R. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161-182

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

Davidson, M. & Sharma, A. (1988). Parametric statistics and levels of measurement. *Psychological Bulletin*, 104,137-144.

Davidson, M. & Sharma, A. (1990). Parametric statistics and levels of measurement: Factorial designs and multiple regression. *Psychological Bulletin*, 107, 394-400.

Davidson, M. & Sharma, A. (1994). ANOVA and ANCOVA of pre- and post- test ordinal data. *Psychometrika*, 59, 593-600.

Dawid, A. (2024). Potential outcomes and decision-theoretic foundations for statistical causality: Response to Richardson and Robins. *Journal of Causal Inference*, 12, 20230058.

Dawid A. (2021). Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9, 39–77.

Dawiid, A. (2020). Causal inference isn't what you think it is. Downloaded from https://www.statslab.cam.ac.uk/~apd/APD%20LSHTM%202020handout.pdf

Dawid, A. (2007). Counterfactuals, hypotheticals and potential responses: A philosophical examination of statistical causality. In F. Russo and J. Williamson (Eds), *Causality and probability in the sciences*. College Publications, Texts In Philosophy Series 5, 503-32, London.

Dawid, A. (2002). Commentary: Counterfactuals: Help or hindrance?, *International Journal of Epidemiology*, 31, 429–430.

Dawid, A. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95, 407–424.

Dawid, A. (1999). Who needs counterfactuals? In Gammerman, A. (Ed.) *Causal models and intelligent data management*. Springer, Berlin, Heidelberg.

de Boor, C. (2001). *A practical guide to splines*. Springer.

de Boer, M., Waterlander, W., Kuijper, L., Steenhuis, I. & Twisk, J. (2015).Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 1-8.

Deke, J. (2014). Using the linear probability model to estimate impacts on binary outcomes in randomized controlled trials. HHS Office of Adolescent Health. Downloaded from https://www.hhs.gov/ash/oah/sites/default/files/ash/oah/oah-initiatives/assets/lpm-tabrief.pdf.

DeMaio, T.J. & Rothgeb, J.M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.). Answering questions: Methodology for determining cognitive and communicative processes in survey research. (pp. 155-176). Jossey-Bass.

Demirtas, H. & Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, 22, 2553-2575.

Depaoli, S. (2021). *Bayesian structural equation modeling*. Guilford.

Depaoli, S., & Clifton, J. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, 22, 327-351.

Derksen, S. & Keselman, H. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45, 265–282.

De Silva, A., Moreno-Betancur, M., De Livera, A., Lee, K. & Simpson, J. (2017). A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: A simulation study. *BMC Medical Research Methodology*, 17, 114–124.

Deshpande, S. (2025). flexBART: Flexible Bayesian regression trees with categorical predictors. Journal of Computational and Graphical Statistics. Downloaded at https://www.tandfonline.com/doi/full/10.1080/10618600.2024.2431072.

Diamantopoulos, A. (2008). Formative indicators: Introduction to the special issue. *Journal of Business Research*, 61, 1201–1202.

Diamantopoulos, A., Riefler, P. & Roth, K. (2008). Advancing formative measurement models. *Journal of Business Research*, 61, 1203–1218.

Didelez, V., Dawid, A. P., & Geneletti, S. (2006). Direct and indirect effects of sequential treatments. In Proceedings of the 22nd conference on uncertainty in artificial intelligence (pp. 138–146). Arlington, VA, USA: AUAI Press

Diehr, P., Martin, D., Koepsell, T. & Cheadle A. (1995). Breaking the matches in a paired t-test for community interventions when the number of pairs is small. *Statistics in Medicine*, 14, 1491-1504.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6, 274–290.

Diez Roux, A. (2002). A glossary of multilevel analysis. *Journal of Epidemiology and Community Health*, 56, 588-594.

Ding, P., Vanderweele, T. & Robins, J. (2017). Instrumental variables as bias amplifiers with general outcome and confounding. *Biometrika*, 104, 291–302.

Dishion, T., McCord, J. & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, 5, 755-766.

Dobson, A. (2008). An introduction to generalized linear models. New York: Chapman and Hall.

Domingue, B., Kanopka, K., Trejo, S., Rhemtulla, M., & Tucker-Drob, E. (2022). Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. *Psychological Methods*, no pagination provided.

Donner, A., Birkett, N. & Buck, C. (1981). Randomization by cluster. Sample size requirements and analysis. *American Journal of Epidemio*logy, 114, 906-914.

Donner A. & Klar N. (2000). *Design and analysis of cluster randomization trials in health research*. Arnold, 2000.

Donohew, L., Helm, D., Lawrence, P. & Shatzer, M. (1990). *Sensation seeking, marijuana use, and responses to prevention messages: Implications for public health campaigns*. Springer.

Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological Methods*, 20, 489-505.

DuGoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*, 49, 284–303.

DuMouchel, W.H. & Duncan, G. J. (1983). Using sampling survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, 78, 535-543.

Dunn, G., Maracy, M., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., et al. (2003). Estimating psychological treatment effects from a randomized controlled trial with both non-compliance and loss to follow-up. *British Journal of Psychiatry*, 183, 323–331.

Dunn, G., Maracy, M., & Tomenson, B. (2005). Estimating treatment effects from randomized controlled trials with noncompliance and loss to follow-up: the role of instrumental variables methods. *Statistical Methods in Medical Research*, 14, 369-395.

Efron, B. (1978). Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, 73, 113-121.

Efron, B., Hastie, T., Johnstone, I & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32, 407-499.

Eldridge S., Ukoumunne, O. & Carlin, J. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review*, 77, 378–394.

Elridge, S. & Kerry, S. (2012). *A practical guide to cluster randomised trials in health services research*. Wiley.

Elwert, F. & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40, 31-53.

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6, 352–370.

Enders, C. K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 313-342). Greenwich, CT: Information Age Publishing.

Enders, C. (2010*). Applied missing data analysis*. Guilford.

Enders, C. (2022). *Applied missing data analysis*. Second Edition. Guilford.

Enders, C., Baraldi, A. & Cham, H. (2014). Estimating interaction effects with incomplete predictor variables. *Psychological Methods*, 19, 39–55.

Enders, C., Du, H. & Keller, B. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods*, 25, 88–112.

Enders, C. & Gottschall, A. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling*, 18, 35–54.

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121–138.

Everitt, B., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster analysis*. Wiley.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517.

Fagan, A. & Lindsey, A. (2014). Gender differences in the effectiveness of delinquency prevention programs: What can be learned from experimental research? *Criminal Justice and Behavior*, 41, 1057-1078.

Fairchild, A. J., MacKinnon, D. P., Toborga, M. P., & Taylor, A. B. (2009). R-squared effect-size measures for mediation analysis. *Behavior Research Methods*, 41, 486–498.

Fairchild, A., Yin, Y., Baraldi, A., Olvera Astivia, L. & Shi, D. (2024). Many nonnormalities, one simulation: Do different data generation algorithms affect study results? *Behavior Research Methods*. Published online.

Falk, C.F. & Biesanz, J. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling*, 22, 24-38.

Falkenström, F., Park, S., & McIntosh, C. N. (2021). Using copulas to enable causal inference from non-experimental data: Tutorial and simulation studies. *Psychological Methods*, In press.

Fan, J. & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society Series B*, 70, 849–911.

Farlow, M., Potkin, S., Koumaras, K., Veach, J. & Mirski, D. (2003). Analysis of outcome in retrieved dropout patients in a rivastigmine vs placebo, 26-week, alzheimer disease trial. *Archives of Neurolology*, 60, 843-848.

Farnam, M. (2011). Comment on David Giles blog at ttps://davegiles.blogspot.com/2011/09/econometrics-and-one-way-streets.html

Faris, P.D., Ghali, W.A.,, Brant, R., Norris, C.M., Galbraith, P.D. and Knudtson, M.L. (2002). Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *Journal of Clinical Epidemiology*, 55, 184-191.

Feaster D., Mikulich-Gilbertson, S. & Brincks, A. (2011). Modeling site effects in the design and analysis of multi-site trials. *American Journal of Drug and Alcohol Abuse*, 37, 383-391.

Feinberg, S. E. (1989). Modeling considerations: Discussion from a modeling perspective. In Kaspryzk, D., Duncan, G., Kalton, G., & Singh, M.P. (Eds.) *Panel surveys*. New York: Wiley.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75, 259–276.

Firpo, S., Fortin, N. & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77, 953-973.

Firpo, S., Fortin, N. & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, 6, 28.

Firpo, S. & Pinto, C. (2016). Identification and estimation of distributional impacts of interventions using changes in inequality measures. *Journal of Applied Econometrics*, 31, 457-486.

Flay B. R. (2002). Positive youth development requires comprehensive health promotion programs. *American Journal of Health Behavior*, *26*, 407–424.

Flay B. R., Allred, C.G., & Ordway, N. (2001). Effects of positive action program on achievement and discipline: Two matched-control comparisons. *Prevention Science*, *2*, 71-89.

Fomby, T. Johnson, S. & Hill, C. (1984). *Applied econometric methods*. Springer.

Food ad Drug Administration (2023). Adjusting for covariates in randomized clinical trials for drugs and biological products: Guidance for industry. Downloaded at https://www.fda.gov/media/148910/download

Fox, J. (1991). *Regression diagnostics: An introduction*. Thousand Oaks, CA: SAGE.

Frangakis, C. & Rubin, D. (1999). Addressing complications of intention-to-treat analysis in the presence of allor-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86, 365–379.

Frangakis, C., Rubin, D. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21–29.

Frank, P. (1961). *Modern science and its philosophy*. Collier Books.

Fried, E. & Flake, J. (2018). Measurement matters. APS Observer. *The American Psychological Association Observer*, 10.

Friedman J. (1991). Multivariate adaptive regression splines (with discussion and a rejoinder by the author). *Annals of Statistics*, 19, 1–67.

Friedman J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189–1232.

Friedman, L., Furberg, C., DeMets, D., Reboussin, D. & Granger, C. (2015). *Fundamentals of clinical trials*. Springer.

Fritz, H., Garcıa-Escudero, L. & Mayo-Iscar, A. (2012). tclust: An R package for a trimming approach to cluster analysis. *Journal of Statistical Software*, 47, 1-26.

Fritz, H., García-Escudero, L. & Mayo-Iscar, A. (2013). A fast algorithm for robust constrained clustering. *Computational Statistics and Data Analysis*, 61, 124–136.

Fuller, W.A. (1984). Least squares and related analyses for complex survey designs. *Survey Methodology*, 10, 97-118.

Fullerton, A. (2009). A conceptual framework for ordered logistic regression models. *Sociological Methods and Research*, 38, 306-347.

Fullerton, A. & Xu, J. (2020). *Ordered regression models*. Chapman & Hall.

Funk, M., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, 173, 761–767.

Furr, M. (2017). *Psychometrics: An introduction*. SAGE.

Furukawa, T.A. & Leucht, S. (2011). How to obtain NNT from Cohen's d: Comparison of two Methods. *PLoS ONE*, 6, e19070.

Gaffert, P., Meinfelder, F., Bosch V. (2015). Towards an MI-proper Predictive Mean Matching, Discussion Paper. Downloaded from https://www.uni-bamberg.de/ fileadmin/uni/fakultaeten/sowi_lehrstuehle/statistik/Personen/Dateien_Florian/properPMM.pdf

Galimard, J., Chevret, S., Curis, E. et al. (2018). Heckman imputation models for binary or continuous MNAR outcomes and MAR predictors. *BMC Medical Research Methodology*, 18, 90-103.

Galimard J., Chevret, S., Protopopescu, C. & Resche-Rigon, M. (2016). A multiple imputation approach for MNAR mechanisms compatible with Heckman's model. *Statistics in Medicine*, 35, 2907–2920.

García-Escudero, L., Gordaliza, A., Matran, C., & Mayo-Iscar, A. (2008). A general trimming approach to robust cluster analysis. *Annals of Statistics*, 36, 1324–1345.

García-Escudero, L., Gordaliza, A., Matrán, C. et al. (2011). Exploring the number of groups in robust model-based clustering. *Statistics and Computing*, 21, 585–599.

Garfield, S. (1999). Brief psychotherapy: the role of common and specific factors. *Clinical Psychology & Psychotherapy*, 4, 122-132.

Garrido, M., Dowd, B., Hebert, P. & Maciejewski, M. (2016). Understanding treatment effect terminology in pain and symptom management research. *Journal of Pain and Symptom Management*, 52, 446–452.

Garson, D. (2014). *Ordinal regression*. Statistical Associates "Blue Book" Series

Gatsonis, C. and Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, 106, 516-524.

Gelman, A. (2017). Causal inference using Bayesian additive regression trees: Some questions and answers. Downloaded from https://statmodeling.stat.columbia.edu /2017/05/18/causal-inference-using-bayesian-additive-regression-trees-questions/

Gelman, A., Carlin, J. B., & Stern, H. S. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

Gelman, A., Hill, J. & Vitari, A. (2021). *Regression and other stories*. Cambridge University Press.

Gelman, A., Hill, J. & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5, 189–211.

Geweke J. (1992). Evaluating theaAccuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernado, J. Berger, A. Dawid & A. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–193. Oxford University Press.

Gerhard, C., Büchner, R. D., Klein, A. G., & Schermelleh-Engel, K. (2017). A fit index to assess model fit and detect omitted terms in nonlinear SEM. *Structural Equation Modeling*, 24, 414–427.

