

BMJ Open Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study

Chenglin Ye,^{1,2} Joseph Beyene,¹ Gina Browne,^{1,3} Lehana Thabane^{1,2}

To cite: Ye C, Beyene J, Browne G, *et al*. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ Open* 2014;**4**:e005362. doi:10.1136/bmjopen-2014-005362

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-005362>).

Received 30 March 2014
Revised 27 May 2014
Accepted 2 June 2014



CrossMark

¹Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

²Biostatistics Unit, St Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada

³School of Nursing, McMaster University, Hamilton, Ontario, Canada

Correspondence to
Dr Lehana Thabane;
thabanl@mcmaster.ca

ABSTRACT

Objective: Randomised controlled trials (RCTs) are often considered as the gold standard for assessing new health interventions. Patients are randomly assigned to receive an intervention or control. The effect of the intervention can be estimated by comparing outcomes between groups, whose prognostic factors are expected to balance by randomisation. However, patients' non-compliance with their assigned treatment will undermine randomisation and potentially bias the estimate of treatment effect. Through simulation, we aim to compare common approaches in analysing non-compliant data under different non-compliant scenarios.

Settings: Based on a real study, we simulated hypothetical trials by varying three non-compliant factors: the type, randomness and degree of non-compliance. We compared the intention-to-treat (ITT), as-treated (AT), per-protocol (PP), instrumental variable (IV) and complier average casual effect (CACE) analyses to estimate large (50% improvement over the control), moderate (25% improvement) and null (same as the control) treatment effects. Different approaches were compared by the bias of estimate, mean square error (MSE) and 95% coverage of the true value.

Results: For a large or moderate treatment effect, the ITT estimate was considerably biased in all scenarios. The AT, PP, IV and CACE estimates were unbiased when non-compliant behaviours were random. The IV estimate was unbiased when non-compliant behaviours were symmetrically dependent on patients' conditions. The PP estimate was mostly unbiased when patients in the control group did not have access to the intervention. When the intervention was not different from the control, the ITT was less biased than the other approaches. Similar results were found when comparing the MSE and 95% coverage.

Conclusions: The standard ITT analysis under non-compliance is biased when the intervention has a moderate or large effect. Alternative analyses can provide unbiased or less biased estimates. Based on the results, we make some suggestions on choosing optimal approaches for analysing specific non-compliant scenarios.

BACKGROUND

Randomised controlled trials (RCTs) are often considered as the gold standard for

Strengths and limitations of this study

- We compared different methods to analyse non-compliant data by simulating hypothetical randomised controlled trials.
- Different non-compliant scenarios were generated by three factors: the type, randomness and degree of non-compliance.
- The simulation framework and parameters were built on a real study.
- Patients' prognostic factors and missing data due to withdrawal were not considered in the simulation.

assessing new health interventions where patients are randomly assigned to receive an intervention or control (eg, placebo). Since patients' prognostic factors are expected to balance by randomisation patients' outcomes can be directly compared between groups to infer the effect of a treatment. In many cases, patients may not fully comply with their assigned treatment according to the protocol. Such protocol violation compromises the 'fair' comparison, which is protected by randomisation, and will potentially bias the estimate of treatment effect. Analysing RCTs subject to non-compliance can be challenging. While different analyses have been proposed to deal with non-compliance, the bias of treatment effect estimate is rarely compared among different approaches. Result interpretations also vary depending on the nature of non-compliance and the objective of a trial. Some RCTs, known as pragmatic trials,^{1–3} are primarily designed to guide clinical practice. Their goal is often to assess whether an intervention will work in routine practice. In contrast, non-pragmatic trials usually focus on the biological efficacy of an intervention. Despite the objective, an analysis that provides an unbiased or less biased estimate of treatment effect is always desirable. In this study, we compare common approaches to analyse

non-compliant data in RCTs. The results will provide useful knowledge in choosing optimal methods for different non-compliant scenarios.

This study was motivated by an RCT that compared the integrated care organised through the Children's Treatment Network (CTN) with the usual care directed by parents for managing children with special health-care needs.⁴ The CTN coordinated community resources to deliver comprehensive health services for the target children and their families. The hypothesis was that the target children's health outcomes would improve by receiving integrated, proactive and necessary services tailored for them. While the use of RCTs in assessing CTN-like interventions has been promising, non-compliant rates are generally high in these trials. This is largely due to the complexity of implementing multidisciplinary interventions in real-life settings.

