Epidemiologic Reviews
© The Author(s) 2022. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 44, 2022
https://doi.org/10.1093/epirev/mxac006
Advance Access publication:
September 14, 2022

Review

Econometric Issues in Prospective Economic Evaluations Alongside Clinical Trials: Combining the Nonparametric Bootstrap With Methods That Address Missing Data

Ali Jalali*, Rulla M. Tamimi, Sterling M. McPherson, and Sean M. Murphy

* Correspondence to Dr. Ali Jalali, Department of Population Health Sciences, Weill Cornell Medicine, 425 East 61st Street, Suite 301, New York, NY 10065 (e-mail: alj4004@med.cornell.edu).

Accepted for publication September 7, 2022.

Prospective economic evaluations conducted alongside clinical trials have become an increasingly popular approach in evaluating the cost-effectiveness of a public health initiative or treatment intervention. These types of economic studies provide improved internal validity and accuracy of cost and effectiveness estimates of health interventions and, compared with simulation or decision-analytic models, have the advantage of jointly observing health and economics outcomes of trial participants. However, missing data due to incomplete response or patient attrition, and sampling uncertainty are common concerns in econometric analysis of clinical trials. Missing data are a particular problem for comparative effectiveness trials of substance use disorder interventions. Multiple imputation and inverse probability weighting are 2 widely recommended methods to address missing data bias, and the nonparametric bootstrap is recommended to address uncertainty in predicted mean cost and effectiveness between trial interventions. Although these methods have been studied extensively by themselves, little is known about how to appropriately combine them and about the potential pitfalls and advantages of different approaches. We provide a review of statistical methods used in 29 economic evaluations of substance use disorder intervention identified from 4 published systematic reviews and a targeted search of the literature. We evaluate how each study addressed missing data bias, whether the recommended nonparametric bootstrap was used, how these 2 methods were combined, and conclude with recommendations for future research.

missing data mechanism; models, econometric; multiple imputation; nonparametric bootstrap; substance-related disorders; inverse probability weighting

Abbreviations: HRQoL, health-related quality of life; IPW, inverse probability weight; MAR, missing at random; MI, multiple imputation; MNAR, missing not at random; SUD, substance use disorder.

INTRODUCTION

Economic evaluations of treatment interventions or health programs provide necessary information regarding the comparative use of scarce health care and societal resources to maximize public health while limiting the growth of health expenditures when possible. An increasingly popular approach to conducting economic evaluations is to assess the cost and effectiveness of competing health interventions or treatment modalities alongside a randomized controlled trial (1).Trial-based economic evaluations prospectively collect participant-level health economic data at multiple time points throughout the study, such as the participant's health care resource utilization, time commitment for receiving care (including travel time), use of tangible and intangible societal resources, and their health-related quality-of-life (HRQoL) for purposes of generating quality-adjusted life years. These types of economic analyses provide multiple advantages for researchers. They generate data from validated trial instruments instead of relying solely on electronic health records or insurance claims, which, although less subject to recall bias (albeit small for recall periods typically used in trials) (2–6), rarely provide a comprehensive record of health care resource utilization. The collection of detailed, baseline, participant-level data allows for the use of robust methods for addressing missingness and sampling

uncertainty, as determined by the missing data mechanism. Furthermore, participant-level data are linked across measures and, therefore, researchers have the advantage of jointly observing health and economic outcomes of trial participants over time, as compared with decision-analytic models that typically rely on publicly available epidemiologic data from the literature; they also provide improved internal validity and accuracy of cost and effectiveness estimates, including the uncertainty around those estimates.

Regression analysis is an important tool for clinical trial researchers to estimate the causal relationship between intervention and outcome variables longitudinally while controlling for potential confounding factors that may remain after randomization or when intent-to-treat analyses are not appropriate. Missing data due to loss to follow-up, failure to initiate treatment, and participant attrition or nonresponse to a survey item are common challenges of clinical trials and limit an intent-to-treat study design (7, 8). Missing data are a particular problem for clinical trials of substance use disorder (SUD) interventions, because participants with chronic SUD face increased stigma (9, 10), barriers to accessing health care services (11, 12), risk of unstable housing (13, 14), and rates of involvement with the criminal legal system (15), all of which contribute to higher rates of item or unit nonresponse and participant attrition. Although trials of treatment interventions for other chronic conditions may face similar issues, we chose to focus on economic evaluations of SUD interventions, given the extent to which these factors intersect among participants with SUD and contribute to missingness in both clinical and nonclinical data.

Multiple imputation (MI) and inverse probability weights (IPWs) are 2 of the most widely recommended regression-based approaches to reduce missing data bias given common causes of missingness (discussed in the next section) while using all available data from all participants (16–21). Upon estimating the adjusted mean cost or effectiveness for each arm, the nonparametric bootstrap is recommended to generate standard errors while accounting for potential sample selection bias (1, 22). Although these methods have been studied extensively by themselves (23, 24), little is known about how to appropriately combine them or about the relative advantages and disadvantages of approaches used in the literature to maintain unbiased results and computational efficiency.

We provide a review of methods used to address missing variable bias of multivariable-adjusted mean values within the nonparametric bootstrap to evaluate uncertainty. We limited our search to published economic evaluations of SUD interventions in the context of a clinical trial and identified studies that reported missing data or patient attrition in the trial. We discuss the methods used in these studies to address missing data bias, whether recommended guidelines were followed to evaluate uncertainty in predicted mean costs and effects between trial interventions using the nonparametric bootstrap, and how these 2 methods were combined in those studies. For example, investigators may impute missing values or calculate IPWs prior to bootstrapped resampling, embed the MI or calculate IPWs for each participant within the bootstrap, or simply sample complete cases and ignore missing data in the bootstrap procedure. We end our discussion with recommendations for future research on expanding the methodologies used in economic evaluations of SUD interventions alongside clinical trials.