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 249-264.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78.

Gilliland, D., & Melfi, V. (2010). A note on confidence interval estimation and margin of error. *Journal of Statistics Education*, 18, 1–8.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. SAGE.

Glymour, M., Weuve, J., Berkman, L., Kawachi, I. & Robins, J. (2005). When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *American Journal of Epidemiology*, 162, 267–278.

Goldberg, L. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research*, 72, 7-27.

Goldsmith, J., Kitago, T., de la Garza, A. et al. (2021). The proportional recovery rule redux: Arguments for its biological and predictive relevance. *bioRxiv*, 05.20.445022

Gomila, R. (2020). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, 150, 700-709.

Gonzalez, O., & MacKinnon, D. P. (2020). The measurement of the mediator and its influence on statistical mediation conclusions. *Psychological Methods*. Published online at doi: 10.1037/met0000263.

Gonzalez, O. & Valente, M. (2023) Accommodating a latent XM interaction in statistical mediation analysis. *Multivariate Behavioral Research*, 58, 659-674.

Gordon, J. (2017). An experimental therapeutic approach to psychosocial interventions. Downloaded from www.nimh.nih.gov/about/director/messages/2017/an-experimental-therapeutic-approach-to-psychosocial-interventions.shtml

Grace, J.B. (2020) A 'weight of evidence' approach to evaluating structural equation models. *One Ecosystem*, 5, e50452.

Grace, J. & Bollen, K.A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistics*, 15, 191-213.

Graham J., Olchowski, A. & Gilreath, T.D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention. Science*, 8, 206–213.

Graham, J. & Schafer, J. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.) *Statistical strategies for small sample research.* (pp. 1-29). SAGE

Green, S.B., Akey, T., Fleming, K., Hershberger, and Marquis, J. (1997). Effect of the number of scale points on chi square fit indices in confirmatory factor analysis. *Structural Equation Modeling*, 4, 108-120.

Greene, W. H. (2012). *Econometric analysis*. Prentice Hall.

Greene, W. H. (2017) *Econometric analysis*. Pearson.

Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine*, 13, 1665-1677.

Greenland S. (2002). Causality theory for policy uses of epidemiologic measures. In: Murray, C., Salomon, J., Mathers, C. et al. (Eds.). *Summary measures of population health*. Harvard University Press. *Medicine*, 13, 1665-1677.

Greenland, S. (2003). Quantifying biases in causal models: Classical confounding vs. collider-stratification bias. *Epidemiology*, 14, 300-306.

Greenland, S., Maclure, M., Schlesselman, J. J., Poole, C., & Morgenstern, H. (1991). Standardized regression coefficients: A further critique and review of some alternatives. *Epidemiology*, 2, 387–392.

Greenland, S. & Pearce, N. (2015). Statistical foundations for model-based adjustments. *Annual Review of Public Health*, 36, 89–108.

Greenland, S., Robins, J., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science, 14*, 29–46.

Greenland, S., Schlesselman, J. J., & Criqui, M. H. (1986). The fallacy of employing standardized regression-coefficients and correlations as measures of effect. *American Journal of Epidemiology*, 123, 203–208.

Greifer, N. & Stuart, E. (2021). Choosing the estimand when matching or weighting in observational studies. Downlioded from https://arxiv.org/abs/2106.10577.

Griffiths, D. A., & Miller, A. J. (1973). Hyperbolic regression-a model based on two-phase piecewise linear regression with a smooth transition between regimes. *Communications in Statistics*, 2, 561-569.

Grömping, U. (2006). Relative importance for linear regression in R: The Package relaimpo. *Journal of Statistical Software*, 17, 1-22.

Grömping, U. (2007a). Estimators of relative importance in linear regression based on variance decomposition. *American Statistician*, 61, 139-147.

Grömping, U. (2007b). Response to comment by Scott Menard, re: Estimators of relative importance in linear regression based on variance decomposition. *American Statistician*, 61, 280-284.

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *American Statistician*, 63, 308-319.

Grömping, U. (2015). Variable importance in regression models. *WIREs Computational Statistics*, 7, 137-152.

Gruber, S. & van der Laan, M. (2012). tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software*, 51, 1-35.

Grund, S., Lüdtke, O & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, 21, 111–149.

Guo, S., Fraser, M. & Chen, Q. (2020). Propensity score analysis: Recent debate and discussion. *Journal of the Society for Social Work and Research*, 11, 2334-2315.

Haggerty, R. J., & Mrazek, P. J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Committee on Prevention of Mental Disorders, Institute of Medicine. National Academies Press.

Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis, 15*, 965–997.

Hair, J., & Sarstedt, M., Ringle, C. & Gudergan, S. (2018). *Advanced issues in partial least squares structural equation modeling*. SAGE.

Halkidi, M. & Vazirgiannis, M. (2001). Clustering validity assessment: Finding the optimal partitioning of a data set. Downloaded at https://www.researchgate.net/ publication/3940215_Clustering_Validity_Assessment_Finding_the_optimal_partitioning _of_a_data_set

Halkidi, M. & Vazirgiannis, M. (2008). A density-based cluster validity approach using multi-representatives. *Pattern Recognition Letters*, 29, 773-786.

Halpern, J. (2015). A modification of the Halpern-Pearl definition of causality. Cornell University, Computer Science Department, Ithaca, NY. Dowloaded from https://arxiv.org/abs/1505.00162

Hanmer, M. & Ozan, K. (2013). Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models. *American Journal of Political Science*, 57, 263–277.

Hao, L. & Naiman, D. (2007). *Quantile regression*. SAGE.

Hardt, J., Herke, M. & Leonhart, R. (2012). Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. *BMC Medical Research Methodology*, 12, 184.

Harring, J. (2014). A spline regression model for latent variables. *Educational and Psychological Measurement*, 74, 197–213

Harrell, F (2021). Avoiding one-number summaries of treatment effects for RCTs with binary outcomes. Downloaded from https://www.fharrell.com/post/rdist/#disqus_thread.

Harwell, M. (2020). The importance of Type I error rates when studying bias in Monte Carlo studies in statistics. *Journal of Modern Applied Statistical Methods*, 18, eP3295

Hasselblad, V., & Hedges, L. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 117, 167–178.

Haughton, D. M. (1988). On the choice of a model to fit data from an exponential family. *The Annals of Statistics*, 16, 342–355.

Haughton, D. M., Oud, J. H. L. and Jansen, R. A. R. G. (1997). Information and other criteria in structural equation model selection. *Communications in Statistics, Part B - Simulation and Computation*, 26, 1477–1516.

Hayduk, L. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, 74, 905–926.

Hayduk, L., Cummings, G. G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two three – Testing the theory in structural equation models! *Personality and Individual Differences*, 42, 841-50.

Hayduk, L., Cummings, G., Stratkotter, R., Nimmo, M., Grygoryev, K., Dosman, D., Gillespie, M., Pazderka-Robinson, H. & Boadu, K. (2009). Pearl's d-separation: One more step into causal thinking. *Structural Equation Modeling*, 10, 289-311.

Hayduk, L & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators, in structural equation models? *BMC Medical Research Methods*, 12, 1–17.

Hayduk, L., Pazderka, H., Cummings, G. & Boadu, K. (2007). The weird world, and equally weird measurement models: Reactive indicators and the validity revolution, *Structural Equation Modeling*, 14, 280-310.

Hays, W. L. (1994). *Statistics* (5th ed.). Harcourt Brace

Hayes, A. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Second edition. Guilford.

Hayes, A. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Third edition. Guilford.

Hayes, A. F., & Coutts, J. J.* (2020). Use omega rather than Cronbach's alpha for estimating reliability. But.. …". *Communication Methods and Measures*, 14, 1-24.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24, 1918-1927.

Heckman J. (1979). Sample selection bias as a specification error. *Econometrica*. 47,153–161.

Hedges, L. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, 32, 151–179.

Hedges, L. & Citkowicz, M. (2015). Estimating effect size when there is clustering in one treatment group. *Behavior Research Methods*, 47, 1295–1308.

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality and Quantity*, 43, 59–74.

Hennig, C., Meila, M., Murtagh, F. & Rocci, R. (2015). *Handbook of cluster analysis*. Chapman & Hall.

Henseler, J. (2020). *Composite-based structural equation modeling: Analyzing latent and emergent variables*. Guilford.

Hernán, M. (2005). Invited commentary: Hypothetical interventions to define causal effects - afterthought or prerequisite? *American Journal of Epidemiology*, 162, 618–620.

Hernán, M. & Robins J. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, 17, 360-372.

Hernán, M. & Robins, J. (2017). Per-protocol analyses of pragmatic trials. *New England Journal of Medicine*, 377, 1391–1398.

Hernán, M. & Robins, J. (2020). *Causal inference: What if*. Chapman & Hall/CRC

Heeringa, S., West, B. & Berglund, P. (2010). *Applied survey data analysis*. Chapman and Hall/CRC.

Heisig, J. & Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review*, 35, 258-279.

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling*, 16, 1–27.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling*, 14, 361–390.

Hesser, H. (2020). Estimating causal effects of internet interventions in the context of nonadherence. *Internet Interventions*, 21,e1-e10.

Hesser, H., Hedman, E., Lindfors, P., Andersson, E. & Ljótsson, B. (2017). The specific effect of systematic exposure in irritable bowel syndrome: Complier average causal effect analysis using growth mixture modeling. *Psychological Medicine*, 47, 2653–2662.

Hesterberg, T. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *American Statistician*, 69, 371–386.

Hill, j. & Su, Y. (2013). Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *Annals of Applied Statistics*, 7, 1386–1420.

Hill, J., Linero, A. & Murray, J. (2020). Bayesian additive regression trees: A review and look forward. *Annual Review of Statistics and Its Application*, 7, 251-278.

Hoffman, L. & Templin, J. (2011). Systematically varying effects in multilevel models: Permissible or problematic? Presented in Norman, Oklahoma at the Annual Meeting of the Society of Multivariate Experimental Psychology. Downloaded at www.lesahoffman.com/Workshops/SMEP_2011_Hoffman_&_Templin.pdf

Hoffman L. & Walters, R. (2022). Catching up on multilevel modeling. *Annual Review of Psychology*, 73, 659-689.

Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57, 116-131.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.

Holm, A., Ejrnæs, M. & Karlson, K. (2015). Comparing linear probability model coefficients across groups. *Quality and Quantity*, 49, 1823-34.

Holmes, D. (2013). Report triggers quibbles over QALYs, a staple of health metrics. *Nature Medicine*, 19, 248.

Horrace, W. C., & Oaxaca, R. L. (2003). New wine in old bottles: A sequential estimation technique for the LPM. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=383102

Horrace, W. C. & Oaxaca, R. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90, 321-327.

Hosmer, D.W., Hosmer, T., Lemeshow, S., le Cessie, S. & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, 16, 965–980.

Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *American Statistician*, 57, 229–232.

Hox, J. & Maas, C. (2001). The accuracy of multilevel structural equation modeling with pseudo balanced groups and small samples. *Structural Equation Modeling*, 8, 157–174.

Hox, J., Maas, C. & Brinkhuis, M. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, 64, 157-170.

Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 131-137.

Huang, F. (2016) Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education*, 84, 175-196.

Huang, F. (2018). Multilevel modeling myths. *School Psychology Quarterly*, 33, 492–499.

Huang, F. (2019) Alternatives to logistic regression models in experimental studies. *Journal of Experimental Education*. Downloaded from https://www.tandfonline.com/doi/abs/10.1080/00220973.2019.1699769.

Huang, F. (2022). Analyzing cross-sectionally clustered data using generalized estimating equations. *Journal of Educational and Behavioral Statistics*, 47, 101-125.

Huang, F. & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavioral Research Methods*, 54, 1181-1199.

Huang, F., Zhang, B. & Li, X. (2023) Using robust standard errors for the analysis of binary outcomes with a small number of clusters. *Journal of Research on Educational Effectiveness*, 16, 213-245.

Huang, P. (2018). A penalized likelihood method for multi-group structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, 71, 499-522.

Huang, P., Che, H. & Weng, L. (2017). A penalized likelihood method for sturctural equation modeling. *Psychometrika*, 82, 329-354.

Huang, Y. (2018). Joint significance tests for mediation effects of socioeconomic adversity on adiposity via epigenetics. *Annals of Applied Statistics*, 12, 1535-1557.

Huque, M., Carlin, J., Simpson, J. et al. (2019). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology*, 18, 168.

Hume, D. (1975). An enquiry concerning human understanding (3rd ed.) (P. N. Nidditch, Ed.). Oxford, UK: Clarendon Press. (Original work published 1777).

Huo, Y. & Budescu, D. (2009). An extension of dominance analysis to canonical correlation analysis. *Multivariate Behavioral Research*, 44, 688-709.

Iacus, S.M., Porro, G. (2007) Missing data imputation, matching and other applications of random recursive partitioning. *Computational Statistics and Data Analysis*, 52, 773-789.

Iacus, S.M., Porro, G. (2009) Random Recursive Partitioning: a matching method for the estimation of the average treatment effect. *Journal of Applied Econometrics*, 24, 163-185.

IBM Support (2015). Clustering binary data with k-means (should be avoided). Retrieved on August 30, 2015 from http://www-01.ibm.com/support/docview.wss?uid= swg21477401.