Intention to treat (ITT), as treated (AT) and per protocol (PP) are common approaches to analyse non-compliant data in RCTs. The ITT analysis is considered as the gold standard⁵ but can be problematic for some scenarios.⁶ Multiple methods are often recommended for analysing RCTs with substantial non-compliance.⁷⁻⁹ A literature review¹⁰ randomly selected 100 RCTs published in high impact journals in 2008. Of 98 RCTs which reported non-compliance, 46 employed variations of PP analyses in addition to an ITT analysis. Another class of methods to deal with non-compliance includes instrumental variable (IV) and complier average causal effect (CACE) approaches.¹¹⁻¹⁵ A conceptual difference among all these methods is that the ITT, IV and CACE approaches estimate treatment effects by preserving randomisation or accounting for potential confounding, but the AT and PP approaches do not. There are also other proposed approaches to correct for non-compliance in RCTs,¹⁶ for example, G-estimation and inverse probability weighted estimators. However, these methods have not been widely adopted. Therefore, we only included the ITT, AT, PP, IV and CACE methods in our comparison.

Very few studies have compared these five methods on the bias of estimating treatment effects. Bang and Davis¹⁷ had compared ITT, AT, PP and IV methods. They showed that ITT and IV analyses were biased in certain non-compliant cases. However, the authors did not include CACE analysis in their comparison and did not consider the situation where there was no crossover between treatment groups. This scenario is common when a new intervention is only accessible to patients who are offered it. In another study, McNamee¹⁸ compared ITT, AT, PP and IV analyses and concluded that an ITT analysis was not always biased towards the null while AT and PP analyses were generally biased. Sheng and Kim¹⁹ investigated the effect of non-compliance on ITT analysis of equivalence trials and showed that non-compliance did not always favour the null hypothesis, that is, no difference between treatment groups. Hertogh *et al*²⁰ concluded that the IV method could give insight into confounding by non-compliance in RCTs.

Most of the previous studies did not consider different associations of non-compliant behaviours with patients' conditions. For example, patients with certain characteristics may always reject a new intervention. Also, there are partial non-compliant cases where patients receive parts of the intervention even if they did not fully comply with the protocol. In our simulation study, we considered additional non-compliant scenarios that were not considered by previous studies and compared the five common methods by the bias of estimate, the mean square error (MSE) and 95% coverage of the true value. Our objectives were to compare the performance of different approaches in analysing non-compliant RCT data and make recommendations on optimal approaches under specific scenarios.

METHODS

Simulation framework

In the CTN trial, over 50% of the children randomised in the CTN group did not fully comply with the intervention for various reasons. Primary and sensitivity analyses showed that the effect of the CTN was not significant but the estimates varied in direction, magnitude and precision.⁴ This observation prompted us to further investigate the impact of non-compliance on estimating treatment effects.

On the basis of the CTN setting, we simulated hypothetical RCTs where patients were randomly assigned to the intervention or usual care by a 1:1 allocation ratio. The parameters for generating hypothetical patients were estimated from the CTN trial. We simulated different non-compliant scenarios by varying three factors: (1) the type of non-compliers, (2) the randomness of non-compliance and (3) the degree of non-compliance. Our simulation framework is shown in figure 1. The design, conduct and reporting of this study has followed the guideline of designing and reporting simulation studies.²¹

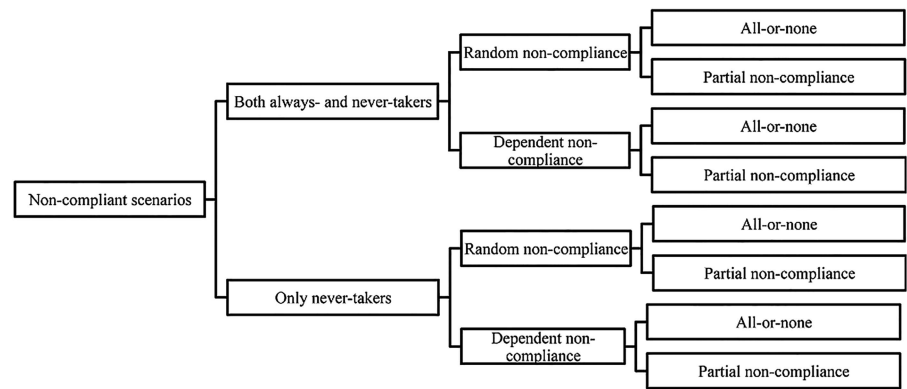
Type of non-compliers

We considered two types of non-compliers which were defined as: never-takers and always-takers.²² Never-takers are patients who will always reject a new intervention if they are offered it. Always-takers will always receive a new intervention even if they are not offered it. Two scenarios were considered. In one scenario, we assumed that non-compliers were either never-takers or always-takers, which mimicked the situation where patients were able to get the intervention elsewhere even if they were not offered it. In the other scenario, we assumed that non-compliers were only never-takers, which mimicked the situation where the intervention was only accessible to patients who were offered it. In addition, we assumed that the intervention and usual care were the only treatment options.