Missing data mechanisms

Addressing missing data requires understanding the mechanisms of the missingness, most specifically whether the data are 1) missing completely at random (MCAR); 2) missing at random (MAR); or 3) missing not at random (MNAR) (18, 25). MCAR implies that missing observations are neither related to nor a function of observed factors, such as patient demographics, disease severity, study site, or provider characteristics, or unobserved confounders. In such cases, listwise deletion (i.e., complete case analysis), theoretically, can be justified, because correcting for missingness would not be necessary to obtain unbiased estimates of covariate-adjusted arithmetic means. However, the proportion of missing data will limit the total sample size for the analysis, reducing degrees of freedom, and directly affecting the power to detect a meaningful difference in cost and effectiveness outcomes. Thus, missing data techniques that allow for the inclusion of all observed data across participants, such as MI or IPW, are still preferrable (1, 17, 26, 27).

MAR describes the case where missing observations are assumed to be related to observed or measured variables but are unrelated to unobserved confounding variables. Both MI and IPW perform well in addressing missing data bias under MAR conditions, though MI has been suggested as a more efficient approach (28). In addition, imputation models that generate new data are not necessarily invariant across outcome measures and must generate data that appropriately fit the observed distribution and characteristics of cost (e.g., strictly positive and asymmetric) and effectiveness (e.g., bounded at 1 for health utility data). MNAR is a more serious problem for researchers, because there are no effective methods for removing the bias, given that the missingness is associated with unobserved factors (27, 29).

Resampling with missing data

The nonparametric bootstrap is recommended for evaluating the distribution of predicted means of commonly reported cost-effectiveness measures, such as the incremental cost-effectiveness ratio, and accounts for sampling uncertainty. The nonparametric bootstrap addresses sampling uncertainty through successive resampling of participants in the trial with replacement, stratified by treatment arm and trial selection criteria (e.g., 1:1 sex ratio), generating a new sample and repeatedly calculating the statistics of interest in the economic evaluation to calculate confidence intervals (1). In the presence of missing data, investigators can either resample complete cases only or include incomplete cases in the resampling procedure to avoid potentially biased statistical inference when data are not MCAR. Combining the method used to address missing data with the nonparametric bootstrap in economic evaluations has received

scant attention. Both MI and IPW can be embedded within the bootstrap, or the data can be imputed and propensity scores for the IPW estimated prior to resampling. IPW estimated within the nonparametric bootstrap is commonly used in applied studies but may produce highly variable propensity scores in trials with a small, randomized sample (30). Embedding MI within the nonparametric bootstrap can be computationally intensive, depending on the number of imputations chosen within each iteration of the bootstrap, and there are no standardized methods to ensure that each multiply imputed data set within the bootstrap is valid (31). In the case of MI embedded within the nonparametric bootstrap, N × M estimations are necessary, where N is the number of bootstrap samples, and M is the number of imputations. Brand et al. (32) recommended in a simulation study that single imputation within the nonparametric bootstrap can be both valid and improve computational efficiency compared with the bootstrap resampling nested within MI or MI within the bootstrap, whereas other researchers have recommended the latter methods based on similar simulation study designs (33).

METHODS

We identified studies from 4 systematic reviews of economic evaluations of opioid use disorder (34–36) and SUD interventions (37). In their systematic review, Murphy and Polsky (34) covered the years 2007 to 2015 and built upon a prior systematic review by Doran (36), in which that author studied economic evaluations of opioid use disorder interventions prior to 2007. Onuoha et al. (35) picked up where Murphy and Polsky left off, covering 2015 through 2019. Jalali et al. (37) reviewed economic evaluations of treatments for SUDs and related conditions in the National Drug Abuse Treatment Clinical Trials Network from 2000 to September 2019.

We first identified all studies within the 4 aforementioned systematic reviews that met the following criteria: 1) a randomized controlled trial study design; 2) applied econometric methods in estimating cost and effectiveness data using participant-level data; and 3) reported missing data in the study sample. These articles were then supplemented with a targeted search of the PubMed/Medline and Google Scholar electronic databases to identify economic evaluations meeting the same criteria for SUDs other than opioid use disorder. The targeted literature search included the intersection of the following or related terms: 1) substance-related disorder; 2) cost, cost-effectiveness, or cost-benefit analysis; and 3) missing, imputation, probability weight, propensity score, bootstrap, parametric, or nonparametric. The PubMed and Google Scholar search terms are available online in the Web Appendix (available at https://doi.org/10.1093/aje/mxac006). Partial economic evaluations (e.g., cost-offset studies) that did not evaluate both cost and effectiveness outcomes were excluded.

All articles meeting the inclusion criteria were read to assess the use of statistical methods to address uncertainty; whether missing data or participant attrition in the trial were reported; the method used for addressing missing data, when possible; and the type of resampling used in combination with the method to address missingness. We then provided a detailed discussion of the identified methodologies.

RESULTS

We identified 29 published economic evaluations that met our inclusion criteria. The trial data in these studies were of treatments for SUDs, including cannabis, stimulants, alcohol, opioids (including heroin), and polysubstance, in a variety of settings among both adolescents and adults. All articles addressed missing data using 1 possible method, including 3 that were categorized as complete case analysis by default (because they did not explicitly address missing data). In 23 studies, researchers used the nonparametric bootstrap method recommended to address statistical uncertainty in the point estimates of outcome variables (1). Some form of complete case analysis for at least 1 missing outcome variable was used in 12 of the articles, 3 applied some variation of the last-observation carried forward, linear interpolation was used in 1 study for nonmonotonic missing data, 6 used IPW, 3 applied baseline observation carried forward, and a parametric method (using mean and variance estimates from complete cases to randomly replace missing cases) was used in 1 study. The rest of the studies applied some form of single-imputation method, including regression adjusted, mean imputation, within-group matching, or even applying cost and effect values based on opinion or plausible assumptions of the authors. No study in this review incorporated MI to address missingness in the analysis.