Imai, K., Keele, L. & Tingley D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15, 309-334.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2010). Causal mediation analysis using R. In H. D. Vinod (Ed.), *Advances in social science research using R, Lecture notes in statistics*. (pp. 129–154). Springer.

Imai, K., Keele, L., Tingley, D., Yamamoto, T. (2010). Advances in social science research using R. In: Vinod, H.D. (Ed.) *Causal mediation analysis using R*, pp. 129–154. Springer.

Imai, K., King, G. & Stuart, E. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society - Series A (Statistics in Society)*, 171, 481–502.

Imbens, G. & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98, 701–712.

Imbens, G. & Rubin, D. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, 65, 555–574.

Imbens, G. & Rubin, D. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25, 305-327.

Insel, T. (2012). Post by former NIMH Director Thomas Insel: Experimental medicine. Downloaded from www.nimh.nih.gov/about/directors/thomas-insel/blog/2012/ experimental-medicine.shtml

Insel, T. (2013). NIMH's new focus in clinical trials. Downloaded from www.nimh.nih.

gov/funding/grant-writing-and-application-process/concept-clearances/2013/nimhs-new-focus-in-clinical-trials.shtml

International Council for Harmonisation (2020). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. SAGE.

Jaccard, J. (2022). Methodological issues in the study of implicit attitudes. In Krosnick, J., Start, T. & Scott, A. (Eds.). *The Cambridge handbook of implicit bias and racism*. Cambridge Press.

Jaccard, J. & Bo, A. (2019). Making outcome measures more useful in clinical trials: Addressing the problem of arbitrary metrics. Silver School of Social Work, New York University, New York, New York.

Jaccard, J., & Brinberg, M. (2021). Monte Carlo simulations using extant data to mimic populations: Applications to the modified linear probability model and logistic regression. *Psychological Methods*, 26, 450–465.

Jaccard, J. & Jacoby (2020). *Theory construction and model building skills: A practical guide for social scientists*. Guilford.

Jaccard, J., Weber, J. & Lundmark, J. (1975). A multitrait-multimethod analysis of four attitude assessment procedures. *Journal of Experimental Social Psych*ology, 11, 149-154.

Jacobson, N. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 121-129.

Jacobucci, R. (2017). regsem: Regularized structural equation modeling. Downloaded from https://arxiv.org/pdf/1703.08489.pdf.

Jacobucci, R., Grimm, K. & McArdle, J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, 23, 555–566.

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2018). Variable selection in structural equation models with regularized MIMIC models. PsyArXiv, 1–40.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer.

James, L. R., & Brett, J. M. (1984). Mediators, moderators and tests for mediation. *Journal of Applied Psychology*, 69, 307-321.

Jeffreys H. (1961) *Theory of probability*. Oxford University Press.

Jia, F (2016). Methods for handling missing non-normal data in structural equation modeling. Doctoral dissertation in the Department of Psychology and the Graduate Faculty of the University of Kansas, Lawrence, Kansas.

Jia, F. & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods*, 51, 2337–2355.

Jiroutek, M., Muller, K., Kupper, L. & Stewart, P. (2003). A new method for choosing sample size for confidence interval-based inferences. *Biometrics*, 59, 580-590.

Jo, B. (2002). Estimating intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, 27, 385–430.

Johnson, J. (2000). A heuristic method for estimating the relative weight of regressors in multiple regression. *Multivariate Behavioral Research*, 35, 1-19.

Johnson, J. & LeBreton, J. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238-257.

Johnston, L., Titov, N., Andrews, G., Spence, J. & Dear, B. (2011) A RCT of a transdiagnostic internet-delivered treatment for three anxiety disorders: Examination of support roles and disorder-specific outcomes. *PLoS ONE*, 6,e28079.

Jones, B. & Kenward, M. (2014). *Design and analysis of cross-over trials*. Chapman & Hall.

Jo, B. (2002a). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, 27, 385-409.

Jo, B. (2002b). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, 7, 178–193.

Jo, B., Asparouhov, T. & Muthén, B. (2008). Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine*, 27, 5565-5577.

Jo, B., Asparouhov, T., Muthén, B., Ialongo, N. & Brown, C. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods*, 13, 1–18.

Jo, B., Ginexi, E. & Ialongo, N., (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science*, 11, 384–396.

Jo, B., Ginexi, E. M., & Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science*, 11, 384–396.

Judd, C.M., McClelland, G.H. & Ryan, C.S. (2009). *Data analysis: A model comparison approach*. Routledge.

Judge, T. & Cable, D. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology*, 89, 428–441.

Judkins, D. & Porter, K. (2015). Robustness of ordinary least squares in randomized clinical trials. *Statistics and Medicine*, 35, 1763-1773.

Kahan, B., Forbes, G., Ali, Y., et al. (2016). Increased risk of type I errors in cluster randomized trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. *Trials*, 1, 438.

Kang, J. & Schafer, J. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22, 523–539.

Kaplan, D. (2008). *Structural equation modeling*. Newbury Park: SAGE.

Karlson, K. B. (2015). Another look at the method of Y-standardization in logit and probit models. Journal of Mathematical Sociology, 39, 29-38.

Krasikova, D., Le, H. & Bachura, E. (2018). Toward customer centric organizational science: A common language effect size indicator for multiple regression and regression with higher order terms. *Journal of Applied Psychology*, 103, 659-675.

Kass, R E., and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90, 928-934.

Kaufman, L. & Rousseeuw, P.J. (1990) *Finding groups in data: An introduction to cluster analysis*. Wiley.

Kazdin, A. E. (2003). Clinical significance: Measuring whether interventions make a difference. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 691–710). American Psychological Association.

Keefe, R., Kraemer, H., Epstein, R., Frank, E., Haynes, G., Laughren, T., ... (2013). Defining a clinically meaningful effect for the design and interpretation of randomized controlled trials. *Innovations in Clinical Neuroscience*, 10, 4S–19S.

Keele, L. & Park, D. (2006). Difficult choices: An evaluation of heterogenous choice models. Downloaded from https://www3.nd.edu/~rwilliam/oglm/ljk-021706.pdf

Kellogg Foundation. (2004). *Logic model development guide*. Battle Creek, MI: Kellogg.

Keller, B. T., & Enders, C. K. (2021). *Blimp user's guide* (Version 3). Retrieved from www.appliedmissingdata.com/multilevel-imputation.html

Kelly, J., Hoeppner, B., Stout, R. & Pagano, M. (2011). Determining the relative importance of the mechanisms of behavior change within Alcoholics Anonymous: A multiple mediator analysis. *Addiction*, 107, 289–299

Kelly, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software*, 20, 1-24.

Kelley, K. (2007b). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software*, 20, 1-24.

Kelley, K. & Maxwell, S. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321.

Kelley, K. & Maxwell, S. (2008). Sample size planning with applications to multiple regression: Power and accuracy for omnibus and targeted effects. In P. Alasuutari, L. Bickman, and J. Brannen (Eds.) *The Sage handbook of social research methods*. pp. 166–92. London: Sage

Kelley, K., & Pomprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, 21, 69-92.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17, 137-152.

Kendall, P. C., Cummings, C. M., Villabo, M. A., Narayanan, M. K., Treadwell, K., Birmaher, B.,… (2016). Mediators of change in the Child/Adolescent Anxiety Multimodal Treatment Study. *Journal of Consulting and Clinical Psychology*, 84, 1–14.

Kenny, D. & Campbell, D. (1999). *A primer on regression artifacts*. Guilford.

Kenny, D. & Judd, C. (2014). Power anomalies in testing mediation. *Psychological Science*, 25, 334-339.

Kenny, D.A. (2018   ). SEM. Web page at http://davidakenny.net/cm/causalm.htm

Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods and Research*, 44, 488-507.

Kenward, M. & Molenberghs, G. (2009). Last observation carried forward: A crystal ball? *Journal of Biopharmaceutical Statistics*, 9, 872–888.

Kenward, M., & Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983–997.

Kenward, M. G. & Roger, J. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, 53, 2583–2595.

Keselman, H., Algina, J., Lix, L., Wilcox, R. & Deering, K. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, 13, 110–129.

Keselman, H.J., Cribbie, R. & Holland, B. (1999). The pairwise multiple comparison multiplicity problem: An alternative approach to familywise and comparisonwise Type I error control. *Psychological Methods*, 4, 58-69.

Keselman, H. J., Othman, A. R. & Wilcox, R. (2013). Preliminary testing for normality: Is this a good practice? *Journal of Modern and Applied Statistical Methods*, 12, 2-19.

Kiefer, C., Woud, M., Blackwell, S. & Mayer, A. (2023). Average treatment effects on binary outcomes with stochastic covariates. Downloaded from https://psyarxiv.com/pj7gf/.

Kim, D. S., & McCabe, C. J. (2022). The partial derivative framework for substantive regression effects. *Psychological Methods*, 27, 121–141.

Kindo B., Wang H. & Peña E. (2016). Multinomial probit Bayesian additive regression trees. *Stat (Int Stat Inst)*, 5, 119-131.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, 30, 666–687.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27, 435–454.

Kirby D. (1984). *Sexuality education: An evaluation of programs and their effects*. Santa Cruz CA: ETR/Network Publications.

Kish, L. (1987). Weighting in Deft2 . *The Survey Statistician*. (pp. 26-30).

Kline, R. (2015). *Principles and practice of structural equation modeling.* Guilford Press. (fourth edition).

Kline, R. (2023). *Principles and practice of structural equation modeling.* Guilford Press. (fifth edition).

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

Koenker, R., & Portnoy, S. (1987). L-Estimation for linear models. *Journal of the American Statistical Association*, 82, 851-857.

Koller, M. (2012). Nonsingular subsampling for S-estimators with categorical predictors. Seminar for Statistics, ETH Zurich, Zurich, Switzerland. Retrieved from http://arxiv.org/pdf/1208.5595.pdf

Koller, M. & Stahel, W.A. (2011), Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics and Data Analysis,* 55, 2504–2515.

Koller, M. & Stahel, W.A. (2017) Nonsingular subsampling for regression S-estimators with categorical predictors, *Computational Statistics*, 32, 631-646.

Koopman, J., Howe, M., Hollenbeck, J. & Sin, H. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*, 100, 194-202.

Korpon, D. (2014). Dominance analysis utilizing the Bayesian framework in linear and logistic regression models. Doctoral dissertation, Towson University.

Kott, P.S. (1991). A model based look at linear regression with survey data. *American Statistician*, 45, 107-112.

Koul, A., Becchio, C. & Cavallo, A. (2018). Cross-validation approaches for replicability in psychology. *Frontiers in Psychology*, 9, 1117.

Kowarik, A & Templ, M. (2016). Imputation with the R package VIM. *Journal of Statistical Software*, 74, 1–16.

Kraemer, H. C. (2014). A mediator effect size in randomized clinical trials. *lnternational Journal of Methods in Psychiatric Research*, 23, 401-410.

Kramer, H. C. & Fairburn, C.G. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, *59*, 877-883.

Kraemer, H. & Gibbons, R. (2009). Why does the randomized clinical trial methodology so often mislead clinical decision making? Focus on moderators and mediators of treatment. *Psychiatric Annals*, 39, 736-745.

Kraemer, H. & Gibbons, R. (2009). Why does the randomized clinical trial methodology so often mislead clinical decision making? Focus on moderators and mediators of treatment. *Psychiatric Annals*, 39, 736-745.

Kraemer H., Kiernan, M., Essex, M. & Kupfer, D. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology*, 27(2S), S101-S108.

Kraemer, H., Mintz, J., Noda, A., Tinklenberg, J. & Yesavage, J. (2006).Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry*, 63, 484-489.

Kreft I. G. G. (1996). Are multilevel techniques necessary? An overview, including simulation studies. Unpublished manuscript, California State University, Los Angeles.

Krieger, N., & Smith, G. (2016). The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology, *International Journal of Epidemiology*, 45, 1787–1808.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25, 178–206.

Kuha, J. (2004). AIC and BIC : Comparisons of assumptions and performance. *Sociological Methods and Research*, 33, 188-208.

Kuhn M. (2008). Building predictive models in R using the caret Package. *Journal of Statistical Software*, 28, 1-26.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C. & Engelhardt, A. (2011). caret: Classification and regression training. R package version 6.07-001.

Kupper, L. (1990). Erratum. *American Journal of Epidemiology*, 131, 578

Kupper, L. & Hafner, K. (1989). How appropriate are popular sample size formulas? *American Statistician*, 43, 101-105.

Lachin, J. (2016). Fallacies of last observation carried forward. *Clinical Trials*, 13, 161–168.

Lachowicz, M., Preacher, K. & Kelley, K. (2018). A novel measure of effect size for mediation analysis. *Psychological Methods*, 23, 244-261.

LaHuis, D., Jenkin,s D., Hartman, M. , Hakoyama S. & Clark, P. (2020). The effects of misspecifying the random part of multilevel models. *Methodology*, 16, 224-240.

Lai, M. & Tse, (2024). Are factor scores measurement invariant? Psychological Methods. Advance online publication.

Landerman, L, Land, K. & Pieper, C. (1997). An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods and Research*, 26, 3-33.

Lang, J., Rothman, K. & Cann, C. (1998) That confounded p-value. *Epidemiology*, 9, 7-8.

