Journal of Clinical Epidemiology 108 (2019) 121−131

Journal of Clinical Epidemiology

REVIEW

An overview of statistical methods for handling nonadherence to intervention protocol in randomized control trials: a methodological review

Mohammod Mostazir[a,*], Rod S. Taylor[b], William Henley[b], Ed Watkins[c]

[a]College of life and Environmental Sciences (CLES), University of Exeter, EX4 4QG, Exeter, United Kingdom
[b]Institute of Health Research, University of Exeter Medical School, University of Exeter, EX1 2LU, Exeter, United Kingdom
[c]Mood Disorders Research, School of Psychology, College of Life & Environmental Sciences, University of Exeter, EX4 4QG, Exeter, United Kingdom

Accepted 4 December 2018; Published online 7 December 2018

Abstract

Objective: To undertake a methodological review of statistical methods used in randomized controlled trials (RCTs) for handling intervention nonadherence.

Study Design and Setting: Bibliographic databases were searched using predefined search terms.

Results: A substantive number of identified studies (56%) were excluded as they only used naive per protocol analysis for handling nonadherence. Our review included 58 articles published between 1991 and 2015. A total of 88 methodological applications were made by these studies. The two most used methods were complier average causal effect (56%) and instrumental variable (23%) predominantly with the use of maximum likelihood (ML) estimators. These alternative applications typically produced treatment effects greater than the intention-to-treat effect but as their standard errors were larger there was no statistical difference between the methods.

Conclusion: A substantive proportion of RCTs rely on naive per protocol for handling nonadherence. Recent years have seen an increasing number of applications of more appropriate statistical methods, in particular complier average causal effect and instrumental variable methods. However, these later methods rely on strong underlying assumptions that may be vulnerable to violation. More empirical studies are needed that directly compare the usability and performance of different statistical methods for nonadherence in RCTs.

Keywords: Nonadherence; Noncompliance; Randomized controlled trial; Methodological review; Causal effect modeling; Statistical methods

1. Introduction

Randomized controlled trials (RCTs) and systematic reviews of RCTs provide the highest level of evidence for assessing the effects of health care interventions [1]. Researchers, however, still face challenges when undertaking RCTs. One of these is the nonadherence/noncompliance of trial participants to the intervention(s) protocol to which they are randomized.

Nonadherence has been shown to be associated with poorer patient outcomes, including higher mortality [2]. A meta-analysis across 569 trials estimated an average treatment nonadherence rate of 25% [3], whereas another study reported a rate of 23% [4]. Current reporting guidelines for RCT (consolidated standards of reporting trials [CONSORT]) recommend the intention-to-treat (ITT) approach, that is, outcomes are compared according to original group allocation regardless of whether participants received the intervention according to the protocol or not [5,6]. By doing so, ITT evaluates effectiveness of an intervention by mirroring the nonadherence to treatment that may occur in real-world practice. Although this may be true, it is argued that by ignoring nonadherence, ITT underestimates the ''true (or causal) effect'' of the intervention because the analysis is diluted by noncompliers [7−9].

A commonly used approach by analysts to handling nonadherence is per protocol (PP) analysis where the outcomes of intervention are compared according to initial random allocation but excluding those participants who do not adhere to the intervention protocol [10,11]. A systematic review of 100 RCTs identified 47% studies to have adopted some form of PP analysis [12]. The PP approach is prone to serious selection bias as it fails to preserve the original randomization, and causality of treatment effect cannot be claimed [13]. ''As treated'' (AT) analysis is another variant of non-ITT analysis [14,15] that classifies participants

Conflict of interest statement: None declared.

* Corresponding author. Tel.: +44 1392 726629; fax: +44 (0) 1392 724003.
E-mail address: m.mostazir@exeter.ac.uk (M. Mostazir).

https://doi.org/10.1016/j.jclinepi.2018.12.002
0895-4356/

What is new?

Key findings

- A large proportion of RCTs continue relying on naïve per protocol (PP) method for handling intervention nonadherence.
- Several statistical applications based on causal framework are now available to more appropriately adjust treatment effect for nonadherence in RCTs data.
- Complier Average Causal Effect (CACE) and Instrumental Variable (IV) are the two widely used method for handling nonadherence.

What this adds to what is known?

- Based on causal framework, these methods can be applied to various study settings.
- Maximum likelihood-based CACE and IV are more appropriate approaches to handling nonadherence in RCTs.

What is the implication and what should change now?

- Unlike PP method, identified methods are important unbiased alternatives to ITT when adherence to treatment is suboptimal. However, given they also suffer from strong underlying assumptions, these methods should always be reported in addition to ITT analysis and regarded as a sensitivity analysis.

according to the intervention they receive regardless to their adherence to the trial protocol and like PP analysis is subject to selection bias [16,17].

Several statistical methods have been developed for estimating causal treatment effects that take account of intervention nonadherence without introducing the biases inherent to PP or AT analyses. The statistical framework for causal inference in RCTs was developed by Rubin, referred to as Rubin's causal model where each participant is assumed to have a set of counterfactual outcomes [18−21]. Under the Rubin's causal model framework, several methods developed for handling nonadherence, including instrumental variable (IV) approach from the field of econometrics [22], complier average causal effect (CACE) by Rubin [14,23], and structural mean models (SMMs) by Robins [24]. To our knowledge there has been no comprehensive review of the use of these statistical methods and their pros and cons.

We undertook a methodological review of RCTs that described statistical methods for handling nonadherence to intervention protocol. Given the bias associated with the methodology, we excluded studies that utilized PP analysis alone. The aims were to: (1) assess the range of statistical methods reviewed and applied in RCTs to handle nonadherence; (2) review the relative pros and cons of these statistical methods; and (3) make a pooled comparison of the treatment effects estimated by ITT and proposed statistical methods for handling nonadherence.