Randomness of non-compliance

Non-compliant behaviours could be random or dependent on patients' conditions. In particular, we considered

Figure 1 The simulation framework.



six scenarios of dependent non-compliant behaviours that were studied by McNamee¹⁸:

- Patients with good conditions would always get the intervention while patients with poor conditions would always reject it;
- Patients with good conditions would always get the intervention;
- Patients with poor conditions would always reject the intervention;
- Patients with good conditions would always reject the intervention while patients with poor conditions would always get it;
- Patients with good conditions would always reject the intervention;
- Patients with poor conditions would always get the intervention.

Patients' conditions were considered to be positively associated with their outcomes under usual care. We assumed that good condition represented an outcome score of at least 0.5 SDs above the group mean under usual care (assuming that a high score was a better outcome). Poor condition represented an outcome score of at least 0.5 SDs below the group mean under usual care. When there were no always-takers, only scenarios C and E were considered.

Degree of non-compliance

Degree of non-compliance referred to the proportion of interventional components that a patient did not receive according to the protocol. The simplest case was all-or-none where compliers received 100% of the intervention and non-compliers received none of it. For multifaceted interventions, patients were likely to receive some components of the intervention even if they did not fully comply with the protocol. In addition, patients might only receive parts of the intervention because of the intervention fidelity. For example, a systematic review showed that many interventions of integrated care did not actually deliver all services as planned due to complexity of implementation.²³ Conversely, we also use the term 'degree of compliance' for the proportion of interventional components that a patient received according to the protocol. Non-compliance and compliance are used throughout this paper.

All-or-none and partial non-compliance were considered in our simulation. For all-or-none case, we considered two compliance levels: all components ($d=1$) or none ($d=0$) of the intervention. A study reported that non-compliance rate could be as high as 30–40% for a treated population.²⁴ Therefore, we randomly selected 30% of patients to receive a treatment opposite to what they were assigned for. For partial non-compliance, we considered four compliance levels: none ($d=0$), one-third ($d=1/3$), two-thirds ($d=2/3$) or all ($d=1$) components of the intervention. These four levels have been studied in a previous simulation study.¹⁷

Simulation procedures

We employed a modified simulation model from the previous study.¹⁷ Let Y_1 and Y_0 be a pair of counterfactual outcomes for a patient if he or she were in the intervention and the usual care groups, respectively. In practice, we can only observe one of the counterfactual outcomes because we can never observe both outcomes for any patient at the same time. Thus, causal inference is often made at population level instead of patient level. By adopting a marginal view, we define the causal effect for the treatment of interest (δ) as

$$\delta = \mu_1 - \mu_0,$$

where μ_1 and μ_0 were the means of Y_1 and Y_0 , respectively.

We chose $\mu_0=59$ to be the effect of usual care. The effect of usual care was estimated from the CTN trial with an SD of 10. For the effect of treatment (μ_1), we chose three different cases: $\mu_1=89$ for a 50% improvement over the usual care; $\mu_1=74$ for a 25% improvement over the usual care and $\mu_1=59$ for no difference from the usual care. Each case was simulated separately. We then generated individual patient's counterfactual outcomes through a normal distribution:

$$Y_k \sim \text{Normal}(\mu_k, 10^2) \text{ and } k = 0, 1.$$

Thus, good condition was defined for a patient with $Y_0 > 64$ (half SDs above the group mean under usual care) and poor condition was defined for a patient with $Y_0 < 54$ (half SDs below the group mean under usual

care). A group indicator Z (1=intervention and 0=usual care) was generated for each patient from a Bernoulli distribution with equal probability of 0.5 of being assigned to either group. The observed outcome for a patient was calculated by

$$y_i = d_i Y_1 + (1 - d_i) Y_0,$$

where d_i was the degree of treatment compliance with the protocol for patient i . For all-or-none case, d was either 1 or 0. For partial compliance, d took a value of 0, 1/3, 2/3 or 1.