Of those economic evaluations that applied the nonparametric bootstrap, 7 used compete case resampling, 11 followed an imputation-before-resampling approach to combining these methods, 6 did not report the use of resampling, and 5 conducted IPW within (or after) resampling. Table 1 summarizes results of these findings.

DISCUSSION

Complete case analysis

Three studies did not explicitly address missingness in our review (38–40). These studies were early economic evaluations (prior to 2000) and predate some of the computational improvements in resampling and imputation methods in popular statistical software. These studies also used statistical tests of unadjusted mean differences, instead of adjusted costs and effectiveness using regression analysis. For the purpose of our review, they were categorized as applying complete case analysis to address missingness by default.

Complete case analysis, in which only observations with complete data in all relevant variables are used, is the least rigorous approach to addressing missing data and is rarely recommended, given that it will only produce unbiased estimates when data are MCAR (41). However, the potential bias from missingness may drop as the percentage of missing data approaches 0, or the sample size increases (42). Of the 8 studies identified in our review in which authors only used complete case analysis, 4 reported proportions of missing data small enough to potentially justify this approach. Olmstead and Petry (43) reported missing data on health

Table 1. Review of Missing Data and Nonparametric Bootstrap Methods in Economic Evaluations

First Author, Year (Reference No.)	Effectiveness Measure[a]	Nonparametric Bootstrap	Methods Addressing Missing Data[b]	Resampling Method Combination[c]
Kraft, 1997 (40)	Abstinence	0	Complete case analysis	N/A
Avants, 1999 (38)	Abstinence, treatment retention	0	Complete case analysis	N/A
Hartz, 1999 (39)	Abstinence	0	Complete case analysis	N/A
Doran, 2003 (53)	Abstinence	1	Regress to baseline and mean imputation	Imputation before resampling
Doran, 2004 (64)	Abstinence	0	Mean imputation	N/A
Dijkgraaf, 2005 (52)	QALYs	1	Last-observation carried forward	Imputation before resampling
Harris, 2005 (51)	QALYs, abstinence	1	Last-observation carried forward	Imputation before resampling
Doran, 2006 (65)	Abstinence	1	Parametric imputation	Imputation before resampling
Bell, 2007 (47)	Abstinence, treatment retention	1	Complete case analysis	Complete case resampling
Olmstead, 2007 (44)	Abstinence	1	Complete case analysis	Complete case resampling
Sindelar, 2007 (45)	Abstinence	1	Complete case analysis, imputation by assumption	Complete case resampling
Fals-Stewart, 2008 (56)	Abstinence	0	Single imputation, worst-observation carried forward	N/A
Olmstead, 2009 (43)	Abstinence	1	Complete case analysis	Complete case resampling
Polsky, 2010 (68)	QALYs	1	Inverse probability weighting	Probability weights within resampling
Lo Sasso, 2012 (62)	Abstinence, illegal activity	1	Regression-based single imputation	Imputation before resampling
Ruger, 2012 (48)	Abstinence, treatment retention	1	Complete case analysis, imputation by assumption	Complete case resampling
Tran, 2012 (69)	QALYS	0	Inverse probability weighting	N/A
Byford, 2013 (46)	QALYs, abstinence	1	Complete case analysis	Complete case resampling
Murphy, 2015 (70)	QALYs, abstinence	1	Inverse probability weighting	Probability weights within resampling
Drost, 2016 (66)	Abstinence	1	Complete case analysis, subgroup sensitivity analysis	Complete case resampling
Goorden, 2016 (58)	QALYs	1	Complete case analysis, linear interpolation	Imputation before resampling
Murphy, 2016 (71)	QALYs, abstinence	1	Inverse probability weighting	Probability weights within resampling
Busch, 2017 (50)	Abstinence, treatment retention	1	Mean imputation, imputation by assumption	Imputation before resampling
Heslin, 2017 (60)	QALYs	1	Complete case analysis, subgroup mean and regression imputation	Imputation before resampling
Dunlop, 2017 (54)	Abstinence	1	Baseline observation carried forward	Imputation before resampling
Murphy, 2017 (72)	QALYs, abstinence	1	Inverse probability weighting	Probability weights within resampling
Dunlap, 2018 (55)	Abstinence	1	Baseline observation carried forward	Imputation before resampling
Murphy, 2019 (49)	QALYs, abstinence	1	Inverse probability weighting	Probability weights within resampling
Rains, 2019 (61)	Abstinence, adverse events	1	Regression-based single imputation	Imputation before resampling

Abbreviations: N/A, not applicable; QALY, quality-adjusted life year.
[a] Abstinence included longest days of abstinence, time to drug use, and continuous measures of abstinence; adverse event indicates acute psychiatric admissions.
[b] Single imputation indicates regression or matching based on participant characteristics.
[c] Imputation before resampling in the resampling method combination column applied to single, mean, assumption, and carry-forward methods.

care utilization of a single participant out of 142; similarly, Olmstead et al. (44) and Sindelar et al. (45) reported missing health care utilization data on 3 and 5 participants out of 415 and 388, respectively. Another study reported that a trial produced complete economic data for 94.5% of participants (46). However, economic evaluations by Bell et al. (47) and Avants et al. (38) that relied on complete case analysis and reported 18% and 22% missing data, respectively, are less likely to be valid and unbiased.

Imputation methods

Imputation involves the replacement of a missing observation with a single estimate, or multiple "good" estimates in the case of MI, for person-period observations missing in the study. The rigor and concomitant effectiveness of imputation methods used to control for missing data bias vary widely. A number of studies addressed missingness via an informed guess or conservative estimates; for example, Ruger et al. (48) and Sindelar et al. (45) assumed that missing substance use data would be imputed as positive for drug use. This approach could be biased if there are differences in retention to treatment across arms, because it is possible that not all missing data of this sort are the result of relapse; especially because urinalysis was the only measure of abstinence, unlike other studies that supplemented missing urinalysis data with self-report and medication adherence data (e.g., the Murphy et al. study (49)). Other studies made more qualified assumptions; for example, Busch et al. (50) indicated that for half of the 25% of participants for whom cost data were missing, sufficient information regarding the reason for failure to complete trial assessments was available to make assumptions on costs; for example, participants who were incarcerated were assumed to incur $0 in health care costs, and participants who did not complete cost assessments because they were in inpatient treatment at the time were assumed to incur costs equivalent to 14 days of residential treatment.