2. Methods

We conducted and reported this methodological review in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement [25].

2.1. Literature search strategy

We searched a number of bibliographic databases, that is, EMBASE (OvidSP), PsycInfo (OvidSP), MEDLINE (OvidSP), CINAHL (EBSCOHOST), and Cochrane Library for methodological studies (Wiley Online Cochrane Library) from inception to June 2015. Database-specific Boolean search strategies were developed using key terms, that is, ''intention to treat'', ''as-treated'', ''per protocol'', ''non-adherence'', ''complier average causal effect'', ''CACE'' (and synonyms). The reference lists of the included articles were manually checked. Details of the search strategy are provided in the e-appendix (A).

2.2. Study selection

We included RCTs that reviewed statistical methods for handling nonadherence and applied these methods to actual/simulated trial participant data. Studies were excluded if: (1) they were available only as abstracts/titles and not as a full publication; (2) they adjusted for nonadherence but provided no information on the statistical basis of this method (this included studies that simply stated that they used ''IV'' or ''CACE'' analysis but gave no further methodological details) [26]; (3) they applied statistical methods for handling any potential confounding/bias but this was unrelated to nonadherence to intervention protocol [27].

2.3. Data extraction

A database was compiled that captured information on characteristics of included studies, that is, title, authors, journal, year of publication, population disease area, type of intervention, randomizing unit, study duration, type of outcomes, sample size, and estimated treatment effect by ITT method and by the proposed methods. Detailed information was extracted on the method of statistical analysis applied, that is, name of the statistical method/framework, statistical estimators/algorithm applied to implement the technique, and any advantages/disadvantages of these statistical method as stated by authors.

2.4. Data analysis and presentation

A descriptive approach was taken to data presentation using frequencies, means, and medians. Pooled comparison

of direct treatment effects across studies was not feasible as studies had varied outcomes. For comparison of treatment effect between ITT and the proposed methods, we compared whether the treatment effect by ITT was larger or smaller compared to the effect estimated by the proposed method (coded ''yes/no'') and presented the results in frequency (%). Furthermore, absolute z-statistic ([treatment effect/standard error [s.e.]) was calculated for each method application and the pooled mean z-statistic was compared between ITT and proposed methods. This pooled comparison accounted for within-study variance by subtracting each proposed method z-statistic from ITT z-statistic before calculating pooled mean z-statistic, that is,

$$\frac{\sum_{i=1}^{n}\left(z_{pi}-z_{ITTi}\right)}{n},$$

where Z is the z-statistic for proposed (p) or ITT method for individual study (i) and n is the total number of method applications. The pooled z-statistic was also used to compare treatment effect between IV vs. CACE method by metaregression accounting for nonadherence rate. Applications made on simulated data and applications involving Bayesian method were excluded from these comparisons. Authors presented information in various formats, that is, presenting coefficient and 95% confidence interval (CI), presenting coefficient and s.e or presenting coefficient and the P-value only. We derived required statistic applying appropriate formulae [28−30], where applicable. Where results from several models were presented, that is, model comparisons from sensitivity analyses, for data extraction, we considered the optimum model suggested by the author. Analyses were undertaken using statistical software Stata, version 15 [31].