In the CTN trial, 450 patients were needed to detect a minimum clinically important difference (MCID) of 15 with 80% statistical power and 5% α . Using the same MCID, we chose to simulate 500 participants in each hypothetical trial. We estimated the SE of the treatment effect to be 1.53 from the CTN trial. Based on this estimate, at least 816 simulations were needed to produce an effect estimate within 1% accuracy of the MCID by the standard formula.²¹ To have sufficient power, we chose to generate 1000 simulations per scenario. The steps of simulation are shown in figure 2.

Statistical analysis

This section describes the different methods that we compared. The estimated treatment effect was expressed as the difference of the mean score between groups. The methods were compared by the bias, MSE and 95% coverage.²¹ Bias is defined as

$$\text{Bias} = \bar{\hat{\delta}} - \delta,$$

where $\bar{\hat{\delta}}$ is the average estimate of interest over all iterations and δ is the true value. MSE is calculated by

$$\text{MSE} = (\bar{\hat{\delta}} - \delta)^2 + (\text{SE}(\hat{\delta}))^2,$$

where $\text{SE}(\hat{\delta})$ is the SD of the empirical distribution of the estimates from all iterations. Out of all iterations, 95% coverage is the number of times the 95% CIs include δ . CI is calculated by normal approximation.

Intention to treat

In the ITT approach, patients are analysed by how they were randomised regardless of their actual compliance with treatment. The treatment effect was estimated by

$$\hat{\delta}_{\text{ITT}} = \bar{y}_1 - \bar{y}_0$$

where \bar{y}_1 and \bar{y}_0 were the mean outcome scores of the intervention and usual care groups, respectively.

As treated

The AT approach compares patients by the treatment they actually received. The treatment effect was estimated by

$$\hat{\theta}_{\text{AT}} = \bar{y}_{\text{treated}} - \bar{y}_{\text{untreated}}.$$

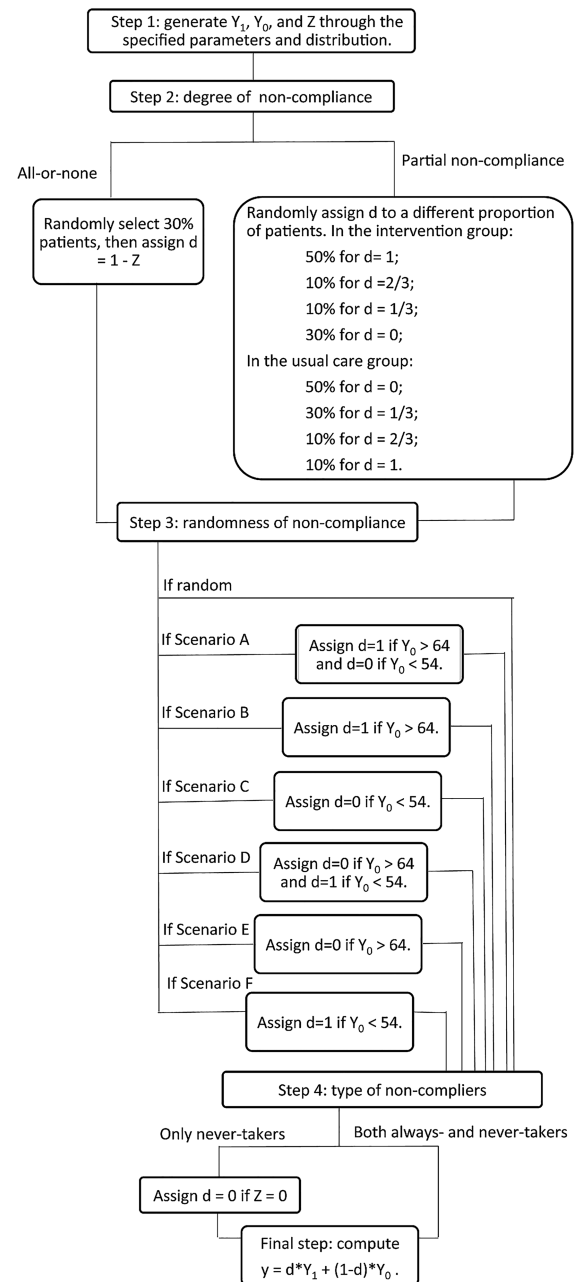


Figure 2 Summary of the simulation steps. Y_1 =the counterfactual outcome for a patient in the intervention group. Y_0 =the counterfactual outcome for a patient in the usual care group. Z =randomisation indicator. d =the degree of compliance with the intervention. y =the simulated outcome for a patient. Scenario A: patients with good conditions will always get the intervention and those with poor conditions will always reject it. Scenario B: patients with good conditions will always get the intervention. Scenario C: patients with poor conditions will always reject the intervention. Scenario D: patients with good conditions will always reject the intervention and those with poor conditions will always get it. Scenario E: patients with good conditions will always reject the intervention. Scenario F: patients with poor conditions will always get the intervention.

