OXFORD

Validity of Data Collected from Randomized Behavioral Clinical Trials During the COVID-19 Pandemic

Constance A. Mara, PʜD and James L. Peugh, PʜD

Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine

All correspondence concerning this article should be addressed to Constance A. Mara, PʜD, Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, 3333 Burnet Ave., MLC 7039, Cincinnati, OH 45229, USA. E-mail: constance.mara@cchmc.org

The COVID-19 pandemic and efforts to mitigate the spread and impact of the virus have drastically altered day-to-day operations of research trials. In the United States, widespread "lockdowns" or "stay at home" orders have halted research operations at many institutions. In some cases, institutions have halted in-person visits that are not of medical necessity. This has left many randomized behavioral clinical trials in unprecedented positions.

McDermott & Newman (2020) recently discussed maintaining the integrity of trials during this pandemic, providing practical suggestions to minimize the disruption that the current pandemic has on ongoing trials. For example, they suggest adapting study protocols for trials that are already ongoing by prioritizing collection of primary outcomes, exploring alternate measures, and transitioning to mail-in or other remote methods of outcome collection. They also provide guidance on retaining study staff and the recruitment and engagement of study participants remotely. Guidelines from the FDA (2020) have also been released to support investigators running trials at this time, emphasizing participants' safety. These recommendations suggest examining whether it might be appropriate to delay assessments for ongoing trials or modify intervention delivery modalities. These guidelines recognize that delays in assessments, halting ongoing recruitment, and withdrawing participants from the trial may be necessary depending on the nature of the trial, the intervention, the timing of the disruption, and safety concerns. Investigators are encouraged to

consult with institutional review boards and study sponsors to discuss anticipated changes in protocols and document in detail all protocol deviations and missing data information. However, no studies to date have specifically addressed threats to the validity of the data collected during the pandemic, nor how modifications to study protocols to maintain study integrity could pose significant threats to the validity of the data collected from these trials.

Randomized behavioral clinical trials will face unique challenges during this time. The COVID-19 pandemic is a textbook example of a "history effect," which can pose a substantial threat to study validity (in-depth discussions on study validity can be found throughout the literature, e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979). A history effect refers to events that occur in the environment that can change the conditions of a study, affecting its outcome. History effects have the potential to differentially change how individuals and study groups respond to treatments or interventions. How this history effect impacts a particular study will depend on the timing of the pandemic onset within a specific study timeline, the overall length of the study, and the magnitude of the modifications and alterations needed to preserve the study. Indeed, the modifications and alterations implemented to preserve trials during this time may inadvertently compound the history effect by introducing additional threats to data validity, such as maturation (e.g., pediatric participants experiencing significant developmental maturing between a lag in

assessment points due to study delays or suspension during lockdowns), differential drop out (e.g., families not able to access remote interventions due to no internet access or unable to continue in-person assessments because they are medically high risk), selection bias (e.g., underserved or high-risk participants excluded from trial recruitment), etc. (Campbell & Stanley, 1963; Cook & Campbell, 1979). Investigators must carefully consider how the internal and external validity of their study may be impacted while operating during this pandemic. Our specific goals in this article are to outline areas where necessary protocol changes in response to the pandemic could threaten the internal or external validity of the study and provide practical solutions that can help mitigate these threats to validity. Many of the suggestions provided below are our opinions as quantitative psychologists in the field of pediatric psychology and how we are advising the 20+ pediatric psychology research labs with extramural funding that we collaborate with.

Efforts to minimize the pandemic's impact on both protocol deviations and data validity can be employed in three areas: (1) recruitment protocol changes, (2) intervention delivery changes, and/or (3) changes to collection of outcomes and other data. Changes in each of these areas pose differing levels of threat to internal validity and external validity.

1. Changes to recruitment: Changes to recruitment procedures are less likely to impact the validity of the data, assuming intervention and data collection procedures are unchanged. If a study recruited participants in person but delivered the intervention at home, via telehealth, app-based interventions, or other remote options, and outcomes were collected online, only the in-person recruitment would need to be modified to maintain study activities with little impact to data validity. However, one may need to consider selection bias in this case: Do the individuals who consent to participate during the lockdown differ systematically from those who do not? Is it possible that the pool of eligible participants that can be recruited remotely differs from the eligible participant pool when recruiting in person?