Carry-forward imputation. A number of studies between 2005 and 2018 used prior observations to replace missing data over time, including methods of last-observation carried forward, baseline observation carried forward, or alternatives (51–56). For example, the last-observation carried forward method of imputing missing data was used by Dijkgraaf et al. (52) on HRQoL assessments to calculate average quality-adjusted life years by treatment arm. This method involved applying the last value observed to subsequent incomplete observations. Doran et al. (53) reported imputation of missing heroin-use data based on "pre-trial behavioral patterns." Their approach to evaluate abstinence is commonly referred to as regression toward a value or baseline observation carried forward, where missing data are replaced with patient baseline observations, the mean, or worst or best outcome value. Baseline observation carried forward also was used in studies by Dunlop et al. (54) and Dunlap et al. (55), and Fals-Stewart and Lam (56) imputed values based on the worst outcome observed for each participant. The carry-forward approaches and their variants are not recommended, because they require investigators to assume that the intervention ceased to have an affect on outcome variables for participants with missing data, thereby implying the participant's health returns to preintervention values or that the short-term effects of the intervention are sustained. The corresponding assumption for cost is that health care utilization patterns observed in a prior period remain constant over the missing period. We have not found supportive research in the use of these approaches to addressing missing data in economic evaluations. Moreover, additional bias can be introduced if there is differential attrition or missingness between arms, or if differences exist between participants with and without missing observations (57). For instance, Dijkgraaf et al. (52) observed differential missingness by study intervention, indicating that the last-observation carry-forward method applied in their study may not have been a valid approach, because potential confounding factors leading to differential missingness were not addressed using last-observation carried forward.

Linear imputation. The last-observation or baseline carry-forward methods are less intuitive approaches when data are missing nonmonotonically (e.g., missing data in the middle of the trial and complete cases at end points). Goorden et al. (58) used linear interpolation to address nonmonotonic missing HRQoL data among persons with cannabis use disorder. The linear interpolation method described by Goorden et al. (58) simply fits a linear line between the complete data points and imputes the midpoint value (if the observation period is equidistant) to replace missing data for each participant. Goorden et al. (58) reported that 51% of HRQoL participant data was missing at the midpoint of the trial but that subsequent participant HRQoL data were available at the study endpoint, with only 7% missing. This approach can be a valid method of imputing nonmonotonic missing HRQoL data if, and only if, the HRQoL of each participant evolves linearly over time, on average, and there is little to no heterogeneity in HRQoL patterns among study participants. However, a recent study of HRQoL among persons with opioid use disorder demonstrated nonlinear and heterogenous quality of life patterns (59). Although Goorden et al. (58) also reported nonmonotonically missing health care cost data (36.5% missing at month 6, 3.5% missing at month 12), they imputed missingness by assuming $0 for missing costs. Goorden et al. (58) argued that an imputation model to estimate missing costs would not be feasible because there were many 0 observations in the complete case data. However, if HRQoL and costs were correlated over time, the differential methods of addressing missing data across costs and effectiveness measures may have biased the cost-effectiveness outcome (i.e., the incremental cost-effectiveness ratio). For example, if participants who incur higher costs (e.g., use more care) over the trial period fare better in terms of HRQoL than those with lower costs, then the results will not account for the additional costs required to gain an additional unit of effectiveness, casting doubt on the cost-effectiveness of the treatment.

Regression-based single imputation. Some studies published between 2012 and 2019 used regression-based single-imputation methods to address missing data (60–62).

Unlike last-observation carry forward or its variants explained above, regression-based imputation uses available participant data in a statistical model to impute missing outcome values. This approach is superior to the last-observation carry forward and linear interpolation because it can preserve the correlation among participant characteristics, costs, and effectiveness. However, there are 2 principal disadvantages of regression-based single imputation. First, unbiased estimates of cost and effect will only be a good approximation of missingness if the statistical model for the imputation is the "correct" model (63). MI is less affected by this limitation compared with single imputation, because it produces multiple approximations for each missing value instead of just a single imputed value. Second, although single-imputation regressions can preserve the observed correlation between cost and effectiveness of an intervention, it can overestimate the relationship between these 2 measures over the study period. For example, an intervention may require providing medications continually to participants, thereby incurring costs over time, but the intervention may only lead to short-term effectiveness. In this example, the correlation between cost and effectiveness is not consistent over the study period. Single imputation may not identify this change.

Other imputation methods. Additional eccentric methods of imputation were observed in our review, including methods that were motivated by limited research resources. Doran et al. (64, 65) evaluated resource costs by using a retrospective chart review of 50% of study participants in each arm and imputing missing follow-up data using either site-specific averages (64) or parametric imputation based on clinical review (65). In their 2006 article (65), Doran et al. used parametric values from the distribution of the observed data (i.e., mean and variance) to randomly generate values for missing data from that distribution. This method relies on the assumption that the 50% of participant data collected from retrospective review is a good representation of the entire population of participants randomized—in other words, the unobserved data are MCAR.