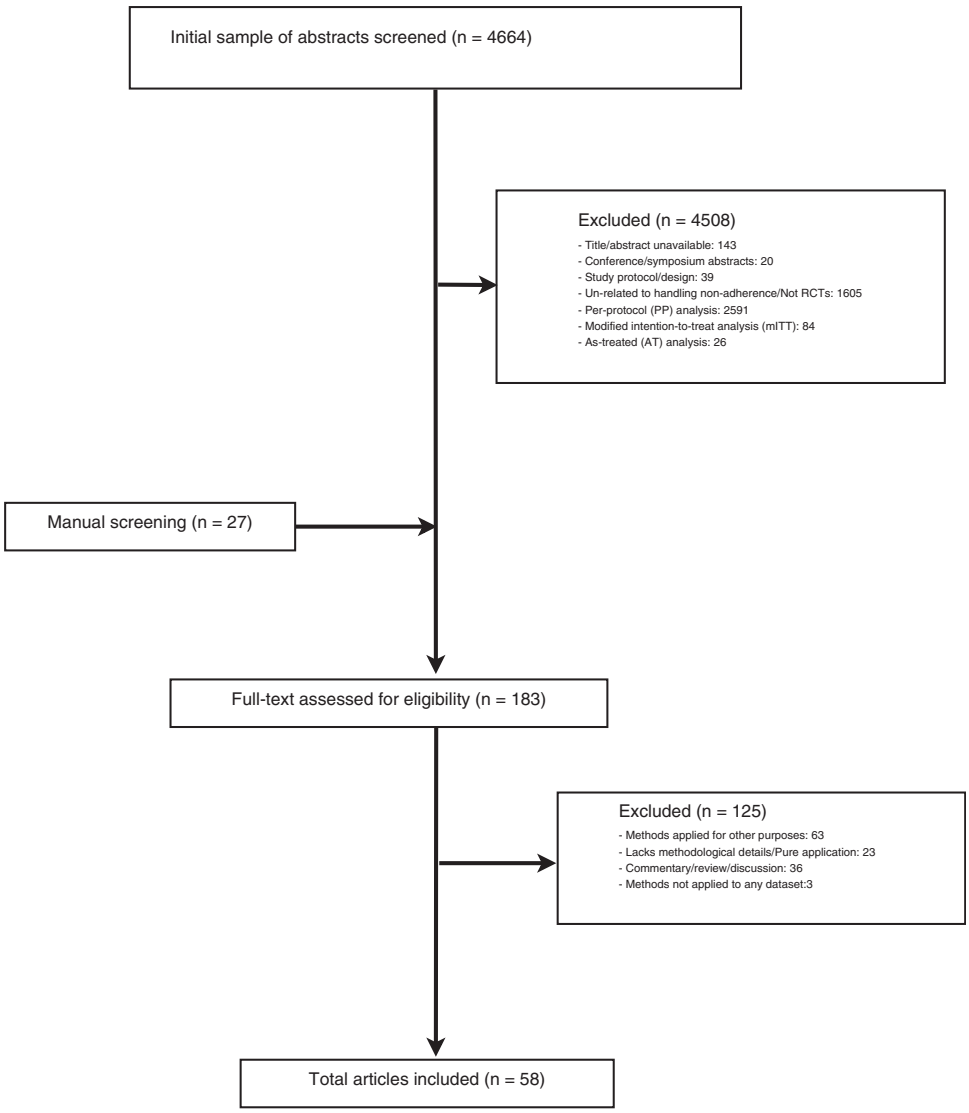

Fig. 1. Flow of studies through inclusion and exclusion process.

Table 1. Summary of included study characteristics

Characteristics	Number	Percent (%)
Number of articles	58	100
Year of publication (n = 58)	–	–
1991-1999	12	21
2000-2007	21	36
2008-2015	25	43
Journals (n = 58)	–	–
Statistics in medicine	20	34
Biometrics	8	14
Biostatistics	5	9
Journal of the American Statistical Association	4	7
Controlled Clinical Trials	2	3
Journal of the Royal Statistical Society	2	3
Psychological Methods	2	3
American journal of epidemiology	1	2
Biometrical Journal	1	2
Biometrika	1	2
British Journal of Psychiatry	1	2
Clinical trials	1	2
Family process	1	2
Health Services & Outcomes Research Methodology	1	2
Journal of Biopharmaceutical Statistics	1	2
Journal of Clinical Epidemiology	1	2
Journal of Educational and Behavioral Statistics	1	2
Psychological Medicine	1	2
Statistica Sinica	1	2
Statistical Methods in Medical Research	1	2
The American journal of drug and alcohol abuse	1	2
The annals of statistics	1	2
Study population clinical area (n = 58)	–	–
Mental health	17	29
Cardiology	7	12
Infectious disease	5	9
Oncology	4	7
Others	20	34
Simulation study/NA	5	9
Follow-up duration (n = 58)	–	–
1 yr	11	19
1-2 yr	19	33
>2 yr	15	26
Not reported	13	22
Type of intervention (n = 58)	–	–
Drug	21	36
Psychotherapy	8	14
Behavioral	8	14
Other	16	28
Simulation study (N/A)	5	9
Type of outcome (n = 58)	–	–
Binary	29	50
Continuous	19	33
Count/Time-to-event	10	17
Sample size (n = 58)	–	–
<400	19	33
400-1,000	14	24
>1,000	18	31
Not reported/NA	7	12
Randomizing unit (n = 58)	–	–
Individual	46	79
Cluster	5	9
Meta-analysis	1	2
Simulation study (N/A)	6	10
% Nonadherence: Median (range)	–	34% (2% to 78%)

3. Results

3.1. Selection of included RCTs

The literature searches resulted in a total of 4,664 titles/abstracts, of which 58 were eligible for inclusion (Fig. 1). A total of 2,591 (56%) of the abstracts were excluded because authors exclusively relied on PP analysis to deal nonadherence. A small number of studies were also excluded for applying AT analysis (26 studies, 0.56%) and modified ITT (84 studies, 1.8%) as both are forms of PP analysis [12]. The other reasons for exclusions were applications that were non-RCTs, unrelated to handling nonadherence or lacked methodological details.

3.2. Characteristics of included RCTs

Detailed description of included studies is given in the e-Appendix (B). Summary of study characteristics is presented in Table 1. Most included studies were published in statistical/methodological journals. Studies were undertaken across a wide range of patient and intervention types, study sizes, and duration and were applied across a range of outcome types (continuous/binary/count/time-to-event).

A significant rise was observed in the published literature in this area of methodological research since 1999. Per year average rate of publication (incidence rate ratio [IRR]) was higher at the later periods compared to the 1991−1999 period (IRR for 2000−2007: 1.56, 95% CI: 1.00 to 2.45, $P < 0.05$; IRR for 2008−2015: 1.72, 95% CI: 1.11 to 2.65, $P < 0.01$).

3.3. Statistical methods and estimators used in included RCTs to handle nonadherence

A total of nine methods for handling treatment nonadherence were described across the included studies (Table 2a). Some of these studies applied more than one method using different estimators resulting in a total of 88 statistical method applications. Studies that were judged to be variants of a common statistical approach were grouped under broader approach, that is, ''Longitudinal Complier Average Causal Effect'' or ''Complier Average Causal Effect within Effect Class (ECACE)'' and were grouped under ''Complier Average Causal Method (CACE)'' as they broadly have similar statistical framework. The two most common statistical methods applied were the CACE (49/88 applications, 56%) and IV (20/88 applications, 23%). Authors applied the term ''IV'' for both ''method'' and ''estimator'' where ''IV-method'' refers to the causal instrumental variable framework [23] and ''IV-estimator'' refers to implementation of a particular method applying two-stage least square (2SLS) estimator [32]. A total of 10 estimators (Table 2b) were identified; the most common being maximum likelihood (ML) estimator in 33% ($n = 29$), method of moment (MOM) base estimator in 22% ($n = 19$), IV estimator in 19% ($n = 17$), and Bayesian estimators were used in 10% ($n = 9$) of the applications.