Langkamp, D., Lehman, A. & Lemeshow, S. (2010). Techniques for handling missing data in secondary analyses of large surveys. *Academy of Pediatrics*, 10, 205–210

Lapka, C., Jupka, K., Wray, R. & Jacobsen, H. (2008). Applying cognitive response testing in message development and pre-testing. Health Education Research, http://her.oxfordjournals.org/cgi/content/abstract/cym089v1.

Larsen, R. (2011) Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling*, 18, 649-662.

Le, H., Oh, I., Robbins, S., Ilies, R., Holland, E., & Westrick, P. (2011).Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, 96, 113–133.

Lee, B., Lessler, J. & Stuart, E. (2011). Weight trimming and propensity score weighting. *PLoS One*, 6, e18174.

Lee, H., Cashin, A., Lamb, S., Hopewell, S. et al. (2021). A guideline for reporting mediation analyses of randomized trials and observational Studies: The AGReMA Statement. *Journal of the American Medical Association*, 326, 1045-1056.

Lee, M., Lee, G. & (2023). Linear probability model revisited: Why it works and how it should be specified. *Sociological Methods and Research*, 1–15.

Lee, S. (2007). *Structural equation modeling: A Bayesian approach*. John Wiley & Sons.

Lee, S. & Song, X. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57, 131-150.

Lee T. & Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychological Methods*, 26, 466-485.

Leeper, T. (2021). An introduction to 'margins.' Downloaded from https://cran.r-project.org/web/packages/margins/vignettes/Introduction.html

Lefcheck, (2016). Piecewise SEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics, *Methods in Ecology and Evolution*, 7, 573-579.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *American Statistician*, 55, 187-193.

Lenth, R.V. (2008). Java applets for power and sample size. Downloaded from www.stat.uiowa.edu/~rlenth/Power/. Division of Mathematical Sciences, the College of Liberal Arts or The University of Iowa.

Leon, A., Davis, L. & Kraemer, H. (2011).The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, 45, 626-629.

Leth-Steensen, C. & Gallitto, E. (2016). Testing mediation in structural equation modeling: The effectiveness of the test of joint significance. *Educational and Psychological Measurement*, 76, 339–351.

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica*, 65, 1201-1213

Lewis D. (1979). Counterfactual dependence and time's arrow, *Noûs*, 13, 455-476.

Lewis, D. (2000). Causation as influence. *Journal of Philosophy*, 97, 182–197.

Levy, R., & Hancock, G. R. (2007). A framework of statistical tests for comparing mean and covariance structure models. *Multivariate Behavioral Research*, 42, 33–66.

Leyrat, C., Morgan, K., Leurent, B. & Kahan , B.(2018). Cluster randomized trials with a small number of clusters: which analyses should be used? *International Journal of Epidemiology*, 47, 321–331.

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of Mahalanobis distance. *Journal of Experimental Social Psychology*, 74, 150-156.

Li, C. (2018). Probability of superiority SEM (PS-SEM): Detecting probability based multivariate relationships in behavioral research. *Frontiers in Psychology*, 9.

Li, J. & Lomax, R. (2017). Effects of missing data methods in SEM under conditions of incomplete and nonnormal data. *Journal of Experimental Education*, 85, 231-258.

Li, P. & Stuart, E. (2019). Best (but oft-forgotten) practices: missing data methods in randomized controlled nutrition trials. *American Journal of Clinical Nutrition*, 109, 504–508.

Li, X. & Shen, C. (2020). Doubly robust estimation of causal effect: Upping the odds of getting the right answers. *Circulation: Cardiovascular Quality and Outcomes*, 13, e006065

Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest." R News, 2(3),18-22.

Liengaard, B., Becker, J., Bennedsen, M., Heiler, P., Taylor, L. & Ringle, C. (2025). Dealing with regression models' endogeneity by means of an adjusted estimator for the Gaussian copula approach. *Journal of the Academy of Marketing Science*, 53, 279–299.

Lindeman, R.H., Merenda, P.F. and Gold, R.Z. (1980). *Introduction to bivariate and multivariate analysis*. Scott & Foresman.

Linero, A. & Zhang, Q. (2022). Mediation analysis using Bayesian tree ensembles. *Psychological Methods*. Online advance copy.

Liu, L. & Bentler, P. (2011). Quantified choice of RMSEAs for evaluation and power analysis of small differences between structural equation models. *Psychological Methods*, 16, 116–126.

Liu, Y. & Pek, J.(2024). Summed versus estimated factor scores: Considering uncertainties when using observed scores. Psychological Methods. Advance online publication.

Lin, L., Huang, P. &d Weng, L. (2017) Selecting path models in SEM: A comparison of model selection criteria. *Structural Equation Modeling*, 24, 855-869.

Linden, A. & Karlson, K. (2013). Using mediation analysis to identify causal mechanisms in disease management interventions. *Health Services and Outcomes Research Methodology*, 13, 86–108.

Linden, A. & Yarnold, P. (2017). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 24, 353-361.

Little, R. (1991). Inference with survey weights. *Journal of Official Statistics*, 7, 405-424.

Little, T., Bontempo, D., Rioux, C. & Tracy, A. (2022). On the merits of longitudinal multiple group modelling: an alternative to multilevel modelling for intervention evaluations. *International Journal of Research & Method in Education*, 45, 5, 437-449.

Little R., D'Agostino, R., Cohen, M., Dickersin, K., Emerson, S. et al. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 367, 1355–1360.

Little, R. & Rubin, D. (1987). *Statistical analysis with missing data*. Wiley. See also newer edition in 2002.

Little, R. & Rubin, D. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*, 21, 121-145

Little, R. & Yau, L. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods*, 3, 147-159.

Little, T., Rioux, C., Odejimi, O. & Stickley, Z. (2022). *Parceling in structural equation modeling: A comprehensive introduction for developmental scientists*. Cambridge University Press.

Little, T., Siegers, D., & Card, N. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, 13, 59-72.

Little, T. (2013). *Longitudinal structural equation modeling*. New York: Guilford.

Lockhart, R., Taylor, J. Tibshirani, R. & Tibshirani, R. (2014). A significance test for the lasso, *Annals of Statistics*, 42, 413–468.

Lodi, S., Sharma, S., Lundgren, J., Phillips, A. Cole, S. Logan, R....(2016). The per-protocol effect of immediate vs. deferred ART initiation in the START randomized trial. *AIDS*, 30, 2659–2663.

Loeys, T., Moerkerke, B. & Vansteelandt, S. (2015). A cautionary note on the power of the test for the indirect effect in mediation analysis. *Frontiers in Psychology*, 5.

Loh, W. & Dongning, R. (2022). Improving causal inference of mediation analysis with multiple mediators using interventional indirect effects. *Social and Personality Compass*, 16, e12708.

Loh, W., Moerkerke, B., Loeys, T. & Vansteelandt, S. (2022). Disentangling indirect effects through multiple mediators without assuming any causal structure among the mediators. *Psychological Methods*, 27, 982-999.

Lohr, S. L & Liu, J. (1994). A comparison of weighted and unweighted analyses in the National Crime Victimization Survey. *Journal of Quantitative Criminology*, 10, 343-360.

Lonati, S., Rönkkö, M. & Antonakis, J. (2024). Normality assumption in latent interaction models, *Psychological Methods*, In press.

Long, J.S. (1997). *Regression models for categorical and limited dependent variables*. SAGE.

Long, J.S. & Ervin, L. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician*, 54, 217-224.

Long, J. S. & Freese, J. (2006). Regression models for categorical dependent variables using Stata. Second Edition. Stata Press.

Long, J.S. & Mustillo, S.(2018).Using predictions and marginal effects to compare groups in regression models for binary outcomes. *Sociological Methods and Research*, 50, 1284-1320.

Lozano, L., M., García-Cueto, E. and Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4, 73-79.

Lubke, G. & Muthén, B. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, 11, 514-534.

Luborsky, L. (1976). Helping alliances in psychotherapy: The groundwork for a study of their relationship to its outcome. In J. L. Cleghorn (Ed.) *Successful psychotherapy*, pp. 92–116. Brunner/Mazel

Lüdtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13, 203-229.

Lui, K. (2016). *Crossover designs: Testing, estimation, and sample size*. Wiley.

Luque-Fernandez, M., Schomaker, M., Rachet, B., & Schnitzer, M. (2018). Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Statistics in Medicine*, 37, 2530-2546.

Lunceford, J. & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23, 2937–2960.

Luo, W. & Azen, R. (2013). Determining predictor importance in hierarchical linear models using dominance analysis. *Journal of Educational and Behavioral Statistics*, 38, 3-31.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13, 203–229.

McCaffrey, D., Griffin, B., Almirall, D., Slaugher, M., Ramchand, R. & Burgette, L. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, 32, 3388–3414.

McCaffrey D., Ridgeway, G. & Morral, A. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9, 403–425.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107–120.

MacCallum, R.C., Browne, M. W. & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*, 11, 19–35.

MacCallum, R. C., Browne, M. W. and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. P*sychological Methods*, 1, 130-149.

MacCallum, R., Zhang, S., Preacher, K. & Rucker, D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40.

McClelland, G.H. (1975). Psychological processing of evaluative meaning of adverb-adjective combinations. Dissertation Abstracts International, 35(11-B), 5697.

McClelland, G., Irwin, J., Disatnik, D. & Sivan, L. (2017). Multicollinearity is a red herring in the search for moderator variables: A guide to interpreting moderated multiple regression models and a critique of Iacobucci, Schneider, Popovich, and Bakamitsos (2016). *Behavior Research Methods*, 49, 394–402.

McClelland, G. & Judd, C. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376–390.

MacCoun, R. (2013). The puzzling unidimensionality of DSM-5 substance use disorder diagnoses. *Frontiers in Psychiatry*, 1-5.

McGraw, K.O. & Wong, S.P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361-365.

McKelvey, R.D. & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4, 103-120.

MacKinnon, D. P. (2008). *Introduction lo statistical mediation analysis*. Taylor & Francis.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, D. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83-104.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41–62.

MacKinnon, J. & Webb, M. (2013). Wild bootstrap inference for wildly divergent cluster sizes. Downloaded from www.maxwell.syr.edu/uploadedFiles/cpr/events/cpr_camp_ econometrics/papers2013/MacKinnon_Webb.pdf.

MacKinnon, J. & Webb, M. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32, 233-254.

McNeish, D. (2016a). On using Bayesian methods to address small sample problems, *Structural Equation Modeling*, 23, 750-773.

McNeish, D. (2017a) Missing data methods for arbitrary missingness with small samples, *Journal of Applied Statistics*, 44, 24-39.

McNeish, D. (2017b). Challenging conventional wisdom for multivariate statistical models with small samples. *Review of Educational Research*, 87, 1117–1151.

McNeish, D. (2017c) Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research*, 52, 661-670.

McNeish, D. (2017d). Exploratory factor analysis with small samples and missing data. *Journal of Personality Assessment*, 99, 637–652.

McNeish, D, (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23, 412-433.

McNeish, D. (2020). Should We Use F-tests for model fit instead of chi-square in overidentified structural equation models? *Organizational Research Methods*, 23, 487-510.

McNeish, D. (2023). Psychometric properties of sum scores and factor scores differ even when their correlation is 0.98: A response to Widaman and Revelle. *Behavior Research Methods*, 55, 4269–4290.

McNeish, D., & Harring, J. R. (2017). Correcting model fit criteria for small sample latent growth models with incomplete data. *Educational and Psychological Measurement*, 77, 990-1018.

McNeish, D. M., & Stapleton, L. M. (2014). The effect of small sample size on two level model estimates: A review and illustration. *Educational Psychology Review*, 28, 295–314.

McNeish, D. & Stapleton, L. (2016b) Modeling clustered data with very few clusters, *Multivariate Behavioral Research*, 51, 495-518.

McNeish, D., Stapleton, L. & Silverman, R. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22, 114–140.

McNeish, D., An, J., & Hancock, G. R. (2017). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, 100, 43-52.

McNeish, D. & Wolf, M. (2020a). Thinking twice about sum scores. *Behavior Research Methods*, 52, 2287–2305.

McNeish, D. & Wolf, M. G. (2020b). Corrigendum to: Thinking twice about sum scores. *Behavior Research Methods*, 52, 2674.

McNeish, D. & Wolf, M.G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*,28, 61-88.

McNeish, D. & Wolf, M.G. (2022). Dynamic fit cutoffs for one-factor models. *Behavior Research Methods*, 55, 1157-1174.

McNeish, D. (2023). Dynamic fit index cutoffs for categorical factor analysis with Likert-type, ordinal, or binary responses, *American Psychologist*, 78, 1061-1075.

Machado, J. & Santos-Silva, J. (2005) Quantiles for counts. *Journal of the American Statistical Association*, 100, 1226-1237.

Madden, K., Scott, T., McKay, P., Petrisor, B. A., Jeray, K. J., Tanner, S. L., Bhandari, M., & Sprague, S. (2017). Predicting and preventing loss to follow-up of adult trauma patients in randomized controlled trials: An example from the FLOW Trial. *Journal of Bone and Joint Surgery*. 99, 1086–1092.

Madley-Dowd, P. Hughes, R., Tilling, K. & Herona, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, 110, 63-73.

Mallett, K., Turrisi, R., Ray, A., et al. (2011). Do parents know best? Examining the relationship between parenting profiles, prevention efforts, and peak drinking in college students. *Journal of Applied Social Psychology*, 41, 2904-2927.