For patients with partial compliance, the treatment effect was estimated by regressing the degree of compliance (d) on the outcome in a linear regression model.

Also, a different notation, $\hat{\theta}$, was used to differentiate from the estimators of making causal inferences.

Per protocol

The PP approach excludes patients who did not fully comply with treatment protocol. The treatment effect was estimated by

$$\hat{\theta}_{PP} = \bar{y}_{\text{observed compliers in the intervention}} - \bar{y}_{\text{observed compliers in the control}}.$$

Instrumental variable

The IV approach employs the randomisation indicator (Z) as an IV to adjust for the proportion of non-compliant patients. The theory and assumptions of IV analysis have been thoroughly discussed in the literature.^{12 13} We used the standard IV estimator for linear models:

$$\hat{\delta}_{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{d}_1 - \bar{d}_0}.$$

The Fieller's theorem was used to calculate the SE of the estimate.¹⁷

Complier average causal effect

The CACE method estimates the treatment effect among compliers. The assumptions and the casual framework of CACE have been discussed elsewhere.^{14 15 25} The treatment effect was estimated by

$$\hat{\delta}_{CACE} = E[\bar{y}_1 - \bar{y}_0 | \text{compliers}].$$

There are two general approaches to the CACE inference²⁵: the maximum likelihood approach by expectation-maximisation (EM) algorithm and the Bayesian approach. The CACE analysis in this paper was conducted in Mplus (V.7; Mac OS X 10.6.8) Los Angeles, CA, Muthén & Muthén, which employed the EM algorithm. The rest of the analyses and simulations were performed in R V.2.15.2.

Cut-off points for non-compliance

In practice, investigators often dichotomise patients to be either compliers or non-compliers. A cut-off of 80% is commonly used.^{9 10} Patients are considered to be compliers if they have complied with at least 80% of the intervention according to protocol. A cut-off of 100% has also been used such that patients are considered to be compliers only if they have complied with the entire treatment protocol. Compliers are expected to receive the full effect of an intervention. We conducted a sensitivity analysis to investigate the impact of these two cut-offs on dichotomising compliers. A new compliance indicator for patient *i* was defined as: $t_i = I(d_i \geq 0.8)$ for a cut-off of 80% and $t_i = I(d_i = 1)$ for a cut-off of 100%. The indicator function *I* returned 1 if the condition was satisfied and 0 otherwise. We then performed the same analysis by replacing d_i with t_i for patient *i*.

RESULTS

The estimates by different analyses under the simulated scenarios, and their bias, MSE and 95% coverage are summarised in tables 1 and 2. For a large treatment effect (a mean difference of 30, representing 50% improvement over the usual care), the ITT estimate was considerably biased. The other estimates were unbiased when non-compliant behaviours were random. When there were always-takers and never-takers, the PP and CACE estimates were the least biased if never-taking behaviours were dependent on patients' conditions (scenarios C and E). The IV estimate was the least biased if non-compliant behaviours were symmetrically dependent on patients' conditions (scenarios A and D). When there were only never-takers, the PP estimates were mostly unbiased. Similar results were found for a moderate treatment effect (details not shown). When an intervention was not different from the usual care (details not shown), all estimates were unbiased when non-compliant behaviours were random. The ITT estimate was also unbiased if non-compliant behaviours were symmetrically dependent on patients' conditions. For the remaining scenarios, the ITT estimate was the least biased and the other estimates were biased with a different degree.

The IV estimate generally had a larger MSE than the other estimates. That was because the standard IV estimator was sensitive to non-compliant rates. For example, when non-compliant rate was equal between groups, the denominator was zero and the estimate became undefined. When never-takers and always-takers were considered, the ITT and IV approaches generally had a better 95% coverage than the other approaches. When there were only never-takers, the PP approach had the best 95% coverage. For a large treatment effect, the ITT approach had zero 95% coverage. Overall, the results from comparing the MSE and 95% coverage were consistent with those from comparing the bias of estimates.