Proportion of missingness

The proportion of missing data can affect the validity of methods to address missingness. In an economic evaluation, Drost et al. (66) reported missing data at baseline and follow-up at greater than 70%. These authors argued that imputation for such a high percentage of missing data may lead to type 2 error; consequently, they conducted a complete case analysis with nonparametric bootstrapping. To address potential bias from missingness, however, Drost et al. (66) supplemented their study by conducting multiple sensitivity and subgroup analyses with the nonparametric bootstrap to assess the level of heterogeneity in the study by different subgroups (e.g., sex, age, education, religion, ethnicity). Identifying heterogeneity and examining the variability of results based on subgroup analysis, while informative for stakeholders, does not specifically address missingness. Patterns of missing data can be correlated by subgroups themselves in clinical trials, casting doubt on the validity of the approach used by Drost et al. Furthermore, recent research suggests that MI under an MAR assumption can be a valid approach even in cases of a large proportion of missing data (67).

Shortcomings of MI. When combining MI with the nonparametric bootstrap, 2 potential disadvantages arise when evaluating clinical trial data with a large proportion of missingness. First, the estimates from the MI model may be characterized by high variance, thereby introducing a large degree of uncertainty in the estimates and making it difficult to find statistically significant differences in estimated cost and effectiveness outcomes between study arms when applying the nonparametric bootstrap after MI. Second, it is unclear if the correct imputation model can be determined when there is a large proportion of missingness, because only a minority of the variation in the relationship between outcomes and covariates is observed (63). Although Madley-Dowd et al. (67) argued that the proportion of missing data should not prevent the use of the recommended MI method, they cautioned that bias reductions are only achieved if the "imputation model was correctly specified and included all variables related to missingness" (67, p. 69). More research is needed to assess the validity and robustness of MI as a preferred method in economic evaluations with a large proportion of missing data.

Inverse probability weights

The IPW is a popular statistical approach to addressing missing data in regression analysis, conditional on the assumption of MAR (19). One of the primary objectives of applying weights to a regression model is to rebalance the distribution of participant characteristics and other observed data to reduce potential bias from differential attrition or missingness after randomization. IPWs are calculated in a 2-step process whereby, first, the probability of an observation being a complete case is estimated, most commonly via multivariable logistic regression; then second, it is inverted to and used as a regression weight. Several economic evaluations in our review used IPW between 2010 and 2019 (49, 68–72), and all these studies embedded the IPW within the nonparametric bootstrap.

Shortcomings of IPW. Although IPW provides valid inferences when data are MAR, there are several limitations to this approach in addressing missingness. As noted by Zubizarreta (73), the standard 2-step procedure of estimating IPWs does not guarantee a balanced sample but merely a propensity score of being a complete case. Unlike MI, IPW requires complete data on all covariates to estimate the initial probability model, and bias may result in instances of highly variable weights (20, 73, 74). Additionally, the logistic regression used to estimate each participant's probability of being a complete case can be affected by partial or complete separation bias (75). Partial or complete separation can occur when covariates in the first step, logistic regression, are highly or perfectly correlated with complete cases or missing cases. Because the procedure requires regressing a dichotomous variable on a set of explanatory variables, the estimated odds ratio of the logistic regression

Table 2. Outline of Potential Issues by Analytic Method

Analytic Method	Potential Issues to Consider
Complete case analysis	Not recommended
	Least rigorous approach to addressing missingness
	Produces biased estimates of cost and effectiveness if data are MAR and MNAR
	Reduces statistical power in analysis
	Complete case resampling is not recommended in generating cost-effectiveness acceptability curves.
Carry-forward imputation	Rarely recommended
	Produces biased estimates of cost and effectiveness if data are MAR and MNAR
	Imposes relatively strong assumptions on the longitudinal values of missing data
	Introduces bias if trial data are affected by differential attrition or missingness between treatment arms
	Introduces bias if cost and effectiveness data are correlated
	No standard procedure recommended in combining carry-forward imputation with the nonparametric bootstrap
Linear imputation	Rarely recommended
	Only applicable for nonmonotonically missing longitudinal data
	Produces biased estimates of cost and effectiveness if data are MAR and MNAR
	Assumes missing data evolve linearly between nonmissing data points
	Introduces bias if cost and effectiveness data are correlated
	No standard procedure recommended in combining linear imputation with the nonparametric bootstrap
Single-regression imputation	Rarely recommended
	Requires "correct" imputation model to be identified
	Does not account for uncertainty in imputed values of missing data
	May over or underestimate the correlation between cost and effectiveness variables
	No standard procedure recommended in combining single regression imputation with the nonparametric bootstrap
Inverse probability weighting	Standard 2-step estimation procedure does not guarantee a balanced sample.
	Requires compete data on all covariates to estimate the initial probability model
	Propensity scores may exhibit high variance.
	Probability models (e.g., logit or probit) may be affected by partial separation when dichotomous explanatory variables are highly correlated with incomplete cases.
Multiple imputation	Requires "correct" imputation model to be identified
	No standard method recommended to ensure that each multiply imputed data set within the nonparametric bootstrap is valid
	No standard procedure recommended in combining multiple imputation with the nonparametric bootstrap
	Computation efficiency of multiple imputation embedded within resampling limits its application.

Abbreviations: MAR, missing at random; MNAR, missing not at random.

can be very large (or small) when these covariates are highly correlated with complete or missing cases (76). For example, economic evaluations of SUD interventions often collect criminal activity data for purposes of estimating societal costs. Investigators may rightly assume that criminal activity is an important determinant of being a complete case, because incarceration could result in dropout from a study. However, in a relatively small sample of participants, only a few instances of criminal activity may be observed in the data set, and many of them could be missing cases. In this scenario, the propensity score from the logistic regression model would be very small, but the IPW would be very large, potentially overweighting the participants with criminal activity who were complete cases in the analysis of cost and effectiveness outcomes.

There is a tradeoff, however, to the potential limitation of IPW in addressing missing data in economic evaluations compared with MI. It is general practice in the applied literature to combine the IPW and the nonparametric bootstrap by embedding the IPW estimation within the bootstrap. This standard approach of combining IPWs with the nonparametric bootstrap prevents highly variable estimates

of the distribution of predicted mean cost and effectiveness outcomes, because the probability of the prebootstrap values of the IPWs in a given iteration being large or small is nonzero. To our knowledge, there is no standard procedure of combining MI with the nonparametric bootstrap for use in economic evaluations. Although MI is arguably a more efficient approach to addressing missing data in small samples and data with a high proportion of missingness, this approach requires greater computational capacity when combined with the nonparametric bootstrap. Table 2 summarizes the potential issues that may arise for each analytic method discussed in our review.