Mallinckrodt, B., Abraham, W. T., Wei, M., & Russell, D. W. (2006). Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology*, 53, 372–378.

Mändli, F. & Rönkkö, M. (2023). To omit or to include? Integrating the frugal and prolific perspectives on control variable use. *Organizational Research Methods*, 1–24.

Manor, O., Matthews, S., & Power, C. (2000). Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology*, 29, 149-157.

Manor, O., & Zucker, D. M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational Statistics & Data Analysis*, 46, 801–817.

Mantel, N. (1970). Why stepdown procedures in variable selection. *Technometrics*, 12, 621–625.

Marbach, M. & Hangartner, D. (2020). Profiling compliers and noncompliers for instrumental variable analysis. *Political Analysis*, 28, 435–444.

Marra, G. & Wood, S.N. (2011). Practical variable selection for generalized additive models. *Computational Statistics and Data Analysis*, 55, 2372–2387.

Marsh, L. & Cormier, D. (2001). *Spline regression models*. SAGE.

Marsh, H., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's 1999 findings. *Structural Equation Modeling*, 11, 320-341.

Marsh, H., Wen, Z., Hau, K. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9, 275–300.

Marsh, H., Wen, Z., Hau, K., & Nagengast, B. (2013). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 267–308). IAP Information Age Publishing.

Marshall, A., Altman, D. & Holder, R. (2010). Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: A resampling study. *BMC Medical Research Methodology*, 10, 112-122.

Marshall A., Altman, D., Royston, P. & Holder, R. (2010). Comparison of techniques for handling missing covariate data within prognostic modeling studies: A simulation study. *BMC Medical Research Methodology*, 10, 1-16.

Martin, D, Diehr, P., Perrin, E. & Koepsell T., (1993). The effect of matching on the power of randomized community intervention studies. *Statistics in Medicine*, 12, 329-338.

Martinez-Vizcaino, V. Álvarez-Bueno, C., Sanchez-Lopez, M. et al. (2021). A cluster mediation analysis confirms the validity of the "Fat but Fit" paradigm in children's cognitive function and academic achievement. *Journal of Pediatrics*, 231, 231-238.

Mashreghi, Z., Léger, C. & Haziza, D. (2014) Bootstrap methods for imputed data from regression, ratio and hot-deck imputation. *Canadian Journal of Statistics*, 42, 142–167.

Mauro, R. (1990). Understanding LOVE (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin*, *108*, 314-329.

Maxwell, S. (2000). Sample size and multiple regression analysis. *Psychological Methods*, 5, 434-458.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*, 23–44.

Maxwell, S. E., Cole, D. A. & Mitchell, M. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model *Multivariate Behavioral Research*, *46*, 816–841.

Maxwell, S., Delaney, H. & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Routledge.

Maxwell, S., Kelley, K. & Rausch, J. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537-563.

Maydeu-Olivares, A. (2017a). Assessing the size of model misfit in structural equation models. *Psychometrika*, 82, 533–558.

Maydeu-Olivares, A. (2017b). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, 24, 383–394.

Maydeu-Olivares, A., Shi, D. & Rosseel, Y. (2017). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling*, 25, 1-14.

Maydeu-Olivares, A., Fairchild, A. & Hall, A. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 24, 495–505.

Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of confidence intervals. *Philosophy of Science*, 48, 269–280.

Mayo, D. & Cox, D. R. (2006). Frequentist statistics as a theory of inductive inference. *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, 49, 77–97.

Meade, A., Michels, L., & Lautenschlager, G. (2007). Are internet and paper-and-pencil personality tests truly comparable?: An experimental design measurement invariance study. *Organizational Research Methods*, 10, 322-335.

Meehl, P. (1970). Nuisance variables and the ex post facto design. In M.Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science: Vol. IV*. (pp. 373-402). University of Minnesota Press.

Meehl, P. (1971) High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology*, 77, 143-148.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10, 259–284.

Menard, S. (2007). Comment on 'Estimators of Relative Importance in Linear Regression Based on Variance Decomposition', *American Statistician*, 61, 280–283.

Menard, S. (2011).Standards for standardized logistic regression coefficients. *Social Forces*, 89, 1409–1428.

Menzies, P. (2017). Counterfactual theories of causation. In Stanford encyclopedia of philosophy. Retrieved from https://plato.stanford.edu/archives/win2017/entries/causation-counterfactual.

Merkle, E. C., You, D., & Preacher, K. (2016). Testing non-nested structural equation models. *Psychological Methods*, 21,151-163.

Meyer, D. L. (1991). Misinterpretation of interaction effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin*, 110, 571-573.

Miller, J., & Ulrich, R. (2016). Interpreting confidence intervals: A comment on Hoekstra et al. (2014). *Psychonomic Bulletin & Review*, 23, 124-130.

Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences*, 42, 875–881.

Millsap, R. E. (2013). A simulation paradigm for evaluating model fit. In M. Edwards & R. MacCallum (Eds.), *Current issues in the theory and application of latent variable models* (pp. 165–182). Routledge

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., et al. (2010). CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*, 1–37.

Mooijaart, A. & Satorra, A. (2009). On insensitivity of the chi square model test to non-linear misspecification in structural equation models. *Psychometrika*, 74, 443-455.

Molenberghs, G. & Kenward, M. (2007). *Missing data in clinical studies*. Wiley.

Montgomery, J., Nyhan, B. & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62, 760-775.

Montoya, A. & Hayes, A. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, 22, 6–27

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26, 67–82.

Mood, C. (2017). Logistic regression: Uncovering unobserved heterogeneity. Downloaded fromhttps://www.su.se/polopoly_fs/1.341160.1501927239!/menu/standard/file/Logit2%20%286%29.pdf.

Morey, R., Hoekstra, R., Rouder, J. & Wagenmakers, E. (2016). Continued misinterpretation of confidence intervals: Response to Miller and Ulrich. *Psychonomic Bulletin Review*, 23, 131–140.

Morey, R., Hoekstra, R., Rouder, J., Lee, M. & Wagenmakers, E. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin Review*, 23, 103–123.

Morgan, S., & Winship, C. (2007). *Counterfactuals and causal inference*. Cambridge University Press.

Morimoto, J. (2022). The effect of sample size and missingness on inference with missing data. Downloaded from https://arxiv.org/abs/2112.09275.

Morris, T., White, R. & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14, 75-87.

Morrison, D. M., Gillmore, M. R., Simpson, E. E., Wells, E. A., & Hoppe, M. J. (1996). Children's decisions about substance use: An application and extension of the theory of reasoned action. *Journal of Applied Social Psychology*, 26, 1658–1679.

Moosbrugger, H., Schermelleh-Engel, K., Kelava; A. & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 103-136). Sense Publishers.

Mouelhi, Y., Jouve, E., Castelli, C. & Gentile, S. (2020). How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health and Quality of Life Outcomes*, 18, 1-17.

Muggeo, V. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22, 3055–3071.

Muggeo, V. (2008). Segmented: An R package to fit regression models with broken-line relationships. R News, 8, 20-25. Retrieved on March 2010 from http://cran.r-project.org/doc/Rnews/Rnews_2008-1.pdf

Muller, C. (2019). Linear quantile regression and endogeneity correction. *Biostatistics and Biometrics*, 9, 123-128.

Muller, C. & MacLehose, R. (2014). Estimating predicted probabilities from logistic regression: Different methods correspond to different target populations. International *Journal of Epidemiology*, 43, 962–970.

Murad, H., Fleishman, A., Sadetzki, S., Geyer, O. & Freedman, L. (2003). Small samples and ordered logistic regression: Does it help to collapse categories of outcome? *American Statistician*, 57, 155-160.

Mustillo, S. (2013). The effects of auxiliary variables on coefficient bias and efficiency in multiple imputation. *Sociological Methods and Research*, 41, 335-361.

Mustillo, S. & Kwon, S. (2015). Auxiliary variables in multiple imputation when data are missing not at random. *The Journal of Mathematical Sociology*, 39, 73 - 91.

Muthén, B. (2010). Bayesian analysis in Mplus: A brief introduction. Technical report. Los Angeles: Muthén & Muthén. www.statmodel.com.

Muthén, B. (2011). Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. Downloaded from https://www.statmodel.com /download/causalmediation.pdf

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

Muthén, B. O. & Asparouhov, T. (2015). Estimator choices with categorical outcomes. Downloaded on March 12, 2016 from www.statmodel.com/download/Estimator Choices.pdf.

Muthén, B. & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2, 371–402.

Muthén, B., du Toit, S.H.C. & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Downloaded from https://www.statmodel.com/download/Article_075.pdf.

Muthén, L. & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599-620.

Muthén, B. & Muthén, L. (2005). WLSMV missing data. At http://www.statmodel.com/discussion/messages/22/824.html?1521157428

Muthén, B., Muthén, L., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén and Muthén.

Myers, J., Rassen, J., Gagne, J., Huybrechts, K., Schneeweiss, S., Rothman, K. et al. (2011a). Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology*, 174, 1213–1222.

Myers, J., Rassen, J., Gagne, J., Huybrechts, K., Schneeweiss, S., Rothman, K. et al. (2011b). Myers et al. respond to "Understanding Bias Amplification". *American Journal of Epidemiology*, 174, 1228–1229.

Naimi, A., Kaufman, J. & MacAclehoseE, R. (2014). Mediation misgivings: Ambiguous clinical and public health interpretations of natural direct and indirect effects. *International Journal of Epidemiology*, 43, 1656-1661.

Namboodiri, K. (1984). *Matrix algebra: An introduction*. Thousand Oaks, CA: . SAGE.

National Institute of Mental Health (2019) Strategic research priorities. Downloaded at www.nimh.nih.gov/about/strategic-planning-reports/strategic-research-priorities/index.

National Research Council 2010. *The pevention and treatment of missing data in clinical trials*. National Academies Press.

Nevitt, J. & Hancock, G. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling*, 8, 353-377.

Newgard, C. & Lewis, R. (2015). Missing data: How to best account for what is not known. *Journal of the American Medical Association*, 314, 940–9411.

NieBen, D., Partch, M. Kemperm C. & Rammstedt, B. (2019). An English-language adaptation of the social desirability–gamma short scale (KSE-G). *Measurement Instruments for the Social Sciences*, 2, 1-10.

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, 39, 439–478.

Newman, D. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17, 372-410.

Ng, T., & Feldman, D. (2011). Affective organizational commitment and citizenship behavior: Linear and non-linear moderating effects of organizational tenure. *Journal of Vocational Behavior*, 79, 528-537.

Nguyen, T., Schmid, I. & Stuart, E. (2021). Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. *Psychological Methods*, 26, 255–271.

Nichols, A. & Schaffer, M. (2007). Clustered standard errors in Stata. United Kingdom Stata Users' Group Meetings 2007 from Stata Users Group.

Nilsson, A., Bonander, C., Strömberg, U. & Björk, J. (2021). A directed acyclic graph for interactions. *International Journal of Epidemiology*, 50, 613–619.

Niolon, P., Vivolo-Kantor, A., Tracy, A. et al., (2019). An RCT of dating matters: Effects on teen dating violence and relationship behaviors. *American Journal of Preventive Medicine*, 57, 13–23.

Nordberg, L. (1989). Generalized linear modeling of sample survey data. *Journal of Official Statistics*, 5, 223-239.

Norton, E. (2012). Log odds and ends. National Bureau of Economic Research. Downloaded at https://www.nber.org/system/files/working_papers/w18252/w18252.pdf

Norton, E., Dowd, B. & Maciejewski, M. (2018). Odds ratios — current best practice and use. *Journal of the American Medical Association*, 320, 84-85.

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Nye, C. D., & Drasgow, F. (2011). Assessing goodness of fit: Simple rules of thumb simply do not work. *Organizational Research Methods*, 14, 548-570.

Oberski, D. (2014). lavaan. survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, 57, 1–27.

Oberski, D. & Satorra, A. (2013). Measurement error models with uncertainty about the error variance. *Structural Equation Modeling*, 20, 409-428.

Oehlert, G. W. (1992). A note on the delta method. *American Statistician*, 46, 27–29.

O'Flaherty, M., Kalucza, S. & Bon, j. (2023). Does anyone suffer from teenage motherhood? Mental health effects of teen motherhood in Great Britain are small and homogeneous. *Demography*, 60, 707–729.

O'Keefe, D. (2003). Colloquy: Should familywise alpha be adjusted? *Human Communication Research*, 29, 431-437.

Olejnik, S. & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25, 241–286.

Onwuegbuzie, A. & Daley, C. (2001). Racial differences in IQ revisited: A synthesis of nearly a century of research. *Journal of Black Psychology*, 27, 209-220.

O'Rourke, H. & MacKinnon, D. (2015). When the test of mediation is more powerful than the test of the total effect. *Behavior Research Methods*, 47, 424–442.

Pampel, F. (2021). *Logistic regression: A primer*. SAGE.

Pan, Z. & Kupper, L. (1999). Sample size determination for multiple comparison studies treating confidence interval width as random. *Statistics in Medicine*, 18, 1475-1488.

Pang, M., Schuster, T., Filion K., Eberg, M. & Platt R. (2016). Targeted maximum likelihood estimation for pharmacoepidemiologic research. *Epidemiology*, 27, 570- 577.

