Received: 23 October 2023 | Revised: 19 February 2024 | Accepted: 15 March 2024

DOI: 10.1002/sim.10067

RESEARCH ARTICLE

Statistics in Medicine WILEY

Assessing efficacy in non-inferiority trials with non-adherence to interventions: Are intention-to-treat and per-protocol analyses fit for purpose?

Matthew Dodd[1] | **James Carpenter**[1,2] | **Jennifer A. Thompson**[3] | **Elizabeth Williamson**[1] | **Katherine Fielding**[3] | **Diana Elbourne**[1]

[1]Department of Medical Statistics, The London School of Hygiene & Tropical Medicine, London, UK

[2]The Medical Research Council Clinical Trials Unit (MRC CTU), UCL, London, UK

[3]Department of Infectious Disease Epidemiology, The London School of Hygiene & Tropical Medicine, London, UK

Correspondence
Matthew Dodd, Department of Medical Statistics, The London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK.
Email: matthew.dodd@lshtm.ac.uk

Funding information
Wellcome Trust, Grant/Award Number: 224485/Z/21/Z

Background: Non-inferiority trials comparing different active drugs are often subject to treatment non-adherence. Intention-to-treat (ITT) and per-protocol (PP) analyses have been advocated in such studies but are not guaranteed to be unbiased in the presence of differential non-adherence.

Methods: The REMoxTB trial evaluated two 4-month experimental regimens compared with a 6-month control regimen for newly diagnosed drug-susceptible TB. The primary endpoint was a composite unfavorable outcome of treatment failure or recurrence within 18 months post-randomization. We conducted a simulation study based on REMoxTB to assess the performance of statistical methods for handling non-adherence in non-inferiority trials, including: ITT and PP analyses, adjustment for observed adherence, multiple imputation (MI) of outcomes, inverse-probability-of-treatment weighting (IPTW), and a doubly-robust (DR) estimator.

Results: When non-adherence differed between trial arms, ITT, and PP analyses often resulted in non-trivial bias in the estimated treatment effect, which consequently under- or over-inflated the type I error rate. Adjustment for observed adherence led to similar issues, whereas the MI, IPTW and DR approaches were able to correct bias under most non-adherence scenarios; they could not always eliminate bias entirely in the presence of unobserved confounding. The IPTW and DR methods were generally unbiased and maintained desired type I error rates and statistical power.

Conclusions: When non-adherence differs between trial arms, ITT and PP analyses can produce biased estimates of efficacy, potentially leading to the acceptance of inferior treatments or efficacious regimens being missed. IPTW and the DR estimator are relatively straightforward methods to supplement ITT and PP approaches.

KEYWORDS
adherence, compliance, intention-to-treat, non-inferiority, per-protocol, trial

1 | INTRODUCTION

Non-inferiority trials are used to assess whether an experimental intervention is not worse than a proven comparator by more than a clinically acceptable amount (known as the non-inferiority margin). Despite becoming increasingly common, there are several challenges associated with this trial design, including the choices of study population, control intervention, the magnitude of non-inferiority margin, and outcomes.[1-3] Participants not receiving their randomly assigned intervention according to the protocol (referred to as non-adherence or non-compliance), such as missing several doses of a prescribed medication or not undergoing a surgical procedure as planned, is a particular concern in these studies because it may dilute the size of any effect between trial arms and inflate the risk of falsely concluding non-inferiority.[4,5] The terms adherence and compliance are often used interchangeably, though adherence is preferred here since it better reflects the partnership between participants and healthcare providers.

Non-adherence to interventions occurs frequently in trials and can bias estimates of efficacy in either direction.[5-7] A common approach to handling non-adherence is to assess outcomes among different analysis populations, with consistent results providing greater confidence in the trial conclusions.[8] Within the setting of non-inferiority trials, the intention-to-treat (ITT) and per-protocol (PP) populations have been advocated and are used extensively. However, in the presence of differential (or non-random) non-adherence, where the factors leading to non-adherence are also associated with outcomes, both of these approaches can result in bias in the same direction. Consequently, their agreement does not guarantee that non-inferiority conclusions are valid.[9-11]

More sophisticated statistical methods that attempt to account for the impact of non-adherence have been proposed. A recent systematic review identified a range of techniques that can be applied for this purpose within non-inferiority trials, including: instrumental variable (IV) methods, rank-preserving structural failure time models (RPSFTM) and G-estimation, inverse-probability-of-treatment weighting (IPTW), modeling adherence as a time-varying covariate in a time-to-event analysis, and adjusting for observed adherence within a regression model.[12] Details of such methods, their inherent assumptions, and their advantages and disadvantages were described as part of the review. Crucially, few studies have compared the performance of these alternative methods under different patterns of non-adherence in non-inferiority analyses, and so it remains unclear when they might be applied appropriately.

In the current study, we focus on a setting where the outcome of interest is binary, the desired effect estimate is the absolute risk difference, and non-adherence to the experimental intervention does not result in switching to the control intervention (and vice versa). Despite this type of non-adherence being common in trials comparing different active drugs (where non-adherence typically occurs in the form of missed doses or permanent discontinuation of the assigned medication), research into the impacts of treatment non-adherence has focused predominantly on crossovers from one trial arm to another. As a result, statistical methods that are either only appropriate in the presence of crossovers or can only be applied to time-to-event outcomes are not considered here.[13-15] Instead, multiple imputation (MI), IPTW and doubly-robust (DR) methods will be used to estimate treatment effects under the hypothetical situation that all participants had received 100% of their randomly assigned intervention, and the performance of these techniques compared with ITT and PP approaches. Despite the well-recognized limitations of adjusting for observed adherence within a regression model, this method is also included in order to assess its performance in a non-inferiority setting.[12] The aforementioned techniques will be evaluated under a range of non-adherence scenarios using computer simulations designed to replicate REMoxTB, a non-inferiority trial that assessed the safety and efficacy of novel tuberculosis (TB) regimens.[16] While we have chosen a TB treatment trial as the motivating example for this simulation study, the findings are likely to be relevant to other disease areas where non-inferiority trials with binary outcomes are used to compare different active drugs.

2 | THE REMOXTB TRIAL

For adults with drug-susceptible tuberculosis (DS-TB), effective treatment typically requires several drugs to be taken together for 6 months. Shortened regimens, which could improve adherence, reduce rates of adverse events, and lower healthcare costs are vital and have been assessed in non-inferiority trials.[16-18] However, these trials are often subject to treatment non-adherence, which has been associated with an increased risk of unfavorable outcomes.[19,20] REMoxTB was a randomized, double-blind, phase III non-inferiority trial that evaluated two 4-month experimental regimens (one isoniazid-based and the other ethambutol-based) in comparison with a standard 6-month control regimen among adults with newly diagnosed, previously untreated DS-TB.[16] Each experimental regimen consisted of four TB drugs prescribed

daily for 17 weeks, followed by 9 weeks of placebo, whereas the control regimen consisted of daily TB medications for 26 weeks. The primary efficacy endpoint, a composite unfavorable outcome of treatment failure or recurrence within 18 months post-randomization, was assessed using both PP (the primary analysis which excluded participants receiving less than ~80% of their allocated regimen) and modified intention-to-treat (mITT) analyses. Further details regarding the prescribed regimens and definitions of the analysis populations, non-adherence, and primary endpoint are provided in Data S1 (Table S1). The two experimental regimens failed to achieve non-inferiority in both the PP and mITT analyses. However, between 8% and 11% of the participants within each group met the protocol definition of treatment non-adherence and it is unclear how this may have influenced the trial results.

We conducted a simulation study based on the REMoxTB trial to assess the performance of different statistical methods that can be used to account for the impact of non-adherence to interventions in non-inferiority trials. By simulating various non-adherence scenarios, the study aimed to: (i) identify effective methods to address potential bias introduced by treatment non-adherence in non-inferiority trials comparing active drugs, (ii) assess how each method affects the probability of correctly concluding non-inferiority (statistical power) or falsely claiming non-inferiority (type I error), (iii) assess the impacts of unobserved confounding, treatment-adherence interactions, and misspecification of the adherence and outcome models on the performance of each method, and (iv) identify which methods are most appropriate for use in non-inferiority trials comparing active drugs in the presence of treatment non-adherence.