2. Changes to intervention delivery: Any modification to the intervention protocol poses a non-negligible threat to the internal validity of the study. Changes to intervention procedures and protocols likely mean pre-pandemic participants received an intervention that may not be comparable to those who received a modified intervention during the pandemic. However, this is highly contingent on the degree of modifications needed. There is an extant empirical research evaluating the equivalence of remote versus face-to-face interventions. In many cases, they have not been found to differ significantly. For example, Wood et al. (2016) found that in-person and telemedicine visits were equally effective at improving adherence to diabetes management regimens in pediatric Type 1 diabetes patients. However, there are times when this is not true. For example, a review by Kuster et al. (2017) examined computer-based versus in-person intervention for stress reduction. They found mixed results in their review, with some studies showing no difference between in-person and virtual interventions, while some studies showed superiority of the in-person interventions. Another review of in-person versus remote versions of interventions to reduce alcohol consumption in college students found that in-person modes were generally superior to computer-based modes (Wagner et al., 2014). In some instances, the remote version of the interventions has shown improved outcomes over in-person modes. For example, one study included in this review found that the short-term efficacy of internet-based and face-to-face cognitive behavioral therapy for depression were equivalent, but the longer-term treatment effect was only maintained by the online group (Wagner et al., 2014). Wade et al. (2020) summarized findings on the efficacy of telepsychotherapy for children and families from 14 clinical trials and report that telepsychotherapy resulted in greater therapeutic alliance, satisfaction, and convenience when directly compared to face-to-face approaches. In summary, researchers cannot assume the equivalence of intervention delivery modes without explicitly examining the efficacy of each mode of delivery specific to a particular intervention.

3. Changes to the collection of outcomes: Changing how outcomes and assessments are collected may pose a threat to the internal validity of the data. In some cases, there is simply a lack of evidence for the validity of a measure in different modes (e.g., a measure may have been validated for pen-and-paper or online administration, but there is no available validity evidence for telephone administration). Changing the mode of data collection (e.g., in person switched to online or via telephone) can change how people respond to outcomes (e.g., measurement mode effect; Hox et al., 2015). Researchers cannot assume that all modes or forms of an assessment are equivalent from a psychometrics perspective. Measurement equivalence (also called invariance) means that groups can be compared on their mean scores because the questionnaire measures identical constructs with the same structure across all groups. If this is not established, group means cannot be compared, as groups respond differently to the items (Mellenbergh, 1989; Meredith, 1993; Millsap & Meredith, 2007). In some cases, measurement invariance has been shown across delivery modes (e.g., Varni et al., 2009), but not in others (e.g., Magnus et al., 2016). The presence of study personnel while outcomes are being collected in clinic versus at home answering questions alone can result in participants responding differently due to social desirability or satisficing (satisficing occurs when respondents provide quick, "good enough" answers rather than carefully considered answers, Coyne et al., 2005; Fang et al., 2014; Hamby & Taylor, 2016; Link & Mokdad, 2005). Satisficing, in particular, can lead to measurement non-invariance across assessment modes, as individuals who satisfice have low motivation and use suboptimal response strategies compared to those who are not satisficing (Barge & Gehlbach, 2012; Kaminska et al., 2010; Krosnick, 1991). For any particular study, evidence of measurement invariance must be shown when collecting mixed modes of outcome measures within the same trial (Coons et al., 2009).

In addition to protocol changes that can impact data validity, investigators must also consider the impact that this pandemic has had on children and their families during this time. For example, there is a loss of physical contact between children and their larger support network of extended family, teachers, and friends during lockdowns. There is increased stress within the household overall, and with school closures, a lack of structure in day-to-day routines. Rosenthal & Thompson (2020) highlight increased risk factors for child abuse during the pandemic because of a loss of connection to, and isolation from, those in their larger support network. The authors further note that access to mental health services for children and their caregivers may be limited. A study of Chinese children during the lockdowns found higher rates of depressive symptoms and anxiety symptoms compared with previous prevalence studies (Xie et al., 2020). This study also noted that serious infectious diseases have been shown to negatively impact children's mental health and well-being. Particularly for behavioral interventions and behavioral outcomes in pediatrics, it is possible that how children and their caregivers respond to treatment and to measures has changed as the context in which they are living has changed, which could pose threats to the internal and external validity of the study data. These threats to validity are more difficult to assess and concerns the ecological validity of the results gathered from studies during this time (e.g., are trial results from data collected during a pandemic generalizable to a world where a pandemic is not occurring?).