Papa, A. & Follette, W. (2015). Dismantling studies of psychotherapy. *Encyclopedia of Clinical Psychology*. Wiley.

Park, J., Cardwell, R. & Yu, H. (2020). Specifying the random effect structure in linear mixed effect models for analyzing psycholinguistic data. *Methodology*, 16, 92–111.

Park I. & Lee, H. (2004). Design effects for the weighted mean and total estimators under complex survey sampling. *Survey Methodology*, 3, 183–193.

Park, S. & Esterling, K. (2021). Sensitivity analysis for pretreatment confounding with multiple mediators. *Journal of Educational and Behavioral Statistics*, 46, 85–108.

Park, S. & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, 31, 567–86.

Pavlov, G., Shi, D. & Maydeu-Olivares, A. (2021). Using the standardized root mean squared residual (SRMR) to assess exact fit in structural equation models. *Educational and Psychological Measurement*, 81, 110-130.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearl, J. (2010). An introduction to causal inference. *International Journal of Biostatistics*, 6, 1-59.

Pearl, J. (2011). Invited commentary: understanding bias amplification. *American Journal of Epidemiology*, 174, 1223-1227.

Pearl, J., Glymour, M. & Jewell, N.P. (2016). *Causal inference in statistics*. Wiley.

Pearl, J. & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Pek, J. & Flora, D. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, 23, 208–225.

Pek, J., & MacCallum, R. C. (2011). Sensitivity analysis in structural equation models: Cases and their influence. *Multivariate Behavioral Research*, 46, 202–228.

Pelleg, D. & Moore, A. (2000). X means: Extending K-means with efficient estimation of the number of clusters. Retrieved on August 4, 201 from http://www.aladdin.cs.cmu.edu/papers/pdfs/y2000/xmeans.pdf.

Perepletchikova, F. (2011) On the topic of treatment integrity. *Clinical Psychology*, 18, 148–153.

Perepletchikova, F., Treat, T.A. & Kazdin, A.E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75, 829-841.

Perezgonzalez, J. D. (2017). The fallacy of placing confidence in confidence intervals - A commentary. Downloaded from https://doi.org/10.31234/osf.io/kvxc4

Peterson, B. L., & Harrell, F. E. (1988). Partial proportional odds models and the LOGIST procedure. Paper presented at SAS Users Group International, Thirteenth Annual Conference, Orlando, FL.

Petty, R., Fabrigar, L., Wegener, D. & Priester, J. (1996). Understanding data when interactions are present or hypothesized. *Psychological Science*, 7, 247-252.

Peugh, J., Strotman, D., McGrady, M., Rausch, J. & Kashikar-Zuck, S. (2017). Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer. *Journal of School Psychology*, 60, 217-224.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61, 317-337.

Pfeffermann, D. & Sverchkov, M. (2009). Inference under informative sampling. *Sample Surveys: Inference and Analysis*, 29, 455-489.

Pischke, J. S. (2012). Probit better than LPM?. Downloaded from http://www.mostly harmlesseconometrics.com/2012/07/probit-better-than-lpm/

Plano Clark, V. (2010). The adoption and practice of mixed methods: U.S. trends in federally funded health related research. *Qualitative Inquiry*, *16*, 428–440.

Pornprasertmanit, S., Wu, W. and Little, T. (2013). Taking into account sampling variability of model selection indices: A parametric bootstrap approach. *Multivariate Behavioral Research*, 48, 168-169.

Porter, K., Miratrix, L., Hunter, K. & Htet, Z. (2023a). Power Under multiplicity project. Downloaded from https://cran.r-project.org/web/packages/PUMP/PUMP.pdf.

Porter, K., Miratrix, L., Hunter, K. & Htet, Z. (2023a). Demonstration of the PUMP package. Downloaded from https://cran.r-project.org/web/packages/PUMP/vignettes/pump_demo.html.

Porter, S. (2015). Quantile regression: Analyzing changes in distributions instead of means. In M.B. Paulsen (Ed.), *Higher education: Handbook of theory and research*. Spinger.

Pratt, J.W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In: Pukkila, T. and Puntanen, S. (Eds.): Proceedings of second Tampere conference in statistics, University of Tampere, Finland, 245-260.

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, 66, 825–852.

Preacher, K. J., Cai, L., & MacCallum, R. C. (2007). Alternatives to traditional model comparison strategies for covariance structure models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 33-62). Mahwah, NJ: Lawrence Erlbaum Associates.

Preacher, K. & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Method*s, 16, 93-115.

Preacher, K. & Merkle, E. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, 17, 1–14.

Preacher, K. & Selig, J. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6, 77–98

Prentice, D.A. & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, 112, 160-164.

Price, L. (2016). *Psychometric methods: Theory into practice*. Guilford.

Price, R., van Ryn, M. & & Vinokur, A. D. (1992). Impact of preventive job search intervention on the likelihood of depression among the unemployed. *Journal of Health and Social Behavior*, 33, 158–167.

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36, 672–683.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, 69, 167–190.

Raftery, A.E. (1995). Bayesian model selection in social research (with discussion). *Sociological Methodology*, 25, 111-195.

Ramaswamy, V., DeSarbo, W., Reibstein, D. & Robinson, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12, 103−124.

Raskin, A., Schulterbrandt, J., Reatig, N., et al (1967). Factors of psychopathology in interview, ward behavior, and self-report ratings of hospitalized depressives. J*ournal of Consulting Psychology*, 31, 270-278.

Raudenbush, S. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. SAGE.

Raush, J., Maxwell, S. & Kelley, K. (2003). Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design. *Journal of Clinical Child and Adolescent Psychology*, 32, 467-486. ·

Raykov, T. (2011). On testability of missing data mechanisms in incomplete data sets. *Structural Equation Modeling*,18, 419-429.

Raykov, T., Lichtenberg, P. and Paulson, D. (2012). Examining the missing completely at random mechanism in incomplete data sets: A multiple testing approach. *Structural Equation Modeling*, 19, 399-408.

Raykov, T. & Marcoulides, G. (2014). Identifying useful auxiliary variables for incomplete data analyses: A note on a group difference examination approach. *Educational and Psychological Measurement*, 74, 537-550.

Reichardt, C.S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford.

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*,169, 12-22.

Rhemtulla, M., van Bork, & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, 25, 30–45.

Ridout, M., Demétrio, C. & Firth, D. (1999). Estimating intraclass correlation for binary data. *Biometrics*, 55, 137-148.

Rijnhart, J., Valente, M., Smyth, H. & MacKinnon, D. (2023). Statistical mediation analysis for models with a binary mediator and a binary outcome: The differences between causal and traditional mediation analysis. *Prevention Science*, 24, 408–418.

Ritter, S., Jewell, N. & Hubbard, A. (2014). R package multiPIM: A causal inference approach to variable importance analysis. *Journal of Statistical Soft*ware, 57, 1–29.

Ritter, N. & Vance, C. (2011). The phantom menace of omitted variables: A comment. Economic Papers, Ruhr-Universität Bochum (RUB), Department of Economics Universitätsstr. 150, 44801 Bochum, Germany.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modeling*, 7, 1393–1512.

Robins, J. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. J. Green, N. L. Hjort, and S. Richardson (Eds.), *Highly structured stochastic systems* (pp. 70–81). Oxford University Press.

Robins, J. & Greenland, S. (1992). Identifiability and exchangeability of direct and indirect effects. *Epidemiology*, 3, 143-155.

Robins, J., Sued, M., Lei-Gomez, Q. and Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when inverse probability weights are highly variable. *Statistical Science*, 22, 544-559.

Robins, J. M., van der Vaart, A., & Ventura, V. (2000). Asymptotic distribution of p values in composite null models. *Journal of the American Statistical Association*, 95, 1143-1156.

Roger, J. (2012). Missing data in clinical trials: A data interpretation problem with statistical solutions? *Clinical Investigation*, 2, 19–24.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1, 27–42.

Rohrmann, B. (2015). Designing verbalized rating scales: Sociolinguistic concepts and psychometric findings from three cross-cultural projects. Downloaded from http://www.rohrmannresearch.net/pdfs/vqs-projects.pdf

Rönkkö, M. (2020). Missing data mechanisms. View video at https://www.youtube.com/watch?v=TxcLeMsZ7Bk

Rönkkö, M. & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods,* 25, 6–47.

Rönkkö, M., McIntosh, C., Antonakis, J. & Edwards, J. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management*, 47, 9–27.

Rönkkö, M., Lee, N., Everman, J., McIntosh, C., & Antonakis, J. (2023). Marketing or methodology? Exposing the fallacies of PLS with simple demonstrations. *European Journal of Marketing*, 57, 1597-1617.

Rosato, R., Pagano, E., Testa, S., Zola, P. & Cuonzo, D. (2021). Missing data in longitudinal studies: Comparison of multiple imputation methods in a real clinical setting. *Journal of Evaluation in Clinical Practice*, 27, 34-41.

Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Rosenbaum, P. & Rubin, D (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79, 516–524.

Rosenberger, W. & Lachin, J. (2015). *Randomization in clinical trials: Theory and practice*. Wiley.

Rosenthal, R. (1995). Methodology. In. A. Tesser (Ed.), *Advanced social psychology* (pp.17-50). McGraw-Hill.

Rosnow, R. & Rosenthal, R. (1989a). Definition and interpretation of interaction eflects. *Psychological Bulletin*, 105, 143-146.

Rosnow, R. & Rosenthal, R. (1989b). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276-1284.

Rosnow, R. & Rosenthal, R. (1991). If you're looking at the cell means. you're not looking at only the interaction (unless all main effects are zero). *Psychological Bulletin*, 110, 574-576.

Rosnow, R. & Rosenthal, R. (1995). Some things you learn aren't so: Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science*, 1, 4-9.

Rosnow, R & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221-237.

Ross, W. & Creyer, E. (1993). Interpreting Interactions: Raw means or residual means? *Journal of Consumer Research*, 20, 330-338.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. Wiley.

Rousseuw, P. J. & van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85, 663-669.

Royse, D., Thyer, B. & Padgett, D. (2015). *Program evaluation: An introduction to an evidence-based approach*. Cengage.

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, 13, 19 –30.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688–701.

Rubin, D. (1976). Inference and missing data. Biometrika, *63*, 581 –592.

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization, *Annals of Statistics*, 6, 34–58.

Russell, B. (1931). *The scientific outlook*. Allen & Unwin.

Russo, F., Wunsch, G. & Mouchart, M. (2011). Inferring causality through counterfactuals in observational studies: Some epistemological issues. *Bulletin of Sociological Methodology*, 111, 43-64.

Rutterford, C., Copas, A. & Eldridge, S. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology*, 44, 1051–1067.

Ryu, E. & West, S.G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *BStructural Equation Modeling*, 16, 583-601.

Sackett, D.L. (1979). Bias in analytic research. *Journal of Chronic Diseases*, 32, 51-63.

Salivar, E., Knopp, K., Roddy, M. K., Morland, L. A., & Doss, B. D. (2020). Effectiveness of online OurRelationship and ePREP programs for low-income military couples. *Journal of Consulting and Clinical Psychology*, 88, 899–906.

Saltelli, A., Tarantola, S., Campolongo, F. & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. Wiley.

Salovey, P. & Mayer, J. (1989). Emotional intelligence. *Imagination, Cognition and Personality*, *9*, 185-211.

Sampson, A. (1974). A tale of two regressions. *Journal of the American Statistical Association*, 69, 682-689.

Samuels, M. & Lu, T. (1992). Sample size requirements for the back-of-the-envelope binomial confidence interval. *American Statistician*, 46, 228-231.

Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8, 448–467.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26, 393–415.

Saris, W.E., A. Satorra and W.M. Van der Veld (2009). Testing structural equation models for detection of misspecifcations. *Structural Equation Modeling*, 16, 561-582.

Satten, G. & Kupper, L. (1990). Sample size requirements for interval estimation of THIS odds ratio. *American Journal of Epidemiology*, 131, 177-184.

Saunders, C. & Blume, J. (2018). A classical regression framework for mediation analysis: Fitting one model to estimate mediation effects. *Biostatistics*, 19, 514–528.

Savalei, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, 72, 910–932.

Savalei, V., Brace, J. & Fouladi, R. (2022). We need to change how we compute RMSEA for nested model comparisons in structural equation modeling. Downloaded from https://psyarxiv.com/wprg8/

Savalei, V., & Falk, C. F. (2014). Robust two-stage approach outperforms robust full information maximum likelihood with incomplete nonnormal data. *Structural Equation Modeling*, 21, 280–302.

Savalei, V., & Reise, S. P. (2019). Don't Forget the Model in Your Model-based Reliability Coefficients: A Reply to McNeish (2018). *Collabra: Psychology*, 5, 36.

Schafer, J. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall.

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

Schäfer, T. & Schwarz, M.A. (2019) The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, 10-813, 1-13.

Schenker, N. & Taylor, J. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics & Data Analysis*, 22, 425-446.

Schomaker, M. & Heumann, C. (2018). Bootstrap inference when using multiple imputation. *Statistics in Medicine*, 37, 2252–2266.

Schönemann, P. H., & Steiger, J. H. (1978). On the validity of indeterminate factor scores. *Bulletin of the Psychonomic Society*, 12, 287-290.

Schuler, M. & Rose, S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology*, 185, 65-73.