3 | METHODS

3.1 | Study design

Computer simulations were used to replicate a two-arm randomized non-inferiority TB trial with a composite unfavorable outcome consisting of treatment failure or recurrence within 18 months following randomization. Participants were randomized in a 1:1 ratio to either a 4-month experimental regimen (EXP) or a standard 6-month control regimen (CON) and the proportions of unfavorable outcomes occurring in the two groups compared by estimating the risk difference. Letting π_1 represent the true probability of unfavorable outcomes with EXP and π_0 the true probability of unfavorable outcomes with CON, the risk difference is defined as $\pi_1 - \pi_0$. This quantity can be estimated using the difference in the observed proportions of unfavorable outcomes among the EXP (p_1) and CON (p_0) groups, $p_1 - p_0$.

Data from REMoxTB were obtained from the Platform for Aggregation of Clinical TB Studies (TB-PACTS; https://c-path.org/programs/tb-pacts/) and used to inform the data generation process. Since the required parameters were similar in the two experimental groups of REMoxTB, data from these arms were combined before estimating the parameters to be used for the EXP group. Furthermore, it was assumed that the 4-month EXP regimen was followed by 2 months of placebo and therefore the adherence levels within this treatment group would be comparable to those observed in the experimental arms of REMoxTB. The simulation study was designed, conducted and reported using the "ADEMP" framework proposed by Morris and colleagues, and all analyses were conducted using Stata SE version 17.0.[21]

3.2 | Data generation

Participant-level data were generated within each simulated data set. First, three binary covariates ($\boldsymbol{C}$) representing age ($C_1 = 0$ represents <30 years, $C_1 = 1$ represents ≥30 years), smoking status ($C_2 = 0$ represents never smokers, $C_2 = 1$ represents ever smokers), and human immunodeficiency virus (HIV) status ($C_3 = 0$ represents HIV negative, $C_3 = 1$ represents HIV positive) were simulated to reflect the characteristics of participants included in the mITT population of REMoxTB. Each of these characteristics were identified as important predictors of both adherence and unfavorable outcomes within the trial data set and therefore act as potential confounders of this association. Participants were randomly allocated to treatment ($Z = 0$ represents allocation to CON, $Z = 1$ represents allocation to EXP) with equal probability and then adherence to the assigned treatment (A, defined as the overall percentage of prescribed doses received) was simulated conditional on the three covariates (see Data S1 for details). Initially, the distributions of adherence were assumed to be similar in the CON and EXP groups, given that the observed distributions in the control and experimental arms of REMoxTB were also comparable. Next, adherence was converted from a continuous variable (0 to 100%) to an ordi-

TABLE 1 Simulation study scenarios.

Element of data generation process	Scenarios explored
Quantity of treatment adherence[a] in EXP group	1) Similar to REMoxTB
	2) Median adherence increased by 3%
	3) Median adherence decreased by 3%
Assumed effect of receiving EXP vs CON (risk difference)	1) 6%
	2) 3%
	3) 0%
Unobserved confounding of adherence-outcome association	1) No unobserved confounding
	2) Unobserved confounding present
Treatment-adherence interaction	1) No interaction assumed
	2) Interaction assumed
Misspecification of the adherence or outcome models	1) No model misspecification
	2) Model misspecification present

Abbreviations: CON, standard 6-month control regimen; EXP, shortened 4-month experimental regimen.
[a]Defined as the overall percentage of prescribed doses received.

nal variable (<80%, 80% to <100%, or 100%) and each participant's probability of an unfavorable outcome estimated as follows:

$$\begin{gathered} y_i \sim Bernoulli(\pi_i) \\ \pi_i = 0.0210 + (TE * x_i) + (0.0265 * c_{1i}) + (0.0640 * c_{2i}) + (0.0875 * c_{3i}) + (0.0570 * a_{1i}) + (0.7000 * a_{2i}), \end{gathered} \tag{1}$$

where $y_i = 1$ represents an unfavorable outcome occurring for participant i ($i = 1, 2, \ldots, n$) with probability π_i, TE is the assumed treatment effect for receiving EXP vs CON, $x_i = 1$ if the ith participant receives EXP and $x_i = 0$ if they receive CON, $a_{1i} = 1$ if the ith participant receives 80% to <100% of doses and $a_{1i} = 0$ otherwise, and $a_{2i} = 1$ if the ith participant receives <80% of doses and $a_{2i} = 0$ otherwise. Different values of TE were explored ranging from 0 (non-inferiority of EXP) to 0.06 (inferiority of EXP) (Table 1). The remaining coefficients were estimated using the REMoxTB data set as described in Data S1. Finally, the outcome ($Y = 0$ represents a favorable outcome, $Y = 1$ represents an unfavorable outcome) was simulated using a Bernoulli pseudo-random variable taking the value one according to the probabilities estimated in Equation (1).

Based on a 15% risk of unfavorable outcomes in the EXP and CON groups and assuming 85% power, a one-sided type I error of 2.5%, and a non-inferiority margin of 6% on the risk difference scale (chosen to emulate REMoxTB), 1280 participants were required per simulated data set. In order to be 95% confident that the type I error would be between 1.8% and 3.2%, 2000 simulated data sets were generated for each non-adherence scenario explored.

A range of non-adherence scenarios were explored by varying different elements of the data generation process one at a time (Table 1). Among the factors varied were the quantity of treatment adherence in the EXP group, the assumed effect of receiving EXP vs CON, and the presence/absence of unobserved confounding of the adherence-outcome association. The level of adherence in the EXP group was increased by adding a random quantity to the percentage of doses received among individuals who did not originally receive all doses. This random quantity was drawn from a normal distribution with a mean of 3.28 and a variance of 1 (simulating >100 000 observations indicated that this would result in a 3% improvement in the median percentage of doses received). In order to reduce the level of adherence in the EXP group, the percentage of doses received was reduced among all individuals originally receiving less than 100% of doses, plus 20% of those originally receiving all doses. The latter were sampled so that participants with risk factors for being non-adherent (>30 years old, ever smokers or HIV positive) were more likely to be selected. After selecting the aforementioned individuals, a random quantity drawn from a normal distribution with a mean of 2.48 and a variance of 1 was subtracted from their original percentage of doses received (simulating >100 000 observations indicated that this would result in a 3% reduction in the median percentage of doses received). Unobserved confounding of the adherence-outcome association was induced by

omitting HIV status, the covariate with the largest effect on the risk of unfavorable outcomes, from the relevant analyses. As well as assuming that the effects of non-adherence on unfavorable outcomes were the same in the CON and EXP groups, a treatment-adherence interaction was also explored. This interaction was produced by using Equation (1) to calculate the risks of unfavorable outcomes among the EXP group, but multiplying the coefficients a_1 and a_2 by a factor of 0.8 when calculating the corresponding risks among the CON group. Consequently, the effect of non-adherence on unfavorable outcomes was assumed to be greater for EXP than CON, which is plausible given that the EXP regimen contains fewer doses in total and so each missed dose is likely to be more important. Finally, misspecification of the models used to predict adherence and outcomes was explored by omitting all interaction terms (see Section 3.4). Further details regarding the data generation process are provided in Data S1.