In the sensitivity analysis, we compared the impact of using a cut-off of 80% or 100% to dichotomise compliant patients. The results showed that dichotomising patients by a cut-off of 80% resulted in less biased estimates than dichotomising patients by a cut-off of 100%. For a null treatment effect, the treatment estimates obtained by applying a cut-off of 80% were less biased than those obtained by directly analysing patients on the observed degree of compliance.

DISCUSSION

Through simulation, we compared different methods of analysing non-compliant RCT data. Our results showed that the ITT approach was the most optimal when estimating a null effect since it provided an unbiased or the least biased estimate in different scenarios. This result was consistent with the general opinion that the ITT estimate is conservative towards the null. However, for the case of a large or moderate treatment effect, the ITT

Table 1 Summary of the results when never-takers and always-takers were allowed (treatment effect=30)

Scenario	Method	All or none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage (%)	Estimate	Bias	MSE	Coverage (%)
Random	ITT	12	-18	330	0	10	-20	403	0
	AT	30	0	1	96	30	0	1	95
	PP	30	0	1	95	30	0	2	96
	IV	30	0	5	96	30	0	7	95
	CACE	30	0	1	95	30	0	2	95
A	ITT	5	-25	651	0	4	-26	689	0
	AT	37	7	50	0	37	7	56	0
	PP	36	6	38	0	37	7	50	0
	IV	29	-1	80	96	28	-2	129	96
	CACE	36	6	38	0	37	7	50	0
B	ITT	10	-20	416	0	8	-22	483	0
	AT	35	5	27	0	35	5	27	0
	PP	35	5	27	0	35	5	27	1
	IV	35	5	34	60	35	5	38	66
	CACE	35	5	27	0	35	5	27	0
C	ITT	7	-23	539	0	6	-24	590	0
	AT	33	3	8	18	34	4	19	2
	PP	31	1	3	80	33	3	9	48
	IV	25	-5	41	73	25	-5	46	80
	CACE	31	1	3	79	33	3	9	47
D	ITT	5	-25	645	0	4	-26	690	0
	AT	23	-7	50	0	23	-7	56	0
	PP	24	-6	39	0	23	-7	51	0
	IV	31	1	174	97	34	4	4895	96
	CACE	24	-6	39	0	23	-7	51	0
E	ITT	10	-20	413	0	8	-22	483	0
	AT	27	-3	8	18	26	-4	19	4
	PP	29	-1	3	78	27	-3	9	50
	IV	35	5	42	74	35	5	49	81
	CACE	29	-1	3	78	27	-3	9	48
F	ITT	7	-23	537	0	6	-24	591	0
	AT	25	-5	26	0	25	-5	27	0
	PP	25	-5	27	0	25	-5	27	0
	IV	25	-5	35	58	25	-5	39	66
	CACE	25	-5	27	0	25	-5	27	0

AT, as treated; CACE, complier average casual effect; ITT, intention to treat; IV, instrumental variable; MSE, mean square error; PP, per protocol.

approach was much more biased than the other approaches. When patients' non-compliant behaviours were purely random, the AT, PP, IV and CACE approaches all provided unbiased estimates. For other non-compliant scenarios that we considered, the choice of optimal method varied. Figure 3 summarises the choices of methods under different scenarios to produce an unbiased or less biased estimate.

Although the ITT method is the most commonly reported analysis, other analyses of non-compliant data may provide a better estimate. Thus, understanding the extent of bias for different analyses is important when choosing an optimal approach and interpreting the results. Our results are limited by a number of factors. First, we did not consider specific prognostic factors in the simulation. Adjusting for prognostic factors may improve the estimation of treatment effect. However, we

did consider different associations between patients' outcomes and non-compliant behaviours. Second, we assumed that the clinical effect of an intervention was proportional to the degree of compliance. This linear association might not represent all real-life situations. Third, we did not consider missing data in the simulation and assumed that non-compliers' outcomes were still collectable. Alternatively, imputation techniques can be applied to handle missing data.²⁶ Finally, we only simulated a subset of general non-compliant scenarios. Thus, our findings may not be generalisable to other scenarios.