Fig. 2a shows the different estimators used for applications of CACE methods. ML base estimators were implemented with expectation maximization algorithm [33], and Bayesian inference (BI) base methods were implemented both with expectation maximization and Markov-Chain Monte-Carlo algorithm [34]. As shown in Fig. 2a, 49% ($n = 24$) of the CACE applications were made using ML estimators. Fig. 2b shows all other statistical methods and the frequencies of the use of their different estimators.

3.4. Pros and cons of statistical methods presented by authors

The remainder of the methods section and Table 3 provide an overview of the statistical basis of the statistical methods and stated pros and cons of these approaches.

3.4.1. Complier average causal effect (CACE)

Based on counterfactual outcome [21], the CACE method was introduced by Angrist et al. for estimating causal effects in the presence of nonadherence [23]. In CACE analysis, the potential adherence classes are stratified into four principal strata based on principal stratification [35], that is, i) ''Compliers'', that is, receive treatment when they are assigned to it, ii) ''Never-takers'', that is, do not receive treatment when they are assigned to it, iii) ''Always-Takers'', that is, always receive the treatment regardless of randomization, and iv) ''Defiers'', that is, always do the opposite of what is assigned and assumed to be nonexistent. In addition to randomization and the stable unit treatment value assumption (SUTVA) [19,36], there are two key assumptions that need to be fulfilled for CACE model to be identified: (1) the effect of treatment assignment on outcomes entirely operates through treatment receipt status of participants, known as ''exclusion restriction'' (ER). ER in other words states that under true randomization, the proportion of noncompliers in the control group (had they been offered the treatment) and their outcomes are similar to the proportion of observed noncompliers and their outcomes in the treatment group; (2) The ''monotonicity'' assumption implies that there are no ''defiers'' meaning no participants will refuse treatment when assigned to treatment and will seek treatment when assigned to control. Although the initial CACE estimator proposed by Angrist et al. was an IV estimator [23], we

Table 2. Statistical methods (a) and their estimators (b) as stated by the authors, applied for handling nonadherence

I #	a) Methods	Method elaboration	Number (%)	b) Estimators	Estimator elaboration	Number (%)
1	CACE	Complier average causal effect model	49 (56)	ML	Maximum likelihood	29 (33)
2	IV	Instrumental variable model	20 (23)	MOM	Method of moments	19 (22)
3	SNMM	Structural nested mean model	7 (8)	IV	Instrumental variable estimator	17 (19)
4	ATR	Adjusted treatment received model	4 (5)	BI	Bayesian inference	9 (10)
5	RPSFTM	Rank preserving structural failure time model	3 (3)	G-estimator	G-estimator	5 (6)
6	C-PROPHET	Rank preserving structural failure time model	2 (2)	Cox-PH	Cox-proportional hazard estimator	4 (5)
7	CALM	Compliers proportional hazards effect of treatment with proportional Hazards model	1 (1)	WLS	Weighted least square	2 (2)
8	Cox-Reg1	Causal accelerated life model	1 (1)	GSMM	Generalized structural mean model estimator	1 (1)
9	Cox-Reg2	Regression adjustment with Cox-model	1 (1)	ISM	Intensity score method	1 (1)
				WGSNM	Weighted generalized structural mean model	1 (1)
	Total		88 (100)			88 (100)