3.3 | Estimand

The ICH E9(R1) addendum provides a structured framework for defining estimands and sensitivity analyses in trials.[22] Where possible, we define each of the five attributes used to construct an estimand in accordance with the mITT approach used in REMoxTB:

i. Treatment—A 4-month experimental regimen (EXP) will be compared to a standard 6-month control regimen (CON) with each participant assigned to receive one of the treatments at random.
ii. Population—Adults with newly diagnosed, previously untreated DS-TB.
iii. Outcome—A binary outcome ('favorable' or 'unfavorable'), where a favorable outcome is defined as participants with culture negative status at 18 months post-randomization, who had not already been classified as having an unfavorable outcome (see intercurrent events below), and whose last positive culture was followed by at least two negative culture results. Culture negative status is defined as two negative culture results without an intervening positive result. Missing values of the outcome due to withdrawals of consent (experienced by 1.8% of REMoxTB participants) or losses to follow-up (0.5%) will be treated as unfavorable outcomes.
iv. Intercurrent events—In this simulation study, the intercurrent event of interest is non-adherence to the randomly assigned treatment which will be handled by estimating the effect of EXP vs CON under the hypothetical scenario that all participants had received 100% of their allocated regimen. Changes or extensions to the assigned regimen (7.4%), non-violent deaths during treatment (0.9%), and TB-related deaths during follow-up (0.1%) will all be treated as unfavorable outcomes. Participants experiencing reinfections with a new strain of TB (2.2%) or non-TB-related death (0.4%) will be classified as having non-assessable (neither favorable nor unfavorable) outcomes and will be excluded from simulations.[23]
v. Population-level summary—The effect of EXP vs CON had all participants been fully adherent to their allocated regimen will be estimated using a risk difference and its corresponding two-sided 95% confidence interval (CI). Non-inferiority of EXP vs CON will be concluded if the upper limit of the 95% CI for the risk difference is less than the non-inferiority margin of 6%.

3.4 | Analysis methods

The MI, IPTW, and DR methods were applied in order to target the estimand described in Section 3.3. Other statistical approaches which do not target this estimand but are commonly employed in non-inferiority trials were also assessed, namely ITT and PP analyses, and adjustment for observed adherence within a regression model. In the subsequent descriptions, the regimen received is denoted by X ($X = 0$ represents receiving CON, $X = 1$ represents receiving EXP) and the percentage of doses received is denoted by A_{100} ($A_{100} = 0$ represents receiving less than 100% of doses, $A_{100} = 1$ represents receiving 100% of doses). Therefore, there are four potential outcomes, where $Y^{X=0,A_{100}=0}$ represents the potential outcome if less than 100% of doses of CON are received, $Y^{X=0,A_{100}=1}$ represents the potential outcome if 100% of doses of CON are received, $Y^{X=1,A_{100}=0}$ represents the potential outcome if less than 100% of doses of EXP are received, and $Y^{X=1,A_{100}=1}$ represents the potential outcome if 100% of doses of EXP are received. These are counterfactual outcomes meaning that, for each individual, only one can be observed in reality. Similarly, potential outcomes depending on randomization (Z) are represented by $Y^{Z=0}$ and $Y^{Z=1}$ for allocation to CON and EXP, respectively.

3.4.1 | Intention-to-treat and per-protocol analyses

The ITT analysis incorporates all participants according to their randomly allocated regimen, regardless of the treatment they actually receive. The effect of Z on Y is estimated as $\Pr\left(Y^{Z=1}=1\right)-\Pr\left(Y^{Z=0}=1\right)$. In contrast, the PP analysis excludes participants who are considered to be non-adherent to their allocated regimen. Three separate definitions of non-adherence were considered based on participants receiving less than 100%, 90%, or 80% of their prescribed doses (denoted PP100, PP90, and PP80, respectively). For instance, the PP100 analysis estimates: $\Pr\left(Y^{Z=1,A_{100}=1}=1\right)-\Pr\left(Y^{Z=0,A_{100}=1}=1\right)$. The ITT and PP analyses were conducted using generalized linear models (GLM) for binary outcomes with identity link functions in order to estimate risk differences.

3.4.2 | Adjustment for observed adherence within a regression model

Adjustment for participants' observed levels of adherence was performed using a logistic regression model in the ITT population. The model contained an indicator variable for allocated treatment and an ordinal variable for the percentage of prescribed doses received (<80%, 80% to <100%, or 100%). Predicted probabilities of the counterfactual outcomes $Y^{Z=0}$ and $Y^{Z=1}$ were obtained for each participant and the difference in the mean predicted probabilities between the EXP and CON groups calculated (ie, a risk difference), along with a 95% CI estimated using delta-method SEs. This approach is referred to as an adjusted ITT analysis henceforth.

3.4.3 | Multiple imputation of outcomes

MI was used to impute outcomes for participants receiving less than 100% of doses under the counterfactual scenario that they had been fully adherent. Log-odds ratios for the effects of the covariates ($\boldsymbol{C}$) and adherence (A_{100}) on unfavorable outcomes were estimated using logistic regression models fitted separately within the CON and EXP groups which contained indicator variables for $\boldsymbol{C}$ and A_{100}, and all of their possible interactions. A new set of coefficients were then drawn from a multivariate normal distribution using the estimated coefficients and variance-covariance matrix from the logistic regression model as the mean and variance, respectively, and these new coefficients used to predict participants' log-odds of an unfavorable outcome. After transforming these log-odds into probabilities using the expit function, a participant-level bias correction was applied (in the form of a second-order Taylor series expansion of the logit function) to account for the fact that the expectation of these predicted probabilities is not equal to the expectation of the predicted probabilities from the initial logistic regression model (see Data S1 for details). Within each stratum of $\boldsymbol{C}$ and A_{100}, the predicted probabilities for those receiving all doses were then used to impute outcomes among those receiving less than 100% of doses. Ten imputations were performed per simulated data set and the estimated risk differences were combined using Rubin's rules.[24] If the participants within a particular stratum of $\boldsymbol{C}$ who received 100% of doses all experienced the same outcome (either all favorable or all unfavorable), a decision was made to impute unfavorable outcomes with risks of 1% and 90%, respectively.

3.4.4 | Inverse-probability-of-treatment weighting

IPTW was used to re-weight the outcomes of participants receiving all doses so that the re-weighted pseudo-population contained no individuals that were non-adherent. First, predicted probabilities of receiving all doses were estimated within each stratum of $\boldsymbol{C}$ using logistic regression models fitted separately in the CON and EXP groups which contained indicator variables for each covariate in $\boldsymbol{C}$ and all of their possible interactions. Weights were calculated as the inverse of these predicted probabilities and the risk difference estimated using a weighted GLM for a binary outcome with an identity link function and robust SEs to account for weighting. The IPTW estimator for a given treatment group is:

$$\hat{\Delta}_{IPTW}=\frac{1}{n}\sum\frac{A_{100_i}Y_i}{\hat{p}_i}, \tag{2}$$

where $\hat{p}_i = \Pr(A_{100_i} = 1|\boldsymbol{C}_i)$. If the participants within a particular stratum of $\boldsymbol{C}$ all received 100% of doses, a decision was made to re-weight these individuals using a weight equal to 1.01 (equivalent to a predicted probability of receiving all doses of 0.99) in order to incorporate some random variability.

3.4.5 | Doubly-robust estimator

The DR estimator was used to combine properties of the MI and IPTW methods. This approach involves three key steps: (i) predicted probabilities of receiving all prescribed doses are estimated separately in the CON and EXP groups as described for the IPTW method, and then inversed to obtain weights, (ii) predicted probabilities of unfavorable outcomes are estimated separately in the CON and EXP groups as described for the MI method, and (iii) each predicted probability of the outcome (from Step ii) is weighted by the predicted probability of receiving all doses (from Step i) in order to produce a weighted average of the two models. The phrase "doubly robust" refers to the fact that this estimator requires at least one of the models used to predict adherence or outcome to be correctly specified, but not both. The DR estimator for a given treatment group is:

$$\hat{\Delta}_{DR} = \frac{1}{n}\sum \frac{A_{100_i} Y_i}{\hat{p}_i} - \frac{1}{n}\sum \frac{\left(A_{100_i} - \hat{p}_i\right)}{\hat{p}_i}\,\hat{Y}_i, \tag{3}$$

where $\hat{Y}_i$ is the predicted probability of an unfavorable outcome for the ith participant (from Step ii). $\hat{\Delta}_{DR}$ was estimated separately in each treatment group using the *teffects* command in Stata and the risk difference calculated using their absolute difference (EXP minus CON). The variance for the risk difference was calculated as the sum of the group-specific variances of $\hat{\Delta}_{DR}$.