Despite the limitations, our study has several strengths. The simulation framework was built on three key factors of non-compliance: the type of non-compliers, the randomness of non-compliance and the degree of non-compliance. These three factors were not considered simultaneously in previous studies. We generated a total

Table 2 Summary of the results when only never-takers were allowed (treatment effect=30)

Scenario	Method	All or none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage (%)	Estimate	Bias	MSE	Coverage (%)
Random	ITT	18	-12	145	0	18	-12	145	0
	AT	28	-2	7	36	30	0	1	94
	PP	28	-2	4	65	30	0	1	94
	IV	30	0	2	95	30	0	2	95
	CACE	30	0	2	95	33	3	12	36
C	ITT	10	-20	389	0	10	-20	387	0
	AT	30	0	1	94	32	2	5	60
	PP	29	-1	3	76	30	0	2	96
	IV	25	-5	33	35	25	-5	31	38
	CACE	26	-4	15	34	30	0	3	94
E	ITT	15	-15	239	0	15	-15	241	0
	AT	26	-4	16	9	28	-2	5	63
	PP	28	-2	6	61	30	0	2	94
	IV	35	5	32	41	35	5	31	39
	CACE	34	4	16	32	36	6	4	8

AT, as treated; CACE, complier average casual effect; ITT, intention to treat; IV, instrumental variable; MSE, mean square error; PP, per protocol.

of 60 scenarios by varying non-compliant factors and the magnitude of treatment effect (ie, null, moderate or large). The findings will help investigators choose the optimal approaches when dealing with similar non-compliant problems. Our results also confirm some

previous findings. For example, the ITT analysis was unbiased if the treatment effect was zero.⁹ All estimates were unbiased if non-compliance was independent of patients' outcomes and the IV estimate was also unbiased when non-compliance was symmetrically dependent on

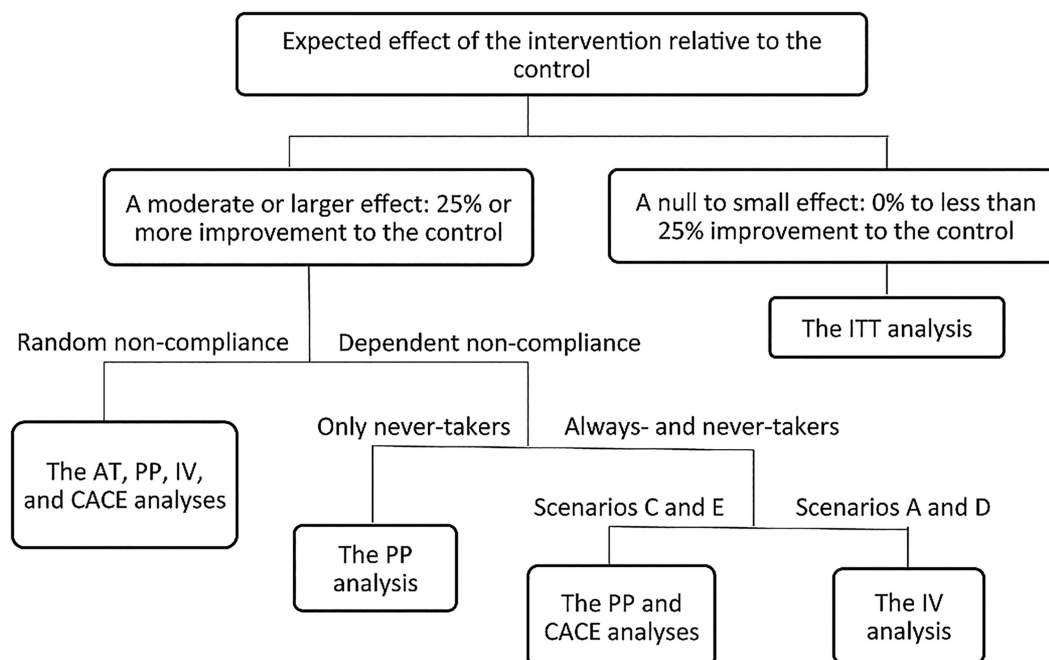


Figure 3 Choosing optimal analyses for different non-compliant scenarios. ITT, intention to treat; AT, as treated; PP, per protocol; IV, instrumental variable; CACE, complier average causal effect. Scenario A: patients with good conditions will always get the intervention and those with poor conditions will always reject it. Scenario C: patients with poor conditions will always reject the intervention. Scenario D: patients with good conditions will always reject the intervention and those with poor conditions will always get it. Scenario E: patients with good conditions will always reject the intervention. A good condition was defined to have an outcome score at least 0.5 SDs above the group mean under usual care. A poor condition was defined to have an outcome score at last 0.5 SDs below the group mean under usual care. In addition, it was assumed that the intervention and usual care were the only treatment option.

patients' outcomes.¹⁷ In addition, we found that the PP estimate was unbiased when there were only never-takers. While the real impact of non-compliance on estimating treatment effect is difficult to generalise, we have compared the performance of common analyses under specific non-compliant scenarios. The results highlight the value of employing multiple approaches to analyse non-compliant data. Our work has considered additional non-compliant scenarios that were not considered by previous studies. It also contributes to the quality assessment of research evidence generated from RCTs subject to non-compliance and provides basis for a more complex evaluation.