CONCLUSION

The goal of an economic evaluation is to provide unbiased estimates of the relative costs and effectiveness between study arms that are as generalizable as possible within the limitations of the study design (1, 77). An economic evaluation conducted alongside a clinical trial relies on prospectively collected resource utilization and cost and effectiveness data to estimate the economic value of new interventions as well as the uncertainty around those estimates. Regression analysis allows for the estimation of the arithmetic mean costs and effectiveness across study arms while controlling for potential confounding variables at all time points. As previously discussed, collection of participant data through trial instruments often results in missing data, which can bias estimates from the study. Investigators should carefully consider the missing data mechanism when developing a statistical analysis plan to address missingness. Complete case analysis is inefficient and only applicable when data are MCAR. Last-observation and baseline carry-forward methods are invalid approaches to address missing data bias when there is differential missingness, and single-imputation methods are not as robust as MI. MI and IPW are 2 widely recommended approaches to addressing missing data bias, but they may require greater computational resources when combined with the nonparametric bootstrap (18, 20, 23).

Although IPW is a popular method to address missing data bias and unequal sampling, IPW faces limitations compared with MI, which has been suggested as a superior method, on average, by some authors (21, 28). Investigators should avoid the use of IPW in economic evaluations of clinical trials with a small sample size or a large proportion of incomplete cases, and use MI. However, imputation methods, in general, require sufficient data to construct a valid statistical model to approximate missing values from complete cases (23, 63), which may be more difficult than simply weighting the treatment effect with IPWs.

In our review of missing data methods in combination with the nonparametric bootstrap, we found that the majority of studies that met our inclusion criteria applied a single-imputation approach prior to bootstrap resampling. The second most popular approach was complete case resampling. No study that imputed missing data in our review embedded the imputation model in the nonparametric bootstrap. A potential barrier to single imputation or MI within the nonparametric bootstrap is likely the computation resources required to conduct such a procedure. Single imputation embedded within the nonparametric bootstrap can be a computationally efficient alternative (32), though other studies have recommended MI-embedded bootstrapping based on similar simulation study designs (33). Insofar as the popularity of IPW appears to have increased over time and little is currently known about the comparative benefits of MI within or before resampling, our findings indicate that IPW embedded within resampling is likely to grow in popularity and define the future direction of the field.

Determining the most appropriate method to address missingness in combination with the nonparametric bootstrap requires consideration of the missing data mechanism, computation resources, and a variety of other potential factors, such as the proportion of missingness in the study. Although embedding the IPW method within the bootstrap resampling procedure is a standard practice in the literature, more research is required with both simulation and real-world clinical trial data to assess the most valid and computationally efficient approach of combining MI and the nonparametric bootstrap, given the advantages of MI compared with IPW outlined in our study. It is, however, always best to prevent the occurrence of missing cases in the trial when possible. Moreover, the scope and complexity of potential statistical issues that arise with the choice of analytic methods discussed in our review warrant broader interdisciplinary collaboration between economists and clinical scientists. We recommend that clinical researchers interested in incorporating an economic evaluation in their study consult and engage with economic investigators early in the trial-design process. Funding agencies can help promote such interdisciplinary collaboration by supporting prospective economic analyses as secondary aims of clinical trial grant proposals.

ACKNOWLEDGMENTS

Author affiliations: Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, United States (Ali Jalali, Rulla M. Tamimi, Sean M. Murphy); and Department of Community and Behavioral Health, Washington State University Elson S. Floyd College of Medicine, Spokane, Washington, United States (Sterling M. McPherson).

This work was funded by National Institutes of Health grants P30DA040500 and R01DA046721.

Data sets generated in this review article are provided in the article.

We thank Joseph Daniel Pérez for copyediting the final manuscript.

This work was presented in part at the Addiction Health Services Research Conference 2021, October 6, 2021, held virtually.