3.5 | Performance measures

To assess the performance of the different analysis methods under each non-adherence scenario, we calculated the following: the mean estimated risk difference, the difference between the mean estimated risk difference and the true risk difference (the bias), the empirical SE (the SD of the estimated risk differences), the mean of the estimated SEs, the loss of information (using the variance of the risk difference assuming perfect adherence as the comparator), the mean squared error (bias-squared plus the empirical SE), and the type I error or power.[21] To assess the probability of a type I error, unfavorable outcomes were generated assuming that EXP was inferior to CON (6% treatment effect) and analyses which concluded non-inferiority were deemed to have committed a type I error. To examine power, unfavorable outcomes were generated assuming that EXP was non-inferior to CON (0% or 3% treatment effect) and analyses which failed to conclude non-inferiority were deemed to have committed a type II error. Power is the probability of avoiding a type II error. Monte Carlo Standard Errors (MCSE) were calculated for the bias and type I error, and used to estimate corresponding 95% CIs.

4 | RESULTS

4.1 | Similar quantities of non-adherence in the control and experimental groups (no unobserved confounding or treatment-adherence interaction)

In simulations where the quantities of non-adherence were assumed to be similar among trial arms, a median of 7.4%, 20.2%, and 72.3% of the participants in each treatment group received 80%, 80% to <100%, and 100% of their prescribed doses, respectively. All of the analysis methods were unbiased when the treatment effect was assumed to be either 3% or 6%, though the type I error rates tended to be slightly higher than the nominal value of 2.5% (Figure 1; Tables S2 and S3). In particular, the PP80 and PP90 approaches appeared to inflate the type I error rate, whereas the ITT analysis was unbiased and maintained a type I error rate close to 2.5%. When no difference in the risks of unfavorable outcomes was assumed between treatment groups, all methods were unbiased and resulted in estimates of power that were at least 85% (the level of power assumed in the sample size calculation; Table S4).

TABLE 2 Performance of analysis methods assuming better adherence in the experimental group (EXP) than the control group (CON), a treatment effect of 6%, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction.

Analysis method	Mean risk of unfavorable outcome, %		Mean risk difference, %	Mean bias, % (95% CI)	Empirical SE[a]	Mean SE	Variance of SEs	Loss of information[b], %	Type I error, % (95% CI)	MSE
	CON	EXP								
ITT	13.99	19.05	5.06	−0.94 (−1.04 to −0.84)	2.02	2.07	0.0027	27.8	6.5 (5.4 to 7.6)	4.95
PP100	7.24	13.49	6.24	0.24 (0.16 to 0.33)	1.89	1.88	0.0064	18.0	1.6 (1.1 to 2.2)	3.64
PP90	8.76	13.90	5.14	−0.86 (−0.95 to −0.77)	1.84	1.84	0.0043	13.5	7.8 (6.6 to 9.0)	4.14
PP80	8.81	13.94	5.13	−0.87 (−0.96 to −0.78)	1.83	1.84	0.0043	12.5	7.7 (6.5 to 8.9)	4.12
Adjusted ITT[c]	13.48	19.81	6.33	0.33 (0.24 to 0.41)	1.86	1.87	0.0039	14.9	1.6 (1.1 to 2.1)	3.56
MI of outcomes	7.63	13.74	6.11	0.11 (0.02 to 0.19)	1.95	1.95	0.0185	22.4	2.0 (1.4 to 2.7)	3.80
IPTW	7.64	13.64	6.00	0.00 (−0.08 to 0.08)	1.93	1.94	0.0081	21.3	2.5 (1.8 to 3.2)	3.74
DR estimator	7.64	13.64	6.00	0.00 (−0.08 to 0.08)	1.93	1.94	0.0071	21.3	2.5 (1.8 to 3.2)	3.74

Abbreviations: CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; MSE, mean squared error; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses; SE, standard error.

[a]The SD of the estimated risk differences.

[b]Using the variance of the risk difference assuming perfect adherence as the comparator.

[c]Adjusted for observed levels of adherence as a covariate.

TABLE 3 Performance of analysis methods assuming better adherence in the experimental group (EXP) than the control group (CON), a treatment effect of 0%, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction.

Analysis method	Mean risk of unfavorable outcome, %		Mean risk difference, %	Mean bias, % (95% CI)	Empirical SE[a]	Mean SE	Variance of SEs	Loss of information[b], %	Statistical power, % (95% CI)	MSE
	CON	EXP								
ITT	14.02	13.07	−0.96	−0.96 (−1.05 to −0.86)	1.94	1.91	0.0033	41.6	95.1 (94.1 to 96.0)	4.69
PP100	7.29	7.49	0.20	0.20 (0.13 to 0.28)	1.67	1.64	0.0073	20.6	93.2 (92.1 to 94.3)	2.83
PP90	8.79	7.92	−0.88	−0.88 (−0.96 to −0.79)	1.66	1.61	0.0049	19.5	98.5 (98.0 to 99.0)	3.51
PP80	8.84	7.95	−0.89	−0.89 (−0.97 to −0.81)	1.65	1.61	0.0048	19.2	98.4 (97.9 to 99.0)	3.52
Adjusted ITT[c]	13.46	13.65	0.20	0.20 (0.12 to 0.27)	1.70	1.66	0.0043	23.8	92.9 (91.8 to 94.1)	2.94
MI of outcomes	7.67	7.71	0.04	0.04 (−0.04 to 0.12)	1.75	1.71	0.0180	27.6	92.3 (91.1 to 93.5)	3.05
IPTW	7.68	7.67	−0.01	−0.01 (−0.09 to 0.06)	1.72	1.71	0.0089	25.3	93.6 (92.5 to 94.6)	2.96
DR estimator	7.68	7.67	−0.01	−0.01 (−0.09 to 0.06)	1.72	1.71	0.0079	25.3	93.4 (92.3 to 94.5)	2.96

Abbreviations: CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; MSE, mean squared error; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses; SE, standard error.

[a]The SD of the estimated risk differences.

[b]Using the variance of the risk difference assuming perfect adherence as the comparator.

[c]Adjusted for observed levels of adherence as a covariate.