Schumacker, R. & Lomax, R. (2016). *A beginner's guide to structural equation modeling*. Routledge.

Schuster, N. Twisk, J., Heymans, M. & Rijnhart, J. (2023) Causal mediation analysis with a binary mediator: The influence of the estimation approach and causal contrast, *Structural Equation Modeling*, 30, 481-490.

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 2, 333-343.

Schwartz, D. & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, 20, 637–648.

Schwartz, D. & Lellouch, J (1967), Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, 20, 637–648.

Seaman, S., Bartlett, J. & White, I. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: An evaluation of statistical methods. *BMC Medical Research Methodology*, 12, 1-13.

Seaman, S. & White, I. (2014). Inverse probability weighting with missing predictors of treatment assignment or missingness. *Communications in Statistics: Theory and Methods*, 43, 3499-3515.

Seaman,S., White, I. & Copas, A. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics*, 68, 129-137.

Sellbom M. & Tellegen A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31, 1428-1441.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. Wiley.

Senn S. (1994). Testing for baseline balance in clinical trials. *Statistics in Medicine*, 13, 1715–1726.

Senn, S. (2012). Dichotomania: An obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical trials. Paper presented at the International Statistical Institute Conference, 55th Session, Downloaded from https://www.researchgate.net/publication/221689734_Dichotomania_an_obsessive_compulsive_disorder_that_is_badly_affecting_the_quality_of_analysis_of_pharmaceutical_trials.

Senn, S. (2013). Seven myths of randomization in clinical trials. *Statistics in Medicine*, 32, 1439–1450.

Serang, S., Jacobucci, R., Brimhall, K., & Grimm, K. (2017). Exploratory mediation analysis via regularization, *Structural Equation Modeling*, 24, 733–744.

Serlin, R. (1993). Confidence intervals and the scientific method: A case for Holm on the range. *Journal of Experimental Education*, 61, 350-360.

Shadish, W. R., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611.

Shen, E., Chou, C., Pentz, M. & Berhane, K. (2014) Quantile mediation models: A comparison of methods for assessing mediation across the outcome distribution, *Multivariate Behavioral Research*, 49, 471-485.

Shi, D., Maydeu-Olivares, A. & Distefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, 53, 676-694.

Shi, D., Lee, T., & Maydeu-Olivares, A. (2018). Understading the model size effect on SEM fit indices. *Educational and Psychological Measurement*, 79, 1–25.

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling*, 25, 21–40.

Shipley, B. (2013). The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology*, 94, 560–564.

Shipley, B. (2003) Testing recursive path models with correlated errors using d-separation. *Structural Equation Modeling*, 10, 214-221.

Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, 7, 206-218.

Shipley, B. & Douma, J. (2020). Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs. *Ecology*, 101, e02960

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422–445.

Sijtsma, K., Ellis, J. & Borsboom, D. (2024). Recognize the value of the sum score: Psychometrics' greatest accomplishment. *Psychometrika*, 89, 84–117.

Skene, S. S., & Kenward, M. G. (2010a). The analysis of very small samples of repeated measurements I: An adjusted sandwich estimator. *Statistics in Medicine*, 29, 2825–2837.

Smid, S., McNeish, D., Miočević, M. & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review, *Structural Equation Modeling*, 27, 131-161.

Smith, G. A. (1986). Observer drift: A drifting definition. *Behavior Analysis*, 9, 127–128. Solomon, P., Cavanaugh, M. & Draine, J. (2009). *Randomized controlled trials: Design and implementation for community-based psychosocial interventions.* Oxford University Press.

Smith, M. & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760

Smith, R. (2012). Stratified, personalised, or precision medicine. *British Medical Journal Opinion*. Downloaded at http://blogs.bmj.com/bmj/2012/10/15/richard-smith-stratified-personalised-or-precision-medicine/

Snijders, T. & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. SAGE.

Snowden, J., Rose, S. & Mortimer, K. (2011). Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *American Journal of Epidemiology*, 173, 731–738.

Sobel, M. & Muthén, B. (2012). Compliance mixture modelling with a zero effect complier class and missing data. *Biometrics*, 68, 1037-1045.

Solomon, P., Cavanaugh, M., & Draine, J. (2009). *Randomized controlled trials: Design and implementation for community-based psychosocial interventions*. Oxford University Press.

Song, X. & Lee, S.(2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika*, 67, 261-288.

Souto1, H. & Neto, F. (2024). K-fold causal BART for CATE estimation. Downloaded from https://arxiv.org/pdf/2409.05665?

Sowa, J. (2000). Processes and causality. Retrieved from www.jfsowa.com/ontology/causal.htm.

Sparapani, R., Spanbauer, C., & McCulloch, R. (2021). Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package. *Journal of Statistical Software*, 97, 1–66.

Speilberger, C. & Vagg, P. (1995). *Test anxiety: Theory, assessment, and treatment*. Taylor and Francis.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*, 64, 583–639.

Stancombe, K. (2019). Web exclusive: Text message reminders curb failure to appear. *The Indiana Lawyer*. April 30. Downloaded from www.theindianalawyer.com/articles/50159-web-exclusive-text-message-reminders-curb-failures-to-appear.

Stanghellinia, E. & Kateri, M. (2023). Exact mediation analysis for ordinal outcome and binary mediator. *Epidemiology*, 33, 840–842.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, 41, 481–520

Steiger, J. & Lind, J. (1980). Statistically based tests for the number of factors. Paper presented at the annual spring meeting of the Psychometric Society, Iowa City, Iowa.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling*, 7, 149-162

Sterba, S. (2017) Partially nested designs in psychotherapy trials: A review of modeling developments, *Psychotherapy Research*, 27, 425-436,

Sterba, S. K., & Pek, J. (2012). Individual influence on model selection. *Psychological Methods*, 17, 582–599.

Sterba, S. Preacher, K., Forehand, R., Hardcastle, E., Cole, D. & Compas, B. E. (2014). Structural equation modeling approaches for analyzing partially nested data. *Multivariate Behavioral Research*, 49, 93–118.

Stevens, S. S. (1951). Mathematics, measurement, psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). Wiley.

Stoeber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17, 222-232.

Strube, M. (1991). Small sample failure of random assignment: A further examination. *Journal of Consulting and Clinical Psychology, 59*, 346-350.

Strube, M. (2015). Small sample failure of random assignment. In R. Cautin and S. Lilienfeld (Eds.) *Encyclopedia of clinical psychology*. Wiley.

Stuart, E., Perry, D., Le, H. & Ialongo, N. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science*, 9, 288-298.

Substance Abuse and Mental Health Services Administration (2014). Developing a logic model to guide program evaluation. Downloaded on March 10, 2012 from www.samhsa.gov/capt/sites/default/files/resources/developing-logic-model-guide.pdf

Swanson, S., Tiemeier, H., Ikram, M. & Hernán, M. (2017). Nature as a trialist? Deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology*, 28, 653-659.

Swanson S., Miller M., Robins, J. & Hernán M. (2015). Selecting on treatment: A pervasive form of bias in instrumental variable analysis. *American Journal of Epidemiology*, 181, 191-197.

Tan, Y. & Roy, J. (2019). Bayesian additive regression trees and the General BART model. *Statistics and Medicine*, 38, 5048–5069.

Taylor, A., West, S. & Aiken, L. (2006). Loss of power in logistic, ordinal logistic, and probit regression when an outcome variable is coarsely categorized. *Educational and Psychological Measurement*, 66, 228-239.

Taylor, A., MacKinnon, B. P., & Tein, J.-Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11, 241-269.

Templ, M., Kowarik, A. & Filzmoser, P. (2011) Iterative stepwise regression imputation using standard and robust methods. *Journal of Computational Statistics and Data Analysis*, 55, 2793- 2806.

Thabane, L., Mbuagbaw, L., Zhang, S. et al. (2013). A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. *BMC Medical Research Methodology*, 13, 92.

Thoemmes, F. & Rose, N. (2014). A cautious note on auxiliary variables that can increase bias in missing data problems. *Multivariate Behavioral Research*, 49, 443–459.

Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, 17, 510-534.

Thomas, E. J. & Rothman, J. (2013). *Intervention research: Design and development for human service*. New York: Taylor Francis.

Thomas, R., Zumbo, B., Kwan, E. & Schweitzer, L. (2014) On Johnson's (2000) relative weights method for assessing variable importance: A reanalysis. *Multivariate Behavioral Research*, 49, 329-338.

Thomas, S. & Heck, R. (2001). Analysis of large scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education*, 42, 517–540

TIBCO (2020). TIBCO Statistica 14.1 user guide. Downloaded from https://docs.tibco.com/pub/stat/14.1.0/doc/html/UserGuide/GUID-A168AF7a-BC67-8CBE-1EE7FE60282D.html.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, Series B 58: 267–288.

Timm, N. & Weisner, A. (2003). Model selection and prediction in multiple linear regression. University of Pittsburgh. Retrieved from ftp://ftp.pitt.edu/users/t/i/timm/Neter/Linpapr.PDF on March, 3, 2012.

Tofighi, D. & Kelley, K. (2020). Indirect effects in sequential mediation models: Evaluating methods for hypothesis testing and confidence interval formation. *Multivariate Behavioral Research*, 55, 188–210.

Tofighi, D. & Kelley, K. (2020). Improved inference in mediation analysis: Introducing the model-based constrained optimization procedure. Psychological Methods, 25, 496–515.

Tofighi, D. & MacKinnon, D. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, 43, 692-700.

Tofighi, D., & MacKinnon, D. P. (2015). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling*, 23, 194-205.

Toh S. & Hernán M. (2008). Causal inference from longitudinal studies with baseline randomization. *International Journal of Biostatistics*, 4, Article 22.

Toh S., Hernández-Díaz, S., Logan, R., Robins, J. & Hernán, M. (2010). Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology*, 21, 528-539.

Tonidandel, S. & LeBreton, J.M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, 26, 1-9.

Torrance, G. E. (1986). Measurement of health state utilities for economic appraisal: A review. Jou*rnal of Health Economics*, 5, 1–30.

Tsiatis, A., Davidian, M., Zhang, M. & Lu X. (2008). Covariate adjustment for two-sample treatment comparison in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, 27, 4658–4677.

Tu, Y. & Gilthorpe, M. (2007). Revisiting the relation Between change and initial value: A review and evaluation. *Statistics in Medicine*, 26, 443–457.

Turner E., Li, F., Gallis, J., Prague, M. & Murray D (2017). Review of recent methodological developments in group-randomized trials: Part 1--Design. *American Journal of Public Health*, 107, 907-915.

Turner, E., Prague, M., Gallis, J., Li, F. & Murray, D. (2017). Review of recent methodological developments in group-randomized trials: Part 2-Analysis. *American Journal of Public Health*, 107, 1078-1086.

Tutz, G. (2020). Modelling heterogeneity: on the problem of group comparisons with logistic regression and the potential of the heterogeneous choice model. *Advances in Data Analysis and Classification*, 14, 517–542.

Uanhoro, J. (2017). A Chi-Square test of close fit in covariance-based SEM. Downloaded at https://www.jamesuanhoro.com/post/2017/11/16/a-chi-square-test-of-close-fit-in-covariance-based-sem/.

Uanhoro, J., Wang, Y., & O'Connell, A. (2019) Problems with using odds ratios as effect sizes in binary logistic regression and alternative approaches. *Journal of Experimental Education*, DOI: Online at https://www.tandfonline.com/doi/full/10.1080/00220973.2019.1693328.

Valente, M., Gonzalez, O., Miocevic, M. & MacKinnon, D. (2016). A note on testing mediated effects in structural equation models: Reconciling past and current research on the performance of the test of joint significance. *Educational and Psychological Measurement*, 76, 889–911.

Valente, M. J., & MacKinnon, D. P. (2017). Comparing models of change to estimate the mediated effect in the pretest-posttest control group design. *Structural Equation Modeling*, 24, 428–450.

van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman and Hall.

van Buuren, S., Boshuizen, H. & Knook, D. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18, 681-694.

van Buuren, S. & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R, Journal of Statistical Software, 45, 1-67.

van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & Van Loey, N. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, 6, 25216.

Vandenbroucke, J., Broadbent, A. & Pearce, N. (2016). Causality and causal inference in epidemiology: The need for a pluralistic approach. *International Journal of Epidemiology*, 45, 1776–1786.

van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & van Loey, N. (2018). Bayesian PTSD-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, 53, 267–291.Vach, W. (1994).

van der Laan M. & Robins, J. (2003). *United methods for censored longitudinal data and causality*. Springer-Verlag.

van der Laan M. & Rose, S. (2011). *Targeted learning: Causal inference for observational and experimental data*. Springer-Verlag.

van der Laan M. & Rubin, D. (2006). Targeted maximum likelihood learning. The International Journal of Biostatistics, 2, Article 11.

van Doorn, J., Verhoef, P.C. & Bijmolt, T. (2007). The importance of non-linear relationships between attitude and behaviour in policy research. *Journal of Consumer Policy*, 30, 75–90.

VanderWeele, T.J.(2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*,20, 18–26.

VanderWeele, T. (2011). Principal stratification—uses and limitations. The International *Journal of Biostatistics*, 7, 1–14.

VanderWeele, T. (2012). Comments: Should principal stratification be used to study mediational processes? *Journal of Research on Educational Effectiveness*, 5, 245–249.