The views expressed in this article are those of the authors and do not reflect those of the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. Glick H, Doshi JA, Sonnad SS, et al. *Economic Evaluation in Clinical Trials*. Second ed. Oxford, UK: Oxford University Press; 2015.
2. Brown JB, Adams ME. Patients as reliable reporters of medical care process: recall of ambulatory encounter events. *Med Care.* 1992;30(5):400–411.
3. Harlow SD, Linet MS. Agreement between questionnaire data and medical records: the evidence for accuracy of recall. *Am J Epidemiol.* 1989;129(2):233–248.
4. Roberts RO, Bergstralh EJ, Schmidt L, et al. Comparison of self-reported and medical record health care utilization measures. *J Clin Epidemiol.* 1996;49(9):989–995.
5. Wallihan DB, Stump TE, Callahan CM. Accuracy of self-reported health services use and patterns of care among urban older adults. *Med Care.* 1999;37(7): 662–670.
6. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev.* 2006;63(2):217–235.
7. Fleming TR. Addressing missing data in clinical trials. *Ann Intern Med.* 2011;154(2):113–117.
8. Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials.* 2000;21(3):167–189.
9. Luoma JB, Twohig MP, Waltz T, et al. An investigation of stigma in individuals receiving treatment for substance abuse. *Addict Behav.* 2007;32(7):1331–1346.
10. van Boekel LC, Brouwers EP, van Weeghel J, et al. Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: systematic review. *Drug Alcohol Depend.* 2013;131(1–2): 23–35.
11. Priester MA, Browne T, Iachini A, et al. Treatment access barriers and disparities among individuals with co-occurring mental health and substance use disorders: an integrative literature review. *J Subst Abuse Treat.* 2016;61:47–59.
12. Hoggatt KJ, Frayne SM, Saechao FS, et al. Substance use disorder-related disparities in patient experiences of primary care. *Health Equity.* 2019;3(1):193–197.
13. Edens EL, Kasprow W, Tsai J, et al. Association of substance use and VA service-connected disability benefits with risk of homelessness among veterans. *Am J Addict.* 2011;20(5): 412–419.
14. North CS, Eyrich-Garg KM, Pollio DE, et al. A prospective study of substance use and housing stability in a homeless population. *Soc Psychiatry Psychiatr Epidemiol.* 2010; 45(11):1055–1062.
15. Erickson SK, Rosenheck RA, Trestman RL, et al. Risk of incarceration between cohorts of veterans with and without mental illness discharged from inpatient units. *Psychiatr Serv.* 2008;59(2):178–183.
16. Abraham WT, Russell DW. Missing data: a review of current methods and applications in epidemiological research. *Curr Opin Psychiatry.* 2004;17(4):315–321.
17. Dziura JD, Post LA, Zhao Q, et al. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med.* 2013;86(3):343–358.
18. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Third ed. Hoboken, NJ: Wiley; 2019.
19. Wooldridge JM. Inverse probability weighted estimation for general missing data problems. *J Econom.* 2007;141(2): 1281–1301.
20. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res.* 2013;22(3):278–295.
21. Vroomen JM, Eekhout I, Dijkgraaf MG, et al. Multiple imputation strategies for zero-inflated cost data in economic evaluations: which method works best? *Eur J Health Econ.* 2016;17(8):939–950.
22. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med.* 2000;19(23):3219–3236.
23. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Econ.* 2012;21(2):187–200.
24. Dwivedi AK, Mallawaarachchi I, Alvarado LA. Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Stat Med.* 2017;36(14): 2187–2205.
25. Rubin DB. Inference and missing data. *Biometrika.* 1976; 63(3):581–592.
26. Briggs A, Clark T, Wolstenholme J, et al. Missing...presumed at random: cost-analysis of incomplete data. *Health Econ.* 2003;12(5):377–392.
27. Leyrat C, Seaman SR, White IR, et al. Propensity score analysis with partially observed covariates: how should multiple imputation be used? *Stat Methods Med Res.* 2019; 28(1):3–19.
28. Faria R, Gomes M, Epstein D, et al. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics.* 2014; 32(12):1157–1170.
29. Leurent B, Gomes M, Faria R, et al. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *Pharmacoeconomics.* 2018;36(8): 889–901.
30. Raad H, Cornelius V, Chan S, et al. An evaluation of inverse probability weighting using the propensity score for baseline covariate adjustment in smaller population randomised controlled trials with a continuous outcome. *BMC Med Res Methodol.* 2020;20(1):70.
31. Nguyen CD, Carlin JB, Lee KJ. Model checking in multiple imputation: an overview and case study. *Emerg Themes Epidemiol.* 2017;14:8.
32. Brand J, van Buuren S, le Cessie S, et al. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. *Stat Med.* 2019;38(2):210–220.
33. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Stat Med.* 2018;37(14):2252–2266.
34. Murphy SM, Polsky D. Economic evaluations of opioid use disorder interventions. *Pharmacoeconomics.* 2016;34(9): 863–887.
35. Onuoha EN, Leff JA, Schackman BR, et al. Economic evaluations of pharmacologic treatment for opioid use disorder: a systematic literature review. *Value Health.* 2021; 24(7):1068–1083.
36. Doran CM. Economic evaluation of interventions to treat opiate dependence: a review of the evidence. *Pharmacoeconomics.* 2008;26(5):371–393.
37. Jalali A, Ryan DA, McCollister KE, et al. Economic evaluation in the National Drug Abuse Treatment Clinical Trials Network: past, present, and future. *J Subst Abuse Treat.* 2020;112S:18–27.
38. Avants SK, Margolin A, Sindelar JL, et al. Day treatment versus enhanced standard methadone services for opioid-dependent patients: a comparison of clinical efficacy and cost. *Am J Psychiatry.* 1999;156(1):27–33.
39. Hartz DT, Meek P, Piotrowski NA, et al. A cost-effectiveness and cost-benefit analysis of contingency contracting-enhanced methadone detoxification treatment. *Am J Drug Alcohol Abuse.* 1999;25(2):207–218.