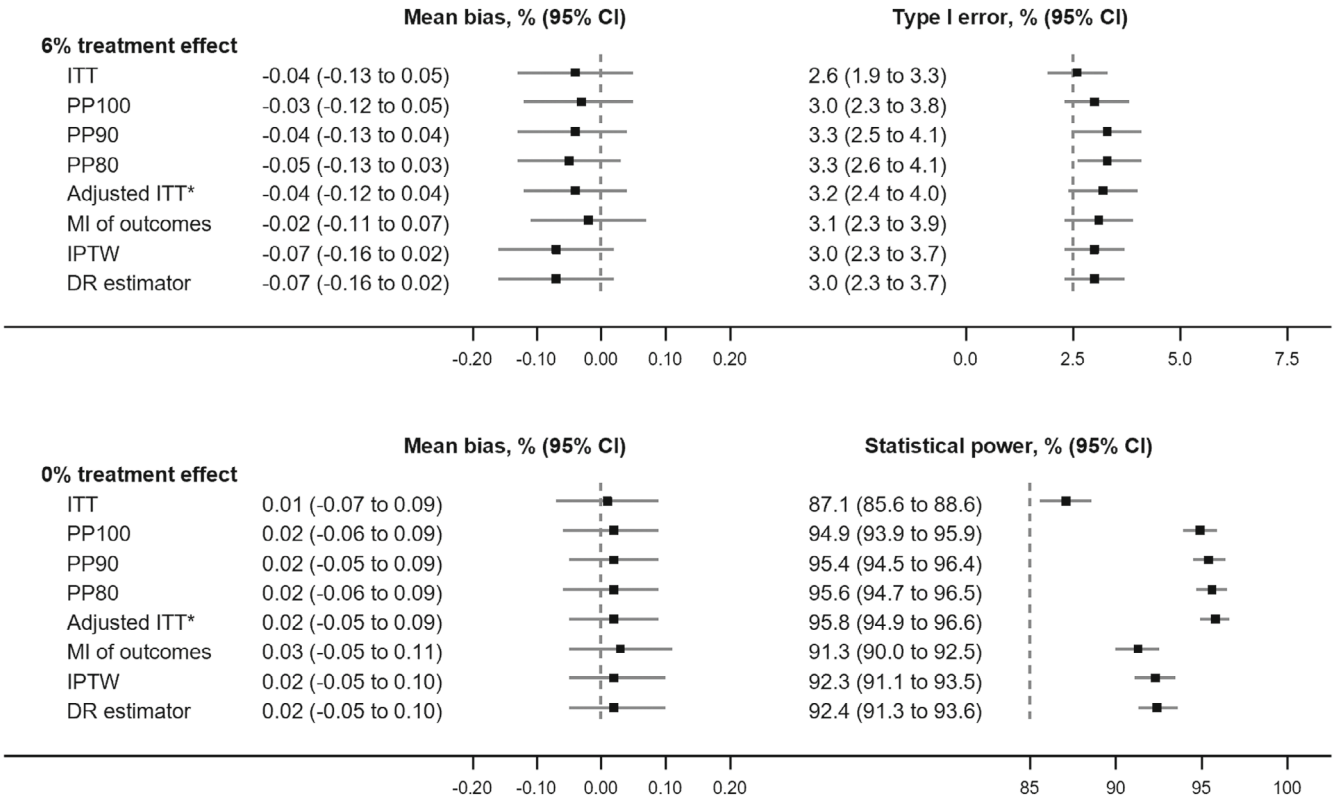

FIGURE 1 Mean bias, type I error, and statistical power of each analysis method assuming similar quantities of non-adherence in the control and experimental groups, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction. CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses. * Adjusted for observed levels of adherence as a covariate.

4.2 | Better adherence in the experimental group (no unobserved confounding or treatment-adherence interaction)

In simulations assuming better adherence in the EXP group, a median of 7.2%, 5.9%, and 86.8% of participants in this arm received 80%, 80% to <100%, and 100% of their prescribed doses, respectively (corresponding values for the CON group were unchanged from Section 4.1). IPTW and the DR estimator were the only unbiased methods under both 3% and 6% treatment effects (Figure 2; Table S5 and Table 2). The MI approach was also unbiased when the treatment effect was assumed to be 3%, but overestimated the treatment effect of 6% by a small amount (bias = 0.11%; 95% CI 0.02% to 0.19%). The ITT, PP, and adjusted ITT analyses were all biased under both 3% and 6% treatment effects, with the magnitude of bias appearing to be similar for the ITT, PP80, and PP90 approaches; the corresponding type I error estimates were inflated approximately 3-fold. Similar patterns were observed when no effect of EXP was assumed, with only the MI, IPTW, and DR methods providing unbiased estimates of the treatment effect, and the ITT, PP80, and PP90 approaches appearing to be biased by similar amounts (Table 3). All of the analysis methods resulted in estimates of power that were greater than 90%.

4.3 | Worse adherence in the experimental group (no unobserved confounding or treatment-adherence interaction)

In simulations assuming worse adherence in the EXP group, a median of 7.6%, 34.5%, and 57.9% of participants in this arm received 80%, 80% to <100%, and 100% of their prescribed doses, respectively (corresponding values for the CON group were unchanged from Section 4.1). The MI, IPTW, and DR approaches were the only unbiased methods under both 3% and 6% treatment effects (Figure 3; Table S6 and Table 4). The ITT, PP, and adjusted ITT analyses were all biased under both 3% and 6% treatment effects, with the magnitude of bias appearing to be similar for the ITT, PP80, and PP90 approaches;

TABLE 4 Performance of analysis methods assuming worse adherence in the experimental group (EXP) than the control group (CON), a treatment effect of 6%, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction.

Analysis method	Mean risk of unfavorable outcome, %		Mean risk difference, %	Mean bias, % (95% CI)	Empirical SE[a]	Mean SE	Variance of SEs	Loss of information[b], %	Type I error, % (95% CI)	MSE
	CON	EXP								
ITT	13.96	20.89	6.93	0.93 (0.83 to 1.03)	2.09	2.11	0.0026	32.5	0.9 (0.5 to 1.4)	5.22
PP100	7.26	12.83	5.57	−0.43 (−0.53 to −0.34)	2.09	2.11	0.0105	32.9	4.3 (3.5 to 5.2)	4.57
PP90	8.75	15.66	6.91	0.91 (0.82 to 1.00)	1.85	1.90	0.0042	14.4	0.6 (0.3 to 0.9)	4.27
PP80	8.80	15.73	6.93	0.93 (0.84 to 1.02)	1.84	1.89	0.0041	13.6	0.6 (0.2 to 0.9)	4.27
Adjusted ITT[c]	14.41	20.24	5.83	−0.17 (−0.25 to −0.09)	1.86	1.86	0.0035	14.7	3.5 (2.7 to 4.2)	3.48
MI of outcomes	7.63	13.70	6.06	0.06 (−0.04 to 0.16)	2.28	2.26	0.0670	43.6	2.6 (1.9 to 3.3)	5.22
IPTW	7.64	13.63	5.99	−0.01 (−0.11 to 0.09)	2.25	2.27	0.0176	42.1	2.7 (1.9 to 3.4)	5.08
DR estimator	7.64	13.63	5.99	−0.01 (−0.11 to 0.08)	2.25	2.24	0.0134	42.0	2.7 (1.9 to 3.4)	5.07

Abbreviations: CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; MSE, mean squared error; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses; SE, standard error.

[a]The SD of the estimated risk differences.

[b]Using the variance of the risk difference assuming perfect adherence as the comparator.

[c]Adjusted for observed levels of adherence as a covariate.

TABLE 5 Performance of analysis methods assuming worse adherence in the experimental group (EXP) than the control group (CON), a treatment effect of 0%, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction.

Analysis method	Mean risk of unfavorable outcome, %		Mean risk difference, %	Mean bias, % (95% CI)	Empirical SE[a]	Mean SE	Variance of SEs	Loss of information[b], %	Statistical power, % (95% CI)	MSE
	CON	EXP								
ITT	13.99	14.90	0.90	0.90 (0.80 to 1.00)	2.04	1.96	0.0032	46.9	73.0 (71.1 to 74.9)	4.97
PP100	7.28	6.86	−0.42	−0.42 (−0.50 to −0.33)	1.85	1.78	0.0117	35.4	93.3 (92.2 to 94.3)	3.59
PP90	8.78	9.64	0.86	0.86 (0.78 to 0.95)	1.74	1.69	0.0051	26.7	85.3 (83.7 to 86.8)	3.76
PP80	8.82	9.71	0.88	0.88 (0.80 to 0.97)	1.73	1.68	0.0049	26.1	85.1 (83.5 to 86.6)	3.77
Adjusted ITT[c]	14.56	14.34	−0.22	−0.22 (−0.30 to −0.15)	1.76	1.70	0.0041	28.5	95.3 (94.4 to 96.2)	3.14
MI of outcomes	7.65	7.75	0.10	0.10 (0.01 to 0.19)	2.06	1.96	0.0551	47.7	80.9 (79.2 to 82.6)	4.24
IPTW	7.66	7.70	0.04	0.04 (−0.04 to 0.13)	1.99	1.96	0.0211	44.3	83.4 (81.8 to 85.1)	3.97
DR estimator	7.66	7.70	0.04	0.04 (−0.05 to 0.13)	1.99	1.94	0.0151	44.3	84.4 (82.8 to 86.0)	3.97

Abbreviations: CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; MSE, mean squared error; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses; SE, standard error.

[a]The SD of the estimated risk differences.