VanderWeele, T.J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.

VanderWeele, T.J. (2016). Mediation analysis: A practitioner's guide. Annual Review of Public Health, 37, 17-32.

VanderWeele, T. (2019). Principles of confounder selection. *European Journal of Epidemiology*.

VanderWeele, T., Zhang, Y. & Limb, P. (2016). Mediation analysis with an ordinal outcome. *Epidemiology*, 27, 651–655.

VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, 25, 300.

van Kesteren, E.& Oberski, D. (2019) Exploratory mediation analysis with many potential mediators. *Structural Equation Modeling*, 26, 710-723.

van Laar, S. & Braeken, J. (2021). Understanding the Comparative Fit Index: It's all about the base!, *Practical Assessment, Research & Evaluation*, 26, 1-23.

Vansteelandt, S., & Daniel, R. M. (2017). Interventional effects for mediation analysis with multiple mediators. *Epidemiology*, 28, 258-265.

Vansteelandt, S. & Keiding, N. (2011) Invited commentary: G-computation - Lost in translation?, *American Journal of Epidemiology*, 173, 739-742.

van t'Veer, A. & Giner-Sorolla, R. (2016). Pre-registration in social psychology: A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2-12.

Vargha, A. & Delaney, H. (2000). A critique and improvement of the "CL" common language effect size statistics of McGraw and Wong, *Journal of Educational and Behavioral Statistics*, 25, 101-132.

Vierron, E. and Giraudeau, B. (2009). Design effect in multicenter studies: Gain or loss in power? *BMC Medical Research Methodology*, 9, 39.

Vig, J., Gehrmann, S., Belinkov, Y, et al.Causal (2021). Mediation analysis for interpreting neural NLP: The case of gender bias. Downloaded from https://arxiv.org/pdf/2004.12265.pdf.

Vilagut, G., Forero, C.G., Barbaglia, G. & Alonso, J. J. (2016). Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): A systematic review with meta-analysis. *PLoS ONE*, 11, 1-17.

Vingerhoets, A. (2013). *Why only humans weep: Unravelling the mysteries of tears*. Oxford University Press.

Vinokur, A., Price, R. & Shul, Y. (1995). Impact of JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology*, 23, 39–74.

Vittinghoff, E. & McCulloch, C. (2006). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology*, 165, 710-718.

von Hippel, P. (2007). Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, 37, 83-117.

von Hippel, P. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology*, 39, 265–91.

von Hippel, P. (2013). The bias and efficiency of incomplete data estimators in small univariate normal samples. *Sociological Methods and Research*, 42, 531-558.

von Hippel, P. (2015). Linear versus logistic probability models: Which is better and when? Philadelphia: Statistical Horizons. Downloaded from https://statisticalhorizons. com/linear-vs-logistic

von Hippel, P. (2016). New confidence intervals and bias comparisons show that maximum likelihood can beat multiple imputation in small samples. *Structural Equation Modeling*, 23, 422-437.

von Hippel, P. (2017). When can you fit a linear probability model? More often than you think. Philadelphia: Statistical Horizons. Downloaded from https://statisticalhorizons. com/when-can-you-fit.

von Hippel, P. (2018). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods and Research*, 49, 699-718.

von Hippel, P. & Lynch, J. (2013). Efficiency gains from using auxiliary variables in imputation. Downloaded from https://www.researchgate.net/publication/258764140_ Efficiency_Gains_from_Using_Auxiliary_Variables_in_Imputation

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-333.

Wallston, K., Strudler-Wallston, B. & Devellis, R. (1978). Development of the multidimensional health locus of control (MHLC) scales. *Health Education Monographs*, 6, 160–170.

Walter, S. & Tiemeier, H. (2009). Variable selection: current practice in epidemiological studies. *European Journal of Epidemiology*, 24, 733–736.

Wang, B., Ogburn, E., & Rosenblum, M. (2019). Analysis of covariance in randomized trials: More precision and valid confidence intervals without model assumptions. *Biometrics*, 75, 1391-1400.

Wang, K. (2018). Understanding power anomalies in mediation analysis. *Psychometika*, 83, 387-406.

Wang, S. & Hu, H. (2022). Impute the missing data using retrieved dropouts. *BMC Medical Research: Methodology*, 22, 82-97.

Wang, X. (2016). Bayesian relative importance analysis of logistic regression models. *Journal of Statistics Applications & Probability Letters*, 3, 53-69.

Wang, Y & Bellemare, M. (2020). Lagged variables as instruments. Downloaded from https://marcfbellemare.com/wordpress/wp-content/uploads/2020/09/WangBellemare LagIVsJuly2020.pdf

Wasserman, L. (1997). Bavesian model selection and model averaging (Working Paper No. 666). Carnegie Mellon University, Department of Statistics.

Wasserstein, R. & Lazar, N. (2016): The ASA's statement on p-values: context, process, and purpose. *American Statistician*, 70, 129-133.

Wasserstein, R., Schirm, A. & Lazar, N. (2019). Statistical inference in the 21st century: A world beyond p < 0.05. *American Statistician*, 73, Supplement 1.

Weed, D. (2016). Commentary: Causal inference in epidemiology: Potential outcomes, pluralism and peer review. *International Journal of Epidemiology*, 45, 1838-1840.

Weisberg, S. (2013). *Applied linear regression*. Wiley

Weissman, M., Sholomskas, D., & Pottenger, M. (1977). Assessing depressive symptoms in five psychiatric populations. *American Journal of Epidemiology*, 106, 203-214.

Welch, C., Petersen, I., Bartlett, J., White, I., Marston, L., Morris, R., et al. (2014). Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Statistics in Medicine*, 33, 3725–3737.

Wen, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as a mediation effect size measure. *Psychological Methods*, 20, 193–203.

Wentz, S. (2018). What quantile regression does and doesn't do: A commentary on Petscher and Logan (2014). *Child Development*, 90, 1442-1452.

West, S, Wu, W., McNeish, D., & Savord, A. (2023). Model fit in structural equation modeling. In R Hoyle (Ed.), *Handbook of structural equation modeling* (2nd Ed.) Guilford, pp.184-205.

White, I.R., Royston, P. & Wood, A.M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30, 377-399.

Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. Wiley.

Whitehead, J (1993). Sample size calculations for ordered categorical data. *Statistics in Medicine*, 12, 2257-2271.

Whittaker, T. (2012). Using the modification index and standardized expected parameter change for model modification. *Journal of Experimental Education*, 80, 26–44.

Wicklin, R. (2023). Bootstrap confidence intervals for the predicted mean in a regression model. At https://blogs.sas.com/content/iml/2023/08/14/bootstrap-ci-predicted-mean-reg.html

Widaman, K., Little, T., Preacher, K. & Sawalani, G. (2011). On creating and using short forms of scales in secondary research. (pp. 39-61.) In K Trzesniewski, M. Donnellan, and R, Lucas (Eds.) *Secondary data analysis: An introduction for psychologists*. American Psychological Association.

Widaman, K. & Revelle, W. (2022). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, 55, 788-806.

Widaman, K. & Revelle, W. (2023). Thinking about sum scores yet again, maybe the last time, we don't know, oh no . . .: A comment on McNeish (2023). *Educational and Psychological Measurement*, 1-23.

Wilcox, R. (1998). How many discoveries have been lost by ignoring modern statistical methods. *American Psychologist*, 53, 300-314.

Wilcox, R.R. (2003). *Applying contemporary statistical techniques*. Academic Press.

Wilcox, R. (2012). *Modern statistics for the social and behavioral sciences: An introduction*. CRC Press.

Wilcox, R. (2017). *Introduction to robust estimation and hypothesis testing*. Academic Press (Fourth edition).

Wilcox, R. (2017). The running interval smoother: A confidence band having some specified simultaneous probability coverage. *International Journal of Statistics: Advances in Theory and Applications*, 1, 21-43.

Wilcox, R. (2021). *Introduction to robust estimation and hypothesis testing*. Academic Press (Fifth edition).

Wilcox, R., Charlin, V. L., & Thompson, K. (1986). New Monte Carlo results on the robustness of the ANOVA F, W, and F* statistics. *Communications in Statistics, Simulation and Computation*, 15, 933–944.

Wilcox, R. & Rousselet, G. (2018). A guide to robust statistical methods in neuroscience. *Current Protocols in Neuroscience*, 82, 1-40.

Wilcox, R. & Tian, T. (2011). Measuring effect size: A robust heteroscedastic approach for two or more groups. *Journal of Applied Statistics,* 38, 1359-1368.

Wilkinson, L., & the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604.

Williams, J. & MacKinnon, D. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling*, 15, 23–51.

Williams, R. (2009) Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods and Research*, 37, 531–559.

Williams, R. (2010). Fitting heterogeneous choice models with oglm. *Stata Journal*, 10, 540–567.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The STATA Journal*, 12, 308-321.

Williams, R. (2012). Ordered logit models: Overview. Downloaded from http://nd.edu/~rwilliam/stats3/L11.pdf.

Williamson, B., Gilbert, P., Carone1, M. & Simon, N. (2021). Nonparametric variable importance assessment using machine learning techniques. *Biometrics*, 77, 9–22.

Willis, G. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. SAGE.

Willitt, J. Singer, J. & Martin, N. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical and methodological recommendations. *Development and Psychopathology*, 10, 395-426.

Winship, C. & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods and Research*, 23, 230-257.

Wodtke, G. & Zhou, X. (2020). Effect decomposition in the presence of treatment-induced confounding: A regression-with-residuals approach. *Epidemiology,* 31, 369-375.

Wolf, M. G. & McNeish, D. (2021). Dynamic Model Fit (version 1.1.0). [Software].www.dynamicfit.app

Wolf, M.G. & McNeish, D. (2023). dynamic: An R package for deriving dynamic fit index cutoffs for factor analysis. *Multivariate Behavioral Research*, 58, 189-194.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Wright, J. (2003). Interpreting health-related quality of life scores: The simple rule of seven may not be so simple. *Medical Care*, 41, 597-598.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161–215.

Yang, S. & Kim, J. (2016). Fractional imputation in survey sampling: A comparative review. *Statistical Science*, 31, 415-432.

Yaremych, H., Preacher, K., & Hedeker, D. (2023). Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychological Methods*, 28, 613–630. (advance online published in 2021).

Yau, L. & Little, R. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, 96, 1232–1244.

Ye F. & Daniel L. (2017). The impact of inappropriate modeling of cross-classified data structures on random slope models. *Journal of Modern Applied Statistical Methods*, 16, 458-484.

Yuan, Y. & Little, R. (2009). Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics*, 65, 478–486.

Yuan, K. H., Tian, Y., & Yanagihara, H. (2015). Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. *Psychometrika*, 80, 379-405.

Yuan, K. H., Yang, M., & Jiang, G. (2017). Empirically corrected rescaled statistics for SEM with small N and large p. *Multivariate Behavioral Research*, 52, 673-698.

Yuan, K. Tian, Y., & Yanagihara, H. (2015). Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. *Psychometrika*, 80, 379–405.

Yuan, K. & Zhong, X. (2013). Robustness of fit indices to outliers and leverage observations in structural equation modeling. *Psychological Methods*, 18, 121–136.

Yuan, Y. & MacKinnon, D. (2014). Robust mediation analysis based on median regression. *Psychological Methods*, 19, 1–20.

Yuan, K., Yang-Wallentin, F. & Bentler, P. (2012).ML versus MI for missing data with violation of distribution conditions. *Sociological Methods and Research*, 41, 598–629.

Yucel, R. M., He, Y., & Zaslavsky, A. (2008). Using calibration to improve rounding in imputation. *American Statistician*, 62, 1–5.

Yung, Y. F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G, Marcoulides and R. Schumacker (Eds.). *Advanced structural equation modeling: Issues and techniques*, pp. 195-226, Hillsdale, NJ: Erlbaum.

Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology*, 115, 929–943.

Zhang, Q & Wang, L. (2017). Moderation analysis with missing data in the predictors. *Psychological Methods*, 22, 649-666.

Zhang, Q., Yuan, K. & Wang, L. (2019). Asymptotic bias of normal-distribution-based maximum likelihood estimates of moderation effects with data missing at random. *British Journal of Mathematical and Statistical Psychology*, 72, 334-354.

Zhang, X. & Savalei,V. (2020) Examining the effect of missing data on RMSEA and CFI under normal theory full-information maximum likelihood. *Structural Equation Modeling*, 27, 219-239.

Zhang, X., & Savalei, V. (2022). New computations for RMSEA and CFI following FIML and TS estimation with missing data. *Psychological Methods*, In press.

Zhang, Z., Gerstein, D. & Friedman, P. (2008). Patient satisfaction and sustained outcomes of drug abuse treatment. *Journal of Health Psychology*, 13, 388–400.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37, 197–206.

Zhou, H., Elliott, M. & Raghunathan, T. (2016) Synthetic multiple-imputation procedure for multistage complex samples. Journal of Official Statistics, 32, 231–256.

Zitzmann, S. Helm, C. & Hecht, M. (2021). Prior specification for more stable Bayesian estimation of multilevel latent variable models in small samples: A comparative investigation
of two different approaches, *Frontiers in Psyc*hology, 11.

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, 14, 305–320.

Zou G.Y. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*, 159, 702-706.

Zucker, D. M., Lieberman, O., & Manor, O. (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society: Series B*, 62, 827–838.