Given what little investigators can do to "control" the history effect we are currently experiencing, recommendations on how to minimize threats to the internal and external validity of ongoing trials within the context of this pandemic history effect are suggested. Suspension of study activities, delays to study start, and changes to mode of intervention or delivery during pandemic lockdowns has been advised in some cases (FDA, 2020; Fleming et al., 2020; McDermott & Newman, 2020) because the study population is not available during the pandemic lockdowns (e.g., high-risk populations, school closures, or institutional restrictions on recruitment), or because the intervention must be delivered in person. In many cases, study suspension may be a forced choice, as many institutions have halted research not related to COVID-19 in the wake of widespread lockdowns. From a validity perspective, suspending studies may be preferred to intervention modification given that modifying an intervention mid-way through a study poses a large threat to internal validity. Specifically, as discussed previously, intervention delivery modes cannot be assumed equivalent. If the protocol or intervention must change in response to the pandemic and statewide containment efforts, investigators may face the reality that the intervention delivered during lockdown is not equivalent to that delivered prior to the pandemic, or post-lockdown. For example, with school-based interventions and group-based therapies, high fidelity may be impossible to achieve when pivoted to remote delivery. In these cases, validity of the trial is best preserved when those interventions can be delivered safely once again as originally intended. Conversely, delays in study timelines could threaten data validity. For individuals who were "mid-intervention" or had not finished all assessment time points, their data need to be flagged and evaluated for inclusion in the overall analysis sample via sensitivity analyses.

To assess the impact of study delay, suspension, or intervention modification on study outcomes, all studies should record and code each observation as occurring pre-pandemic, during lockdown (if research activities are continuing), and post-lockdown to explore subsets of the sample at different stages of the pandemic and control for this variable to some degree in future analyses. If studies continue through the "end" of the pandemic, investigators may need to consider adding this classification to data points collected at that point as well. There may also be instances where a second lockdown occurs for an institution/state over the course of the coming year and a "second lockdown" classification might be needed. For some studies, "during lockdown" may be sufficient to classify data occurring during any lockdowns, as the protocol modifications will be identical each lockdown. The definition and timing of these "classifications" should be discussed and agreed upon by study collaborators. For example, some investigators have coded any participants and timepoints occurring prior to March 16, 2020, as "pre-pandemic" and those occurring on or after March 16, 2020, as "during lockdown." The "post-lockdown" classification will depend on individual state and institutional "re-opening" timelines. Intervention modes can be compared by continuing to deliver the intervention per protocol if possible after lockdowns as well as the modified mode and comparing arms for evidence of equivalent efficacy. Additionally, any data collected from the trial that used different intervention delivery modes can serve as pilot data that informs a future trial by offering key information about dosage, format, timing, and content of interventions. If the mode of outcome measurement was modified, using what data are available to provide some evidence of the invariance of the modes may be possible. Additional guidance on pre-specifying statistical analyses to account for COVID-19 pandemic-related trial disruptions is discussed in Fleming et al. (2020).

It will also be critical to document missing data information for all assessments and participants and

note that those lost to follow-up under normal circumstances may vary from those lost to follow-up due to pandemic effects. Maximum likelihood (ML) estimation and multiple imputation (MI) are the only empirically acceptable missing data handling mechanisms, and both assume missing at random (MAR) by default. Defined, MAR assumes that variables that could explain why missing data are missing (i.e., auxiliary correlate variables; Graham, 2003) are included in the data analysis model so that parameter estimates of interest remain unbiased (e.g., see Holbein et al., 2019). The literature is clear regarding the conditions under which MI is clearly favored over ML estimation (Enders et al., 2018, 2019), otherwise, ML and MI are assumed to be asymptotically equivalent. Furthermore, knowing that acceptable auxiliary correlate variables are difficult to identify for ML missing data handling and that MI is far more flexible and accommodating of a wide range of possible auxiliary correlate variables (Enders et al., 2018, 2019), we recommend researchers consider using MI to address missing data for studies conducted during COVID-19. Researchers should include measures of COVID-19's overall impact on pediatric participants as well as specific impact measures (social isolation, depression, stress, anxiety, frustration) that could explain missing data (discussed below) in their data collection protocols for their MI models.