[b]Using the variance of the risk difference assuming perfect adherence as the comparator.

[c]Adjusted for observed levels of adherence as a covariate.

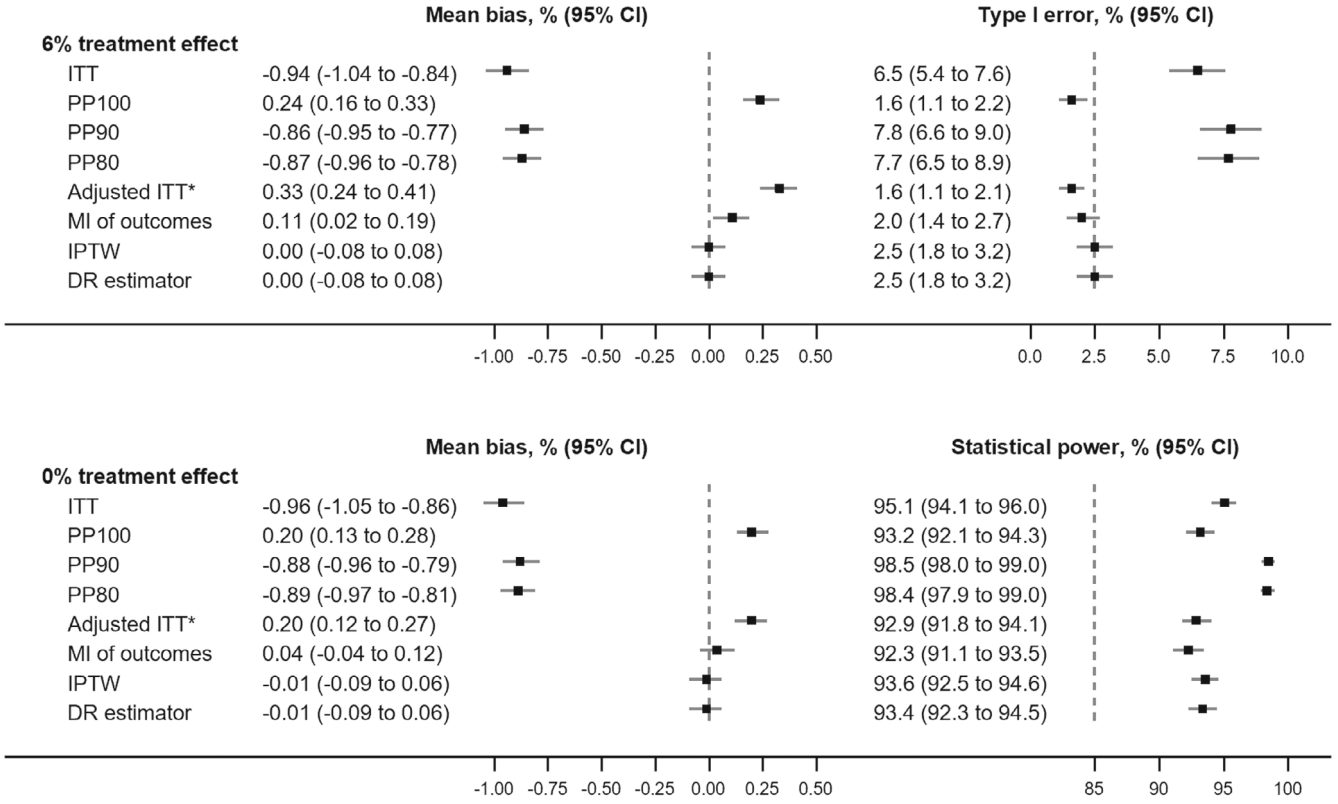

FIGURE 2 Mean bias, type I error, and statistical power of each analysis method assuming better adherence in the experimental group than the control group, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction. CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses. * Adjusted for observed levels of adherence as a covariate.

the corresponding type I error estimates were less than 1%. When no difference in the risks of unfavorable outcomes was assumed between treatment groups, similar patterns were observed, though the MI approach overestimated the treatment effect by a small amount (bias = 0.10%; 95% CI 0.01% to 0.19%) and resulted in a moderate loss of power (80.9%; 95% CI 79.2% to 82.6%). IPTW and the DR estimator remained unbiased and resulted in estimates of power that were close to 85% (Table 5). In contrast, the ITT analysis resulted in a considerable loss of power (73.0%; 95% CI 71.1% to 74.9%).

4.4 | Unobserved confounding of the adherence-outcome association

In the presence of unobserved confounding of the adherence-outcome association, the MI, IPTW, and DR approaches were all unbiased under treatment effects of 0%, 3%, and 6% when the quantities of non-adherence were similar among trial arms (Table S7). In simulations assuming better adherence in the EXP group than the CON group, all three methods overestimated the treatment effects by a small amount, with the corresponding estimates of bias ranging from 0.08% to 0.21% (Table S8). In contrast, the three methods tended to be unbiased in simulations assuming worse adherence in the EXP group than the CON group, except for IPTW and the DR estimator which both underestimated the treatment effect of 3% by a small amount (bias = −0.14%; 95% CI −0.23% to −0.05%) (Table S9). Regardless of the assumed level of adherence in the EXP group, unobserved confounding did not affect the estimated type I error rates substantially (range: 1.9%–3.1%).

4.5 | Presence of treatment-adherence interaction

When the effect of non-adherence on unfavorable outcomes was assumed to be greater for EXP than CON, the ITT, PP, and adjusted ITT analyses tended to be biased, the magnitude of which could be substantial (Figures S1–S3 and