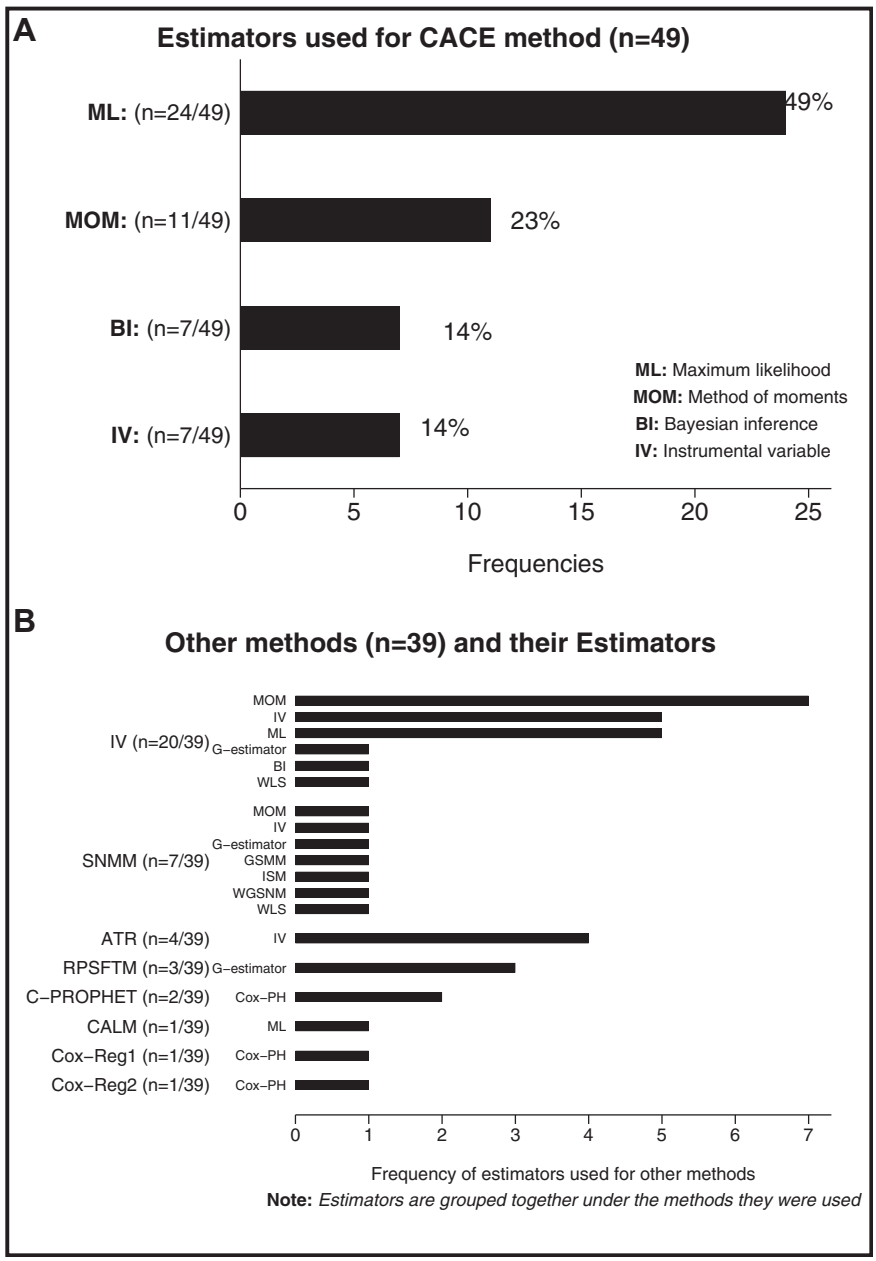

Fig. 2. (A) Use of varied estimators for estimating CACE and (B) other methods.

identified several other estimators for CACE applied into different settings (Fig. 2a). Our findings, across several types of CACE applications, suggest that ML-base estimation was applied more often than other estimators and the reason may be that ML estimates are considered more efficient than 2SLS-based IV estimators [37−39]. We also found applications of IV estimators in combination with ML estimators in estimating CACE, and this combination contributed to substantial methodological development [40]. Missing data add another level of complexity in presence of treatment nonadherence, and in over half of CACE applications (24/47), authors provided guidelines for handling missing data. CACE also has been implemented in cluster randomized trials where intraclass correlation from similar adherence behavior at cluster level may compromise estimated treatment effects [41−44]. When there are multiple arms involved, CACE model may suffer from nonidentifiability issues or may require complex modeling assumptions [45,46], and Bayesian methods may be applied addressing such complexities [47]. The fundamental limitation of the CACE approach is that the underlying assumptions, that is, ER, monotonicity are not easily testable [48−51], and if violated, CACE estimates may be biased [40,51,52].

Table 3. Summary of methods, pros, and cons as stated by the authors

Methods[a]	Frequency	Estimators[a]	Method description	Strengths	Limitations in implementation
CACE	49	BI/IV/ML/MOM	Based on principal stratification and counterfactual outcome rather observed outcome, CACE estimate relates to treatment effect for participants who would have complied with the treatment had they been offered it.	i) In presence of nonadherence, unlike PP/AT, it yields causal effect of treatment on the treated. ii) Randomization-based estimate of efficacy iii) By using pretreatment covariates identifiable CACE model is feasible instead forcing exclusion restriction assumption. iv) Cell specific, that is, ''always-takers'', ''never-takers'' treatment effect estimation is possible.	i) Assumptions are not directly testable and may not reflect the real world scenario under nonadherence ii) CACE estimates can be biased if assumption is violated and the bias can be substantial with low compliance rate iii) The strong assumptions limit the flexibility of CACE modeling in practice iv) Encounters difficulty when number of randomized arms and adherence categories increase.
IV	20	BI/G-estimator/IV/ML/MOM	Instrumental variable method historically grounded in econometric theory where an instrument is an exogenous variable that influences the outcome through adherence-related posttreatment variable only.	i) IV estimate does not require an assumption of homogeneous treatment effects, under exclusion restriction and the monotonicity assumptions ii) Method is not sensitive to differences in baseline risk between compliers and noncompliers iii) IV tends to have superior RMSE unless the compliance rate is low or zero.	i) Outcome may be affected by other means rather solely through treatment received and randomization. ii) Method is most sensitive to violations of exclusion restriction and to the monotonicity assumption when there are few compliers. iii) With 2SLS estimator and low compliance rate, IV can produce large variances iv) Implementation with 2SLS can only be feasible if missing data are ignorable.
SNMM	7	GE/WGSNM/ISM/IV/MOM/ WGSNM/WLS	Method relates to comparing mean of the outcome at an observed compliance level with the mean of the potential outcome at some reference level.	i) Provide randomization-based causal effects at varying level of adherence ii) Cause–effect relationships are established by considering a potential treatment-free outcome for each observational unit.	i) With binary outcomes, in some instances with complex study design, that is, three arms, the estimating equation has no solution. ii) G-estimation can be complex in implementation.
ATR	4	IV	Variation of IV method with error term from endogenous regression added in the model as covariate.	i) Method copes with missing data problems as the first stage uses all randomized participants. Only the second stage is affected with missing data.	i) Method only valid when patients switch between treatment arms, for example, when one arm consists of placebo therapy and a placebo effect is not anticipated.
RPSFTM	3	GE	Estimates parameters of a class of semiparametric failure time models, using a class of rank estimators. These models are the structural version of the ''accelerated failure time model with time-dependent covariates.''	i) Yields valid results for both outcome-dependent and outcome-independent treatment noncompliance ii) designed to consistently estimate causal effects on the treated, without direct assumptions about the compliance selection mechanism.	i) RPSFTM makes a strong noninteraction assumption which in certain settings might be considered biologically implausible ii) G-estimation can be complex in implementation.
C-PROPHET	2	Cox-PH	Randomization-based proportional hazard model for continuous survival data.	i) Model framework is similar to CACE and provides estimates equivalent to CACE.	i) Allows only binary/all or none compliance ii) Under time-varying noncompliance the model maybe biased.