To further understand how the pandemic may impact trial results, if feasible, we recommend adding a COVID-19 impact questionnaire (e.g., https://disasterinfo.nlm.nih.gov/content/files/Epidemic-Pandemic-Impacts-Inventory-FINAL.pdf or https://www.nlm.nih.gov/dr2/CEFIS_COVID_questionnaire_English_42220_final.pdf) in addition to psychometrically sound measures of distress, anxiety, depression, well-being, and/or support to understand the contextual impact of the pandemic on children and their families which will provide information to help researchers assess and mitigate threats to internal and external validity. Specifically, researchers can use this additional data: (a) to measure the degree to which the pandemic has impacted participants' responses to treatment in the trial, (b) to include as auxiliary correlates to help make the default MAR assumption more plausible when handling missing data, (c) to assess possible threats to the external validity of the study, (d) for sensitivity analyses (e.g., comparing treatment responses among those families experiencing high versus low levels of distress), (e) for testing mediation or moderation mechanisms, and (f) to provide valuable information to examine differential impact of the pandemic on study outcomes for high-risk or underserved groups. In particular, low-income children already face economic, educational, health, and social disparities. The pandemic has only magnified these

inequalities through lack of access to normal educational, nutritional, and social resources (Dooley et al., 2020). Understanding the consequences that this pandemic will have on these children, particularly those with chronic health conditions, will be critical in advocating for increased community support and resources for their health and well-being both in the short and long term.

Overall, researchers conducting randomized clinical behavioral trials must carefully consider how protocol changes made to preserve their trials may impact the conclusions that can be made at the end of the study. While it is important to consider the logistical and safety issues of continuing trials, it is just as important to ensure that the data collected during and after the pandemic are valid and useful. It is becoming clear with each passing week that the pandemic will continue for several more months or longer. Perhaps, things will never return to "normal," and researchers will find efficiency, accessibility, and flexibility in a new normal. Behavioral trials during this time have a unique opportunity to understand the impact that the pandemic has on children and their families, particularly those who are medically high risk or face social and economic disadvantages. The full impact of this historical time on research will likely be best understood in hindsight, but researchers can take steps now to gather information in their ongoing trials to understand the short- and long-term consequences of this pandemic. Even while making decisions "on the fly" in the midst of these uncertain and ever-changing conditions, researchers conducting randomized behavioral clinical trials must attempt to maintain the integrity of current trials. Often, the logistics of research operations and fiscal concerns may be at odds with the validity of the study. For example, changing an adolescent group-based exercise intervention to an online-delivery changes many core components of the intervention itself: the ability to make hands-on adjustments, peer interaction and comradery with the group, etc. However, pausing this study may not be feasible due to lack of needed additional funding to extend the study timeline. Investigators could make the choice to compromise the validity of the study to transition to an online intervention to keep study running. In this scenario, collecting response variable data via parallel intervention modes would be helpful in conducting sensitivity analyses and comparing treatment efficacy across the different modalities (in-person group-based versus online-delivery). In general, investigators need to carefully weigh their decisions and, when necessary protocol changes must be made that do threaten validity, these limitations need to be fully disclosed. By incorporating the recommendations above, investigators can mitigate many of the threats to the validity of the data and evaluate the

impact of the pandemic on their study population and trial outcomes.

Conflicts of interest: None declared.

References

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluated the quality of survey data. *Research in Higher Education*, 53, 182–200.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand McNally.

Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., Lenderking, W. R., Cella, D., & Basch, E.; on behalf of the ISPOR ePRO Task Force. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value in Health*, 12, 419–429.

Coyne, I., Warszta, T., Beadle, S., & Sheehan, N. (2005). The impact of mode of administration on the equivalence of a test battery: A quasi-experimental design. *International Journal of Selection and Assessment*, 13, 220–224.

Dooley, D. G., Bandealy, A., & Tschudy, M. M. (2020). Low-income children and the coronavirus disease 2019 (COVID-19) in the US. *JAMA Pediatrics*. [epub]. https://doi.org/10.1001/jamapediatrics.2020.2065

Enders, C. K., Du, H., & Keller, B. T. (2019). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods, 25, 88-112.* https://doi.org/10.1037/met0000228

Enders, C. K., Hayes, T., & Du, H. (2018). A comparison of multilevel imputation schemes for random coefficient models: Fully conditional specification and joint model imputation with random covariance matrices. *Multivariate Behavioral Research*, 53, 695–713.