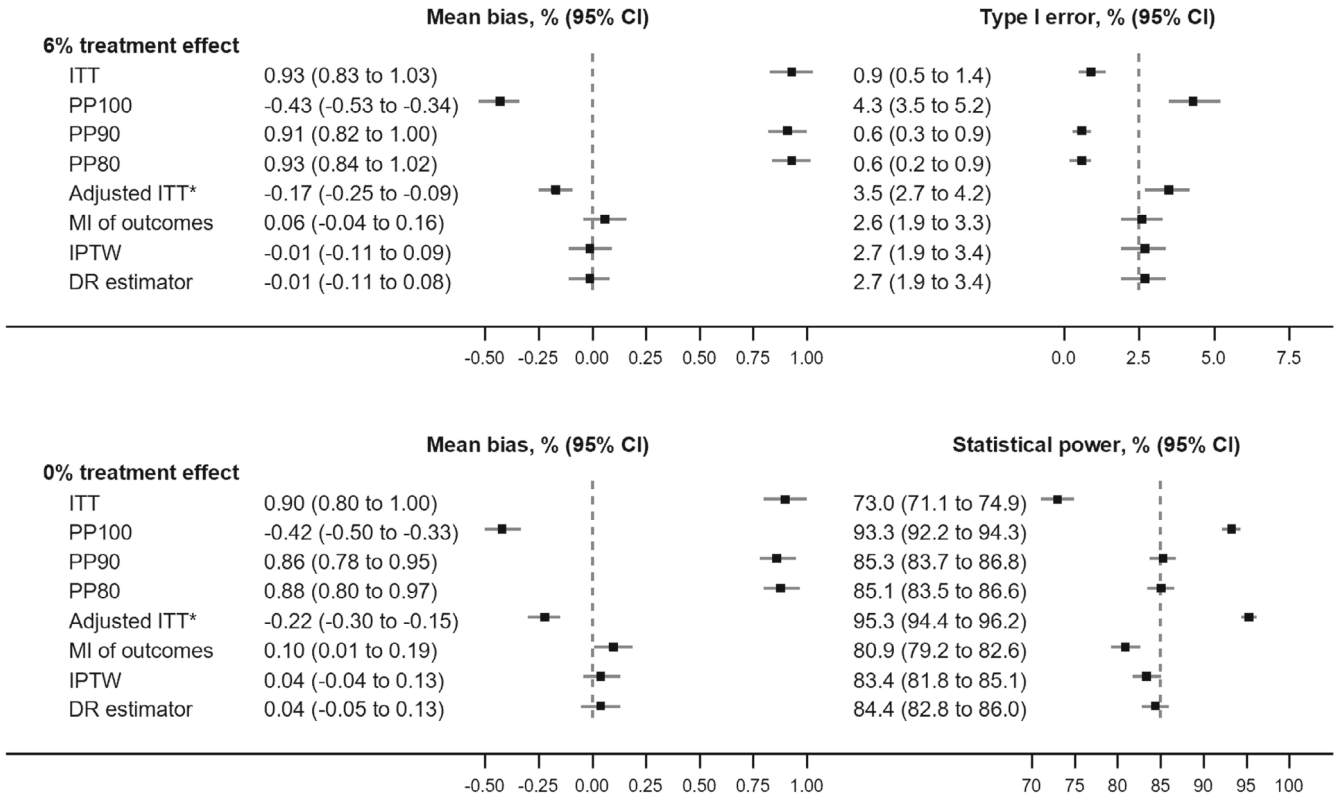

FIGURE 3 Mean bias, type I error, and statistical power of each analysis method assuming worse adherence in the experimental group than the control group, no unobserved confounding of the adherence-outcome association, and no treatment-adherence interaction. CI, confidence interval; DR, doubly robust; IPTW, inverse-probability-of-treatment weighting; ITT, intention-to-treat; MI, multiple imputation; PP100, per-protocol defined as receiving 100% of prescribed doses; PP80, per-protocol defined as receiving at least 80% of prescribed doses; PP90, per-protocol defined as receiving at least 90% of prescribed doses. * Adjusted for observed levels of adherence as a covariate.

Tables S10–S18). For example, the mean bias for the ITT analysis was ~1.2% under all three treatment effects (0%, 3%, and 6%) when the treatment-adherence interaction was combined with similar quantities of adherence between trial arms; the corresponding type I error estimate was roughly eight times smaller than the desired 2.5% and power was estimated to be 16% lower than the desired value of 85% (Tables S10–S12). The ITT, PP, and adjusted ITT analyses were most biased in scenarios that combined the treatment-adherence interaction with worse adherence in the EXP group than the CON group (Tables S16–S18). In contrast, the MI, IPTW, and DR approaches were unbiased and maintained type I error rates close to 2.5% in most of the scenarios explored, resulting in estimates of power that were at least 83%. Results for these three methods were also comparable in the presence of unobserved confounding of the adherence-outcome association, though a small amount of bias remained in some scenarios (Tables S19–S21).

4.6 | Misspecification of the adherence and outcome models (no unobserved confounding or treatment-adherence interaction)

When the adherence and outcome models were misspecified by omitting interaction terms, the IPTW and DR methods remained unbiased and maintained type I error rates close to 2.5% regardless of the assumed level of adherence in the EXP group (Tables S22–S24). In contrast, the MI approach remained unbiased when the quantities of non-adherence were similar in the EXP and CON groups but tended to result in a small amount of bias when adherence levels differed between trial arms (range: −0.23%–0.21%), which consequently under- or over-inflated the type I error rate. All three methods resulted in estimates of power that were at least 83%.

5 | DISCUSSION

Using computer simulations to replicate a real non-inferiority trial with an active control regimen, this study found that when non-adherence differed between trial arms, ITT and PP analyses often resulted in non-trivial bias in the estimated treatment effect which consequently under- or over-inflated the type I error rate. Depending on the patterns of non-adherence, it was possible for ITT and PP analyses to be biased in the same direction by similar amounts, and, generally, the presence of a treatment-adherence interaction exacerbated the issues with these approaches. Adjustment for observed adherence led to similar issues, whereas the MI, IPTW and DR methods were able to correct bias under most non-adherence scenarios but could not always eliminate bias entirely in the presence of unobserved confounding. The bias correction applied as part of the MI approach appeared to work well, but not perfectly; the inclusion of higher order terms from the Taylor series expansion of the logit function may further reduce the magnitude of any bias observed. IPTW and the DR estimator were generally unbiased, maintained desired type I error rates, and did not result in any meaningful losses of statistical power.

Our results are consistent with those of Mo et al. despite differences with the study designs.[5] For instance, when differential non-adherence occurred due to confounding factors which increased the probability of an adverse outcome, Mo and colleagues also found that ITT and PP analyses tended to result in bias that frequently occurred in the same direction. IPTW and IV estimation were able to minimize this bias, but the former approach could only eliminate it entirely if all confounders were appropriately adjusted for. IV methods were not considered in the current study since standard approaches are only able to account for treatment crossovers, a different type of non-adherence to that considered in our study.

Other simulation studies have assessed the performance of RPSFTM and G-estimation for handling treatment crossovers in non-inferiority trials with time-to-event outcomes compared with conventional survival analyses.[13,14] While RPSFTM and G-estimation is not an appropriate analysis method for the binary outcomes considered in our study design, these simulations were consistent with our results and showed that in the presence of differential non-adherence, ITT and PP analyses can result in either conservative or anti-conservative type I error rates. Similar to our conclusions, the authors of both studies deduced that in non-inferiority trials with treatment non-adherence, neither ITT or PP analyses can guarantee the validity of non-inferiority conclusions.

Although IPTW and the DR estimator were the best performing methods in this study, these techniques have some important limitations. First, they can eliminate bias if all confounders can be appropriately adjusted for, but this will often not be possible. Potential confounders of the adherence-outcome association need to be carefully considered at the design stage of trials, relevant data collected as fully as possible, and their effects modeled correctly when predicting adherence and outcomes. For IPTW, misspecification of either of these models, such as the omission of important interactions or incorrectly specifying the functional form of covariates, may lead to treatment groups that are imbalanced with regard to potential confounders and consequently bias in the estimated treatment effect.[25] A key advantage of the DR estimator is that only one of the models used to predict adherence and outcomes needs to be correctly specified, but not necessarily both. Finally, IPTW and the DR methods cannot be used if there are strata of the covariates for which all participants are fully adherent (violating the positivity assumption). To overcome this issue, we imputed a large predicted probability of receiving all doses for individuals within such strata (probability of 0.99), which performed well in the simulations.

ITT analyses play an important role in estimating treatment effects in clinical practice, but we have shown that they pose several limitations in the context of non-inferiority trials. They rely on strong assumptions, such as trial adherence being reflective of real-world behavior, yet adherence is frequently observed to be better in trials.[26] Consequently, non-inferiority conclusions based on an ITT approach may not apply to real-world scenarios with different adherence patterns. In addition, a particular concern in non-inferiority trials is that non-adherence can dilute estimated treatment effects and increase the risk of type I errors (accepting an inferior intervention). In our study, using an ITT analysis inflated the risk of a type I error in some scenarios, particularly when adherence to the experimental regimen was better than the comparator. Instead, we advocate for estimating treatment effects under full adherence to ensure that non-inferiority exists when participants receive their interventions as intended, avoiding any bias caused by non-adherence. This estimand translates easily from a trial to real-world setting and is likely to be of greater interest to patients and healthcare professionals than ITT effects dependent on observed adherence.

In future TB trials, shorter regimens and novel approaches to adherence monitoring may result in improved adherence.[27] One would expect adherence levels to improve as duration of treatment shortens, in which case any bias in ITT and PP analyses should diminish. However, some non-adherence is likely even with shorter regimens, as was seen in a more recent short-treatment TB trial.[28] Therefore, the ITT and PP results of trials assessing shorter regimens are not

guaranteed to be free from bias caused by non-adherence. In addition, recent developments in digital adherence technologies such as digital pillboxes show mixed evidence that they may lead to improved adherence.[29,30] However, they do not guarantee full adherence and so the issues with ITT and PP analyses highlighted by the current study are likely to persist.[31-33] Our findings are also of interest beyond TB trials, where non-adherence to treatment is likely to remain a concern for the foreseeable future.