(*Continued*)

Table 3. Continued

Methods[a]	Frequency	Estimators[a]	Method description	Strengths	Limitations in implementation
Cox-Reg$_{1\text{-}2}$	2	Cox-PH	Cox-proportional hazard model for continuous survival data adjusting for nonadherence.	i) Yield valid causal estimates under random nonadherence.	i) Under nonrandom adherence, estimates can be biased.
CALM	1	ML	Accelerated failure to time (AFT) model for survival data that relates each observed event time in the treated group to a potential event time that would have been observed if the control treatment had been given throughout the trial.	i) Allows cross-over type nonadherence between arms ii) Estimates hazard ratio equivalent to CACE ii) Does not rely on homogenous population.	i) The model may produce extreme values and larger error terms.

Abbreviation: RMSE, root mean square error.
[a] Please refer to Table 2 for the elaboration of the method/estimator acronyms.

3.4.2. Instrumental variable (IV)

An IV is an exogenous variable that influences the outcome solely through a binary posttreatment variable that identifies whether participants adhered to treatment or not [21,22]. Typically in RCTs, an IV is the randomizing variable, and participants' adherence status is the endogenous variable through which outcome is affected. The assumption that outcome solely depends on adherence status is equivalent to the ER assumption discussed in the CACE section above. Therefore, in a two-arm trial design where participants' choice to postrandomization switching between arms is restricted, an IV estimates alternate CACE estimates, given same estimator applied [48]. Typically IV estimators are implemented with 2SLS [32] estimators; however, ML-based IV estimators are also used [53]. In our selected studies, we separately identified ML-base IV estimators where authors were explicit about it. We found IV methods, like CACE, being applied in varied scenarios. However, IV with 2SLS is likely to estimate treatment effect on complete case basis and valid only when missing data are ignorable [32]. When compliance rate is low, 2SLS-base IV estimator produces large effects compared to ITT and produces large variances, which makes it a less attractive estimator [54]. In such scenarios, ML is a more efficient estimator of IV [55]. A variation of IV method is adjusted treatment received method introduced by Nagelkerke [56] with an adjustment made to error terms. The distinction to typical IV method is that in adjusted treatment received, the error terms from first stage endogenous regression is added to the model as a covariate to allow adjustment for any unmeasured confounding.

3.4.3. Other statistical methods

SMM/structural-nested mean model was introduced by Robins [24]. The framework provides causal treatment effect for observed adherence comparing with a conditional reference level of adherence [46]. Linear additive framework is used for continuous outcomes and multiplicative framework is used for binary outcomes. Models are estimated with the G-estimator (GE) proposed by Robins and Tsiatis [57,58]. The appealing aspect of SMM is that causal parameters can be estimated for varying levels of adherence. However, identifying reference level of compliance may be challenging [59]. Another version of SMM applied to accelerated failure time model (time to event survival data with time as outcome) is rank preservative structural failure time model (RPSFTM) [60,61]. They are called rank preserving because they use a class of rank estimators for subjects' failure [61]. In practice, G-estimators have not been widely adopted because of level of complexities involved in implementation [62]. For handling nonadherence in continuous time survival data include Cox-reg$_{(1,2)}$, complier proportional hazard effect of treatment (C-PROPHET) model, and causal accelerated life model (CALM) [63,64]. Cox-reg$_{(1,2)}$ both are adherence adjustment base Cox-regression implemented by Cox-PH estimator. Cox-reg$_1$ is implemented in situations when compliance is ''all-or-nothing'', that is, either patients receive the treatment or they do not, and Cox-reg$_2$ is implemented when compliance is partial. CALM and C-PROPHET both have CACE-like framework where CALM is applied to accelerated failure time model and C-PROPHET is applied to continuous time survival data.

3.5. Comparison of estimated treatment effects

We were able to compare treatment effect for 68/88 applications. Most of the alternative methods (n = 48, 71%) produced treatment effects that were greater than the treatment effect estimated by ITT. For 11 applications (16%), estimates were similar for both ITT and the proposed methods. For all alternative methods, excluding the Bayesian applications, 95% CIs overlapped with the CIs of ITT either at lower or upper bound region.

Sixty-four of 88 applications contributed to the calculation of s.e. and z-statistics. In 83% of the applications (53/64), s.e. for the alternative methods were larger than the s.e. of ITT estimates. After accounting for within study variation, average z-statistic from proposed methods were greater by +0.13 SD (95% CI: −0.99 to 1.71). We found 7 of 58 studies (12%) achieved significant treatment effect

by applying an alternative method, which was not achieved by the ITT method.

In metaregression, when accounted for percent nonadherence rate, z-statistic for IV method was no different than z-statistic from ITT (−0.01, 95% CI: −0.27 ti 0.26) but z-statistic from CACE was greater by +0.18 SD (0.18, 95% CI: −0.01 to 0.35). CACE estimates were higher by the same amount when compared to IV.

4. Discussion

In this review, across 58 studies, a wide variety of statistical methods against ITT were identified for handling treatment nonadherence. The median intervention nonadherence was 38% ranging across studies from 2% to 78%. The two most commonly used methods were CACE and IV. Overall, there was no significant difference between the pooled z-statistics from ITT and the alternative methods. In general, the most of the proposed applications (83%) produced larger error variance compared to the error variance produced by ITT. We note that use of the CACE method resulted in larger z-statistics compared to the IV method when accounting for nonadherence rate.