40. Kraft MK, Rothbard AB, Hadley TR, et al. Are supplementary services provided during methadone maintenance really cost-effective? *Am J Psychiatry.* 1997; 154(9):1214–1219.
41. Demissie S, LaValley MP, Horton NJ, et al. Bias due to missing exposure data using complete-case analysis in the proportional hazards regression model. *Stat Med.* 2003;22(4): 545–557.
42. McNeish D. Missing data methods for arbitrary missingness with small samples. *J Appl Stat.* 2017;44(1):24–39.
43. Olmstead TA, Petry NM. The cost-effectiveness of prize-based and voucher-based contingency management in a population of cocaine- or opioid-dependent outpatients. *Drug Alcohol Depend.* 2009;102(1–3):108–115.
44. Olmstead TA, Sindelar JL, Petry NA. Cost-effectiveness of prize-based incentives for stimulant abusers in outpatient psychosocial treatment programs. *Drug Alcohol Depend*. 2007;87(2–3):175–182.
45. Sindelar JL, Olmstead TA, Peirce JM. Cost-effectiveness of prize-based contingency management in methadone maintenance treatment programs. *Addiction.* 2007;102(9): 1463–1471.
46. Byford S, Barrett B, Metrebian N, et al. Cost-effectiveness of injectable opioid treatment v. oral methadone for chronic heroin addiction. *Brit J Psychiat.* 2013;203(5):341–349.
47. Bell J, Shanahan M, Mutch C, et al. A randomized trial of effectiveness and cost-effectiveness of observed versus unobserved administration of buprenorphine-naloxone for heroin dependence. *Addiction.* 2007;102(12):1899–1907.
48. Ruger JP, Chawarski M, Mazlan M, et al. Cost-effectiveness of buprenorphine and naltrexone treatments for heroin dependence in Malaysia. *PLoS One.* 2012;7(12):e50673.
49. Murphy SM, McCollister KE, Leff JA, et al. Cost-effectiveness of buprenorphine-naloxone versus extended-release naltrexone to prevent opioid relapse. *Ann Intern Med.* 2019;170(2):90–98.
50. Busch SH, Fiellin DA, Chawarski MC, et al. Cost-effectiveness of emergency department-initiated treatment for opioid dependence. *Addiction.* 2017;112(11): 2002–2010.
51. Harris AH, Gospodarevskaya E, Ritter AJ. A randomised trial of the cost effectiveness of buprenorphine as an alternative to methadone maintenance treatment for heroin dependence in a primary care setting. *Pharmacoeconomics.* 2005;23(1): 77–91.
52. Dijkgraaf MGW, van der Zanden BP, de Borgie CAJM, et al. Cost utility analysis of co-prescribed heroin compared with methadone maintenance treatment in heroin addicts in two randomised trials. *BMJ.* 2005;330(7503):1297–300a.
53. Doran CM, Shanahan M, Mattick RP, et al. Buprenorphine versus methadone maintenance: a cost-effectiveness analysis. *Drug Alcohol Depend*. 2003;71(3):295–302.
54. Dunlop AJ, Brown AL, Oldmeadow C, et al. Effectiveness and cost-effectiveness of unsupervised buprenorphine-naloxone for the treatment of heroin dependence in a randomized waitlist controlled trial. *Drug Alcohol Depend*. 2017;174:181–191.
55. Dunlap LJ, Zarkin GA, Orme S, et al. Re-engineering methadone-cost-effectiveness analysis of a patient-centered approach to methadone treatment. *J Subst Abuse Treat.* 2018; 94:81–90.
56. Fals-Stewart W, Lam WKK. Brief behavioral couples therapy for drug abuse: a randomized clinical trial examining clinical efficacy and cost-effectiveness. *Fam Syst Health.* 2008;26(4): 377–392.
57. Manson JE, Shufelt CL, Robins JM. The potential for postrandomization confounding in randomized clinical trials. *JAMA.* 2016;315(21):2273–2274.
58. Goorden M, van der Schee E, Hendriks VM, et al. Cost-effectiveness of multidimensional family therapy compared to cognitive behavioral therapy for adolescents with a cannabis use disorder: data from a randomized controlled trial. *Drug Alcohol Depend*. 2016;162:154–161.
59. Jalali A, Ryan DA, Jeng PJ, et al. Health-related quality of life and opioid use disorder pharmacotherapy: a secondary analysis of a clinical trial. *Drug Alcohol Depend*. 2020; 215:108221.
60. Heslin M, Patel A, Stahl D, et al. Randomised controlled trial to improve health and reduce substance use in established psychosis (IMPaCT): cost-effectiveness of integrated psychosocial health promotion. *BMC Psychiatry.* 2017; 17(1):407.
61. Rains LS, Marston L, Hinton M, et al. Clinical and cost-effectiveness of contingency management for cannabis use in early psychosis: the CIRCLE randomised clinical trial. *BMC Med.* 2019;17(1):161.
62. Lo Sasso AT, Byro E, Jason LA, et al. Benefits and costs associated with mutual-help community-based recovery homes: the Oxford House model. *Eval Program Plann.* 2012; 35(1):47–53.
63. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.
64. Doran CM, Shanahan M, Bell J, et al. A cost-effectiveness analysis of buprenorphine-assisted heroin withdrawal. *Drug Alcohol Rev.* 2004;23(2):171–175.
65. Doran CM, Shanahan M, Digiusto E, et al. Cost-effectiveness analysis of maintenance agonist treatments in the NEPOD. *Expert Rev Pharmacoecon Outcomes Res.* 2006;6(4): 437–446.
66. Drost RM, Paulus AT, Jander AF, et al. A web-based computer-tailored alcohol prevention program for adolescents: cost-effectiveness and intersectoral costs and benefits. *J Med Internet Res.* 2016;18(4):e93.
67. Madley-Dowd P, Hughes R, Tilling K, et al. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63–73.
68. Polsky D, Glick HA, Yang J, et al. Cost-effectiveness of extended buprenorphine-naloxone treatment for opioid-dependent youth: data from a randomized trial. *Addiction.* 2010;105(9):1616–1624.
69. Tran BX, Ohinmaa A, Duong AT, et al. Cost-effectiveness of methadone maintenance treatment for HIV-positive drug users in Vietnam. *AIDS Care.* 2012;24(3):283–290.
70. Murphy SM, McDonell MG, McPherson S, et al. An economic evaluation of a contingency-management intervention for stimulant use among community mental health patients with serious mental illness. *Drug Alcohol Depend.* 2015;153:293–299.
71. Murphy SM, Campbell ANC, Ghitza UE, et al. Cost-effectiveness of an internet-delivered treatment for substance abuse: data from a multisite randomized controlled trial. *Drug Alcohol Depend.* 2016;161:119–126.
72. Murphy SM, Polsky D, Lee JD, et al. Cost-effectiveness of extended release naltrexone to prevent relapse among criminal justice-involved individuals with a history of opioid use disorder. *Addiction.* 2017;112(8):1440–1450.
73. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *J Am Stat Assoc.* 2015;110(511):910–922.

74. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Stat Med.* 2016;35(30):5642–5655.
75. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Stat Med.* 2002;21(16): 2409–2419.
76. Lesaffre E, Albert A. Partial separation in logistic discrimination. *J R Stat Soc Series B Stat Methodol*. 1989; 51(1):109–116.
77. Neumann PJ, Sanders GD, Russell LB, et al. *Cost Effectiveness in Health and Medicine*. Second ed. Oxford, UK: Oxford University Press; 2017.