The current study has several strengths including the use of a real non-inferiority trial to inform the design, the direct comparison of different statistical methods, and the range of non-adherence scenarios explored. However, it has some limitations. First, our simulation study relies on assumptions made in the data generation process and analysis, albeit we feel that these assumptions are plausible. Second, we calculated adherence using the overall percentage of prescribed doses received. Other patterns of adherence, such as the timing of missed doses, may be important.[34,35] Third, adherence was dichotomized (100% vs <100% of doses received) before being included in the models used to predict adherence and outcomes. It is plausible that utilizing a continuous functional form of adherence may improve the ability of the more sophisticated methods to eliminate bias due to treatment non-adherence. For instance, using fractional polynomials or generalized propensity score methods.[36,37] Future work should assess the additional benefits of these approaches. In addition, further research should compare the performance of different statistical methods for handling treatment non-adherence in non-inferiority trials with active control regimens and time-to-event outcomes (where treatment crossovers are often not permitted).

6 | CONCLUSION

In non-inferiority trials where treatment non-adherence differs between arms, ITT and PP analyses can produce biased estimates of efficacy that can occur in the same direction, potentially leading to the acceptance of inferior treatments or efficacious regimens being missed. IPTW and the DR estimator, which are relatively straightforward methods to implement, were able to correct bias under most non-adherence scenarios and should be used to supplement ITT and PP approaches in ongoing non-inferiority trials with active control regimens and binary outcomes. Future studies should ultilize more sophisticated methods for handling non-adherence in the primary analysis with a number of sensitivity analyses conducted (including ITT and PP approaches) in order to explore the impacts of their different assumptions.

ACKNOWLEDGEMENTS

We would like to thank the Critical Path (C-Path) Institute and TB-PACTS program for granting us access to data from the REMoxTB trial. This research was funded in whole, or in part, by the Wellcome Trust [Senior Research Fellowship 224485/Z/21/Z]. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Matthew Dodd https://orcid.org/0000-0002-6207-6604
James Carpenter https://orcid.org/0000-0003-3890-6206
Jennifer A. Thompson https://orcid.org/0000-0002-3068-3952
Elizabeth Williamson https://orcid.org/0000-0001-6905-876X
Katherine Fielding https://orcid.org/0000-0002-6524-3754
Diana Elbourne https://orcid.org/0000-0003-3044-4545

REFERENCES

1. Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy*. 2011;31(9):833-839.
2. Bikdeli B, Welsh JW, Akram Y, et al. Noninferiority designed cardiovascular trials in highest-Impact Journals. *Circulation*. 2019;140(5):379-389.
3. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;308(24):2594-2604.

4. Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. *Trials*. 2012;13:84.
5. Mo Y, Lim C, Mukaka M, Cooper BS. Statistical considerations in the design and analysis of non-inferiority trials with binary endpoints in the presence of non-adherence: a simulation study. *Wellcome Open Res*. 2019;4:207.
6. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med Care*. 2004;42(3):200-209.
7. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ Open*. 2014;4(6):e005362.
8. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006;295(10):1152-1160.
9. Matilde Sanchez M, Chen X. Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Stat Med*. 2006;25(7):1169-1181.
10. Brittain E, Lin D. A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Stat Med*. 2005;24(1):1-10.
11. Clarke PS, Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. *Biostatistics*. 2010;11(4):756-770.
12. Dodd M, Fielding K, Carpenter JR, Thompson JA, Elbourne D. Statistical methods for non-adherence in non-inferiority trials: useful and used? A systematic review. *BMJ Open*. 2022;12(1):e052656.
13. Matsuyama Y. A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. *Stat Med*. 2010;29(20):2107-2116.
14. Wu Y, Zhao L, Hou Y, Li K, Zhou X. Correcting for non-compliance in randomized non-inferiority trials with active and placebo control using structural models. *Stat Med*. 2015;34(6):950-965.
15. Alshreef A, Latimer N, Tappenden P, et al. Statistical methods for adjusting estimates of treatment effectiveness for patient nonadherence in the context of time-to-event outcomes and health technology assessment: a systematic review of methodological papers. *Med Decis Mak*. 2019;39(8):910-925.
16. Gillespie SH, Crook AM, McHugh TD, et al. Four-month Moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med*. 2014;371(17):1577-1587.
17. Merle CS, Fielding K, Sow OB, et al. A four-month gatifloxacin-containing regimen for treating tuberculosis. *N Engl J Med*. 2015;372(17):1677.
18. Jindani A, Harrison TS, Nunn AJ, et al. High-dose Rifapentine with Moxifloxacin for pulmonary tuberculosis. *N Engl J Med*. 2014;371(17):1599-1608.
19. Imperial MZ, Nahid P, Phillips PPJ, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nat Med*. 2018;24(11):1708-1715.
20. Garcia-Cremades M, Solans BP, Strydom N, et al. Emerging therapeutics, technologies, and drug development strategies to address patient nonadherence and improve tuberculosis treatment. *Annu Rev Pharmacol Toxicol*. 2022;62:197-210.
21. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074-2102.
22. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH harmonised guideline: addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials: E9(R1). 2019 https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf
23. Pham TM, Tweed CD, Carpenter JR, et al. Rethinking intercurrent events in defining estimands for tuberculosis trials. *Clin Trials*. 2022;19(5):522-533.
24. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: Wiley; 2004.
25. Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J*. 2021;15(1):14-20.
26. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med*. 2005;353(5):487-497.
27. Zaidi HA, Wells CD. Digital health technologies and adherence to tuberculosis treatment. *Bull World Health Organ*. 2021;99(5):323-323a.
28. Paton NI, Cousins C, Suresh C, et al. Treatment strategy for rifampin-susceptible tuberculosis. *N Engl J Med*. 2023;388(10):873-887.
29. Alipanah N, Jarlsberg L, Miller C, et al. Adherence interventions and outcomes of tuberculosis treatment: a systematic review and meta-analysis of trials and observational studies. *PLoS Med*. 2018;15(7):e1002595.
30. Mohamed MS, Zary M, Kafie C, et al. The Impact of Digital Adherence Technologies on Health Outcomes in Tuberculosis: A Systematic Review and Meta-Analysis *medRxiv*. 2024 2024.2001.2031.24302115.
31. Acosta J, Flores P, Alarcón M, Grande-Ortiz M, Moreno-Exebio L, Puyen ZM. A randomised controlled trial to evaluate a medication monitoring system for TB treatment. *Int J Tuberc Lung Dis*. 2022;26(1):44-49.
32. Liu X, Lewis JJ, Zhang H, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. *PLoS Med*. 2015;12(9):e1001876.
33. Liu X, Thompson J, Dong H, et al. Digital adherence technologies to improve tuberculosis treatment outcomes in China: a cluster-randomised superiority trial. *Lancet Glob Health*. 2023;11(5):e693-e703.
34. Stagg HR, Flook M, Martinecz A, et al. All nonadherence is equal but is some more equal than others? Tuberculosis in the digital era. *ERJ Open Res*. 2020;6(4):00315-2020.
35. Fox WS, Strydom N, Imperial MZ, Jarlsberg L, Savic RM. Examining nonadherence in the treatment of tuberculosis: the patterns that lead to failure. *Br J Clin Pharmacol*. 2023;89(7):1965-1977.

36. Stagg HR, Thompson JA, Lipman MCI, Sloan DJ, Flook M, Fielding KL. Forgiveness is the attribute of the strong: nonadherence and regimen shortening in drug-sensitive tuberculosis. *Am J Respir Crit Care Med*. 2023;207(2):193-205.
37. Austin PC. Assessing the performance of the generalized propensity score for estimating the effect of quantitative or continuous exposures on binary outcomes. *Stat Med*. 2018;37(11):1874-1894.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dodd M, Carpenter J, Thompson JA, Williamson E, Fielding K, Elbourne D. Assessing efficacy in non-inferiority trials with non-adherence to interventions: Are intention-to-treat and per-protocol analyses fit for purpose?. *Statistics in Medicine*. 2024;43(12):2314-2331. doi: 10.1002/sim.10067