We are aware of two previous systematic reviews undertaken to assess the analytical approaches to the handling of treatment protocol nonadherence in RCTs. Dodd et al. [12] summarized the extent to which nonadherence to treatment protocol is reported in RCTs. However, this study did not identify methods apart from the conventional ITT, PP, and AT analysis approaches. Adewuyi et al. [65] studied nonadherence in surgical intervention and reported that 63% of the studies adopted ITT, 21% adopted PP, and 3% adopted AT analysis. Our systematic review is therefore the first to identify and systematically reviewed statistical methods that have been developed to handle nonadherence using a causal inference framework.

The CACE and IV methods are flexible and have been applied across a range of RCT designs. One of the benefits of the CACE application is that cell-specific treatment effect can be obtained, which can provide valuable insights for researchers in relation to various types of adherence, whereas this opportunity is limited for the IV approach. We also found good number of applications of MOM estimators (19/88, 22% applications), but we avoided emphasizing on it as it relies on simple cell means and ignores distributional error terms that can be erroneous [51,66]. According to our findings, both CACE and IV methods are applicable to varieties of scenarios and both rely on strong assumptions that are vulnerable to violations. Unless there are direct ways of testing the assumptions, it is not readily verifiable whether the applied methods captured the true treatment effect or simply inflated the treatment effect influenced by level of adherence.

5. Strengths and limitations

The strength of this study was its use of a systematic review approach to identify studies for inclusion. However, it has a number of limitations. Study selection and data extraction were undertaken by a single reviewer (MM), although the opinion of a second reviewer (RST) was available. We were unable to compare treatment effects estimated by different statistical methods because of their varied outcomes. The comparison of pooled z-statistic may not be an ideal approach, but this provides an indication of location of treatment effect estimated by different methods around the region of significance. We excluded studies that implemented relevant methods purely for application purposes instead of providing methodological guidance for handling suboptimal adherence and also studies that used statistical methods for handling general confounding other than handling nonadherence exclusively, for example, propensity score (PS), inverse probability weighting (IPW). These methods have wider applications in observational studies for adjusting general confounding based on probabilistic weighting, but they directly do not contribute to the formation of causal frameworks for handling nonadherence.

6. Implications for practice and policy

Usually the ITT estimate of a treatment effect will be smaller than the ‘‘true’’ effect because if the treatment works, noncompliance to treatment means suboptimal effects. Therefore, the search for an alternative to ITT is a growing area of interest. Perhaps surprisingly, we found that a large number of RCTs continue to use PP methods despite the major limitations of this approach. CACE and IV methods are two important unbiased alternatives to ITT when adherence to treatment is suboptimal, and this review shows that these methods have been applied to a wide range of RCTs. However, given both suffer from strong underlying assumptions, these methods are always reported in addition to ITT analysis and regarded as a sensitivity analysis.

7. Conclusions

Our review found that the alternative methods for handling nonadherence rely on strong assumptions that may be vulnerable to violations. More empirical studies are needed that directly compare the usability and performance of different statistical methods for nonadherence in RCTs.

Acknowledgment

Authors’ contributions: M.M. undertaken literature search, carried out data extraction, statistical analysis, and drafting the article. R.S.T. reviewed article selection/data extraction process and provided guidance on overall structure of the article. W.H. reviewed the statistical aspects of the article, and E.W. provided constructive inputs in article structure, formats, and presentation of results. All authors

were involved in writing the article and had final approval of the submitted and published versions.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2018.12.002.

References

[1] Charlton BG. Medical practice and the double-blind, randomized controlled trial. Br J Gen Pract 1991;41(350):355−6.

[2] Simpson SH, Eurich DT, Majumdar SR, Padwal RS, Tsuyuki RT, Varney J, et al. A meta-analysis of the association between adherence to drug therapy and mortality. BMJ 2006;333:15.

[3] DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004;42:200−9.

[4] Bannister-Tyrrell M, Miladinovic B, Roberts CL, Ford JB. Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. J Clin Epidemiol 2015;68:525−33.

[5] Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999;319:670−4.

[6] Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.

[7] Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. Clin Trials 2012;9:48−55.

[8] Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011;2(3):109−12.

[9] Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. Clin Pharmacol Ther 1995;57:6−15.

[10] Sedgwick P. Per protocol analysis. BMJ 2010;340:c1825.

[11] Sedgwick P. What is per protocol analysis? BMJ 2013;346:f3748.

[12] Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. Trials 2012;13:84.

[13] Sedgwick P. Intention to treat analysis versus per protocol analysis of trial data. BMJ 2015;350:h681.

[14] Frangakis C, Rubin D. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika 1999;86(2):365−79.

[15] Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ 2001;165(10):1339−41.

[16] Ten Have TR, Normand S-LT, Marcus SM, Brown C, Lavori P, Duan N. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. Psychiatr Ann 2008;38(12):772−83.

[17] Bang H, Davis CE. On estimating treatment effects under noncompliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? Stat Med 2007;26:954−64.

[18] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66(5):688−701.

[19] Rubin DB. Bayesian inference for causal effects: the role of randomization. Ann Stat 1978;6:34−58.

[20] Holland PW. Statistics and causal inference. J Am Stat Assoc 1986;81:945−60.

[21] Marcus SM, Gibbons RD. Estimating the efficacy of receiving treatment in randomized clinical trials with noncompliance. Health Serv Outcomes Res Methodol 2001;2(3−4):247−58.