CONCLUSION

Our simulation shows that the ITT analysis under non-compliance is considerably biased when an intervention has a large effect over the control. Alternative analyses can provide unbiased or less biased estimates. For RCTs subject to non-compliance, we make some suggestions for the choice of analyses under specific scenarios to minimise the bias of estimated treatment effect. Our study also informs the design of further investigations on the issue of non-compliance in RCTs.

Contributors CY conceived the study, designed and performed the simulations, conducted the statistical analyses, interpreted the results and drafted the manuscript; LT advised on the design of the study and revised the manuscript; JB and GB contributed to the interpretation of the results and revision of the manuscript; all authors have read and approved the final manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. CY is supported in part by funding from the Father Sean O'Sullivan Research Center (FSORC) Studentship award, the Canadian Institute of Health Research (CIHR) Training Award in Bridging Scientific Domains for Drug Safety and Effectiveness, and the Canadian Network and Centre for Trials Internationally (CANNeCTIN) programme.

Competing interests The funding organisations have no influence on the submitted work. All authors declare no competing interests.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- McMahon AD. Study control, violators, inclusion criteria and defining explanatory and pragmatic trials. *Stat Med* 2002;21:1365–76.
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Clin Epidemiol* 2009;62:499–505.
- Tunis SR, Stryer DB, Clancy CM. Practical clinical trials. *JAMA* 2003;290:1624–32.
- Ye C, Browne G, Beyene J, et al. A sensitivity analysis of the Children's Treatment Network trial: a randomized controlled trial of integrated services versus usual care for children with special health care needs. *Clin Epidemiol* 2013;5:373–85.
- Hubbard W. ICH: guidance on statistical principles for clinical trials (FDA Docket No. 97D-0174). *Fed Regist* 1998;63:49583–98.
- Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995;57:6–15.
- Nich C, Carroll KM. 'Intention-to-treat' meets 'missing data': implications of alternate strategies for analyzing clinical trials data. *Drug Alcohol Depend* 2002;68:121–30.
- Wright C, Sim J. Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. *J Clin Epidemiol* 2003;56:833–42.
- Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9:48–55.
- Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. *Trials* 2012;13:84.
- Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 1991;47:1213–34.
- Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Stat Methods Med Res* 2005;14:369–95.
- Martens EP, Pestman WR, de Boer A, et al. Instrumental variables: application and limitations. *Epidemiology* 2006;17:260–7.
- Jo B. Estimation of intervention effects with noncompliance: alternative model specifications. *J Educ Behav Stat* 2002;27:385–409.
- Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 2009;65:640–9.
- Robin JM. Correction for non-compliance in equivalence trials. *Stat Med* 1998;17:269–302.
- Bang H, Davis CE. On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Stat Med* 2007;26:954–64.
- McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Stat Med* 2009;28:2639–52.
- Sheng D, Kim MY. The effects of non-compliance on intent-to-treat analysis of equivalence trials. *Stat Med* 2006;25:1183–99.
- Hertogh EM, Schuit AJ, Peeters PH, et al. Noncompliance in lifestyle intervention studies: the instrumental variable method provides insight into the bias. *J Clin Epidemiol* 2010;63:900–6.
- Burton A, Altman DG, Royston P, et al. The design of simulation studies in medical statistics. *Stat Med* 2006;25:4279–92.
- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996;91:444–55.
- Atun R, de Jongh T, Secci F, et al. A systematic review of the evidence on integration of targeted health interventions into health systems. *Health Policy Plann* 2010;25:1–14.
- Düsing R, Lottemoser K, Mengden T. Compliance with drug therapy—new answers to an old question. *Nephrol Dial Transplant* 2001;16:1317–21.
- Yau LHY, Little RJ. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *J Am Stat Assoc* 2001;96:1232–44.
- Ye C, Giangregorio L, Holbrook A, et al. Data withdrawal in randomized controlled trials: defining the problem and proposing solutions: a commentary. *Contemp Clin Trials* 2011;13:318–22.