Fang, J., Wen, C., & Prybutok, V. (2014). The assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. *Computers in Human Behavior*, 30, 335–343.

FDA (2020). *FDA guidance on conduct of clinical trials of medical products during COVID-19 pandemic: Guidance for industry, investigators, and institutional review boards*. US Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/fda-guidance-conduct-clinical-trials-medical-products-during-covid-19-pandemic Retrieved 8 April 2020.

Fleming, T. R., Labriola, D., & Wittes, J. (2020). Conducting clinical research during the COVID-19 pandemic: Protecting scientific integrity. *Journal of the American Medical Association*. [epub]. https://doi.org/10.1001/jama.2020.9286

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, 10, 80–100.

Hamby, T., & Taylor, W. (2016). Survey satisficing inflates reliability and validity measures. *Educational and Psychological Measurement*, 76, 912–932.

Holbein, C. E., Smith, A. W., Peugh, J., & Modi, A. C. (2019). Allocation of treatment responsibility in adolescents with epilepsy: Associations with cognitive skills and medication adherence. *Journal of Pediatric Psychology*, 44, 72–83.

Hox, J. J., De Leeuw, E. D., & Zijlmans, E. A. (2015). Measurement equivalence in mixed mode surveys. *Frontiers of Psychology*, 6, 1–11.

Kaminska, A., McCutcheon, A., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74, 956–984.

Krosnick, J. A. (1991). Response strategies for coping with cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.

Kuster, A. T., Dalsbø, T. K., Luong Thanh, B. Y., Agarwal, A., Durand-Moreau, Q. V., & Kirkehei, I. (2017). Computer-based versus in-person interventions for preventing and reducing stress in workers. *Cochrane Database Systematic Reviews*, 8, CD011899. https://doi.org/10.1002/14651858.CD011899.pub2

Link, M. W., & Mokdad, A. H. (2005). Effects of survey mode on self-reports of adult alcohol consumption. A comparison of mail, web, and telephone approached. *Journal of Studies on Alcohol*, 66, 239–245.

Magnus, B. E., Liu, Y., He, J., Quinn, H., Thissen, D., Gross, H. E., DeWalt, D. A., & Reeve, B. B. (2016). Mode effects between computer self-administrations and telephone interviewer-administrations of the PROMIS pediatric measures, self- and proxy report. *Quality of Life Research*, 25, 1655–1665.

McDermott, M. M., & Newman, A. B. (2020). Preserving clinical trial integrity during the coronavirus pandemic. *Journal of the American Medical Association*.[epub]. https://doi.org/10.1001/jama.2020.4689

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Education Research*, 13, 127–143.

Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika*, 58, 525–543.

Millsap, R. E., & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. In R. Cudeck and R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 131–152). Erlbaum.

Rosenthal, C. M., & Thompson, L. A. (2020). Child abuse awareness month during the coronavirus disease 2019 pandemic. *JAMA Pediatrics*, 174, 812. https://doi.org/10.1001/jamapediatrics.2020.1459

Varni, J. W., Limbers, C. A., & Newman, D. A. (2009). Using factor analysis to confirm the validity of children's self-reported health-related quality of life across different modes of administration. *Clinical Trials*, 6, 185–195.

Wade, S. L., Gies, L. M., Fisher, A. P., Moscato, E. L., Adlam, A. R., Bardoni, A., Corti, C., Limond, J., Modi, A. C., & Williams, T. (2020). Telepsychotherapy with children and families: Lessons gleaned from two decades of

translational research. *Journal of Psychotherapy Integration*, *30*, 332–347.

Wagner, B., Horn, A. B., & Maercker, A. (2014). Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial. *Journal of Affective Disorders*, *152–154*, 113–121.

Wood, C. L., Clements, S. A., McFann, K., Slover, R., Thomas, J. F., & Wadwa, R. P. (2016). Use of telemedicine to improve adherence to American Diabetes Association standards in pediatric type 1 diabetes. *Diabetes Technology & Therapeutics*, *18*, 7–14.

Xie, X., Xue, Q., Zhou, Y., Zhu, K., Lui, Q., Zhang, J., & Song, R. (2020). Mental health status among children in home confinement during the coronavirus disease 2019 outbreak in Hubei Province, China. *JAMA Pediatrics*. [epub]. https://doi.org/10.1001/jamapediatrics.2020.1619