[22] Bloom HS. Accounting for no-shows in experimental evaluation designs. Eval Rev 1984;8(2):225−46.

[23] Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996;91:444−55.

[24] Robins J. Correcting for non-compliance in randomized trials using structural nested mean models. Commun Stat Theory Methods 1994;23(8):2379−412.

[25] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.

[26] Osypuk TL, Tchetgen Tchetgen EJ, Acevedo-Garcia D, Earls FJ, Lincoln A, et al. Differential mental health effects of neighborhood relocation among youth in vulnerable families: results from a randomized trial. Arch Gen Psychiatry 2012;69:1284−94.

[27] Vitaro F, Brendgen M, Giguere C-E, Tremblay RE. Early prevention of life-course personal and property violence: a 19-year follow-up of the Montreal Longitudinal-Experimental Study (MLES). J Exp Criminol 2013;9(4):411−27.

[28] Altman DG, Bland JM. How to obtain the confidence interval from a P value. BMJ 2011;343:d2090.

[29] Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ 2011;343:d2304.

[30] Hackshaw A. Statistical formulae for calculating some 95% confidence intervals. In: A concise guide to clinical trials. Oxford: Wiley-Blackwell; 2009:205−7.

[31] StataCorp. Stata statistical software: release 15. College Station, TX: StataCorp LLC; 2017.

[32] 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(4):369−95.

[33] McLachlan GJ, Krishnan T. The EM algorithm and extensions. New York: John Wiley & Sons, Inc.; 2007.

[34] Gilks WR, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo in Practice. Wiley, London: Chapman & Hall/CRC; 1995:504.

[35] Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics 2002;58:21−9.

[36] Basu D. Randomization analysis of experimental data: the Fisher randomization test. J Am Stat Assoc 1980;75:575−82.

[37] Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. Ann Stat 1997;25:305−27.

[38] Little RJ, Yau L. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. Psychol Methods 1998;3(2):147−59.

[39] Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. Biostatistics 2000;1(1):69−88.

[40] Jo B. Estimation of intervention effects with noncompliance: alternative model specifications. J Educ Behav Stat 2002;27(4):385−409.

[41] Frangakis CE, Rubin DB, Zhou XH. Clustered encouragement designs with individual noncompliance: bayesian inference with randomization, and application to advance directive forms. Biostatistics 2002;3(2):147−64.

[42] Jo B, Asparouhov T, Muthen BO. Intention-to-treat analysis in cluster randomized trials with noncompliance. Stat Med 2008;27:5565−77.

[43] Jo B, Asparouhov T, Muthén BO, Ialongo NS, Brown CH. Cluster randomized trials with treatment noncompliance. Psychol Methods 2008;13(1):1−18.

[44] Lin JY, Have TRT, Elliott MR. Longitudinal nested compliance class model in the presence of time-varying noncompliance. J Am Stat Assoc 2008;103:462−73.

[45] Cheng J, Small DS. Bounds on causal effects in three-arm trials with non-compliance. J R Stat Soc Ser B Stat Methodol 2006;68(5):815−36.

[46] Brumback BA, He Z, Prasad M, Freeman MC, Rheingans R. Using structural-nested models to estimate the effect of cluster-level adherence on individual-level outcomes with a three-armed cluster-randomized trial. Stat Med 2014;33:1490−502.
[47] Long Q, Little RJ, Lin X. Estimating causal effects in trials involving multi-treatment arms subject to non-compliance: a bayesian framework. J R Stat Soc Ser C Appl Stat 2010;59(3):513−31.
[48] Yau L, Little RJ. Inference for the complier-average causal effect for 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.
[49] Huang S, Cordova D, Estrada Y, Brincks AM, Asfour LS, Prado G. An application of the complier average causal effect analysis to examine the effects of a family intervention in reducing illicit drug use among high-risk Hispanic adolescents. Fam Process 2014;53(2):336−47.
[50] Mealli F, Imbens GW, Ferro S, Biggeri A. Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. Biostatistics 2004;5(2):207−22.
[51] Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. Beyond intention to treat: what is the right question? Clin Trials 2014;11:28−37.
[52] Jo B. Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. Stat Med 2002;21:3161−81.
[53] Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. Stat Med 1997;16:1017−29.
[54] Matsui S. Stratified analysis in randomized trials with noncompliance. Biometrics 2005;61:816−823+95.
[55] 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.
[56] Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of noncompliance. Stat Med 2000;19:1849−64.
[57] Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods 1991;20(8):2609−31.
[58] Korhonen PA, Laird NM, Palmgren J. Correcting for noncompliance in randomized trials: an application to the ATBC Study. Stat Med 1999;18:2879−97.
[59] Ma Y, Roy J, Marcus B. Causal models for randomized trials with two active treatments and continuous compliance. Stat Med 2011;30:2349−62.
[60] Cox DR, Oakes D. Analysis of survival data. 1st ed. London: Chapman & Hall/CRC; 1984.
[61] Wu Y, Zhao L, Hou Y, Li K, Zhou X. Correcting for noncompliance in randomized non-inferiority trials with active and placebo control using structural models. Stat Med 2015;34:950−65.
[62] 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.
[63] Odondi L, McNamee R. Performance of statistical methods for analysing survival data in the presence of non-random compliance. Stat Med 2010;29:2994−3003.
[64] Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics 2003;59:100−5.
[65] Adewuyi TE, MacLennan G, Cook JA. Non-compliance with randomised allocation and missing outcome data in randomised controlled trials evaluating surgical interventions: a systematic review. BMC Res Notes 2015;8:403.
[66] O'Malley AJ, Normand SLT. Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. Biometrics 2005;61:325−334+647.