

BMJ Open Statistical methods for non-adherence in non-inferiority trials: useful and used? A systematic review

Matthew Dodd ^{1,2}, Katherine Fielding ³, James R Carpenter ^{1,4},
Jennifer A Thompson ³, Diana Elbourne ^{1,2}

To cite: Dodd M, Fielding K, Carpenter JR, *et al*. Statistical methods for non-adherence in non-inferiority trials: useful and used? A systematic review. *BMJ Open* 2022;**12**:e052656. doi:10.1136/bmjopen-2021-052656

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-052656>).

Received 21 April 2021

Accepted 16 December 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

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

²Clinical Trials Unit, London School of Hygiene & Tropical Medicine, London, UK

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

⁴MRC Clinical Trials Unit, UCL, London, UK

Correspondence to

Matthew Dodd;
matthew.dodd@lshtm.ac.uk

ABSTRACT

Background In non-inferiority trials with non-adherence to interventions (or non-compliance), intention-to-treat and per-protocol analyses are often performed; however, non-random non-adherence generally biases these estimates of efficacy.

Objective To identify statistical methods that adjust for the impact of non-adherence and thus estimate the causal effects of experimental interventions in non-inferiority trials.

Design A systematic review was conducted by searching the Ovid MEDLINE database (31 December 2020) to identify (1) randomised trials with a primary analysis for non-inferiority that applied (or planned to apply) statistical methods to account for the impact of non-adherence to interventions, and (2) methodology papers that described such statistical methods and included a non-inferiority trial application.

Outcomes The statistical methods identified, their impacts on non-inferiority conclusions, and their advantages/disadvantages.

Results A total of 24 papers were included (4 protocols, 13 results papers and 7 methodology papers) reporting relevant methods on 26 occasions. The most common were instrumental variable approaches (n=9), including observed adherence as a covariate within a regression model (n=3), and modelling adherence as a time-varying covariate in a time-to-event analysis (n=3). Other methods included rank preserving structural failure time models and inverse-probability-of-treatment weighting. The methods identified in protocols and results papers were more commonly specified as sensitivity analyses (n=13) than primary analyses (n=3). Twelve results papers included an alternative analysis of the same outcome; conclusions regarding non-inferiority were in agreement on six occasions and could not be compared on six occasions (different measures of effect or results not provided in full).

Conclusions Available statistical methods which attempt to account for the impact of non-adherence to interventions were used infrequently. Therefore, firm inferences about their influence on non-inferiority conclusions could not be drawn. Since intention-to-treat and per-protocol analyses do not guarantee unbiased conclusions regarding non-inferiority, the methods identified should be considered for use in sensitivity analyses.

PROSPERO registration number CRD42020177458.

Strengths and limitations of this study

- This is the first systematic review to identify statistical methods that attempt to account for the impact of non-adherence to interventions in randomised non-inferiority trials.
- A description and critique of the statistical methods identified is provided, along with their target estimands.
- Publications from any year, journal or disease area/patient population were reviewed independently by two authors.
- One author extracted the data from the eligible papers.
- While statistical analysis plans were requested for eligible trials, these could not be obtained for all included trials.

INTRODUCTION

Non-inferiority trials, which assess whether a new intervention is not worse than a proven comparator by more than a clinically acceptable amount, are becoming increasingly common.^{1–3} They are principally used when it is hoped that the new intervention may convey some advantage other than better efficacy (its effect under ideal conditions), such as improved safety, tolerability, convenience or reduced cost.^{4,5}

One of the challenges in these studies, and the focus of this review, is how participants not receiving their randomly assigned intervention according to the trial protocol (termed non-adherence or non-compliance)⁶ should be handled in the statistical analysis.⁶ Examples of non-adherence include not receiving a surgical intervention as planned, not taking all of the prescribed doses of a medication, or not attending all of the sessions of an exercise rehabilitation programme. Such non-adherence is common in trials and has been associated with poorer health outcomes.^{7–9} It can bias estimates of efficacy in either direction and so obtaining an accurate and reliable measure of adherence and accounting for any



non-adherence in the statistical analysis of these studies is essential.^{10 11} Non-adherence may also be linked with missing outcome data if, for example, the trial protocol stipulates that further follow-up is no longer required once adherence drops below a specific threshold or if non-adherent participants become lost to follow-up. The terms adherence and compliance are often used interchangeably, though adherence is preferred here since it is felt to better reflect the partnership between the health-care provider and participant.

A simple approach to handling non-adherence is to define and analyse different analysis sets based on participants' observed levels of adherence, with consistent results providing greater confidence in the trial conclusions.¹ In the setting of non-inferiority trials, the intention-to-treat (ITT) and per-protocol (PP) populations have been advocated and are commonly used.^{4 12 13} However, agreement between the ITT and PP results of these trials does not guarantee that conclusions regarding non-inferiority are free from bias caused by differential, or non-random, non-adherence (where the factors leading to non-adherence are associated with outcomes).¹⁴⁻¹⁶

Standard ITT analyses typically include all participants in their randomised groups irrespective of the intervention actually received.¹⁷ Thus, they reflect the effect of assigning individuals to interventions in clinical practice where not everyone is fully adherent (also known as the 'effectiveness' of an intervention). This approach preserves the balance in known and unknown prognostic factors afforded by randomisation and so any difference in outcomes between study arms can be attributed solely to the experimental intervention.¹⁸ However, in the presence of non-adherence, ITT analyses may yield biased estimates of efficacy (also known as the 'causal effect' of an intervention).¹⁹ In non-inferiority trials, where efficacy and effectiveness may be considered equally important, this can increase the probability of falsely claiming non-inferiority and, therefore, accepting a worse intervention.¹¹

Modified ITT (mITT) analyses are commonly used to address some of the limitations with standard ITT methods.²⁰ This approach allows some randomised participants, such as those who never receive any of the allocated intervention or who are identified as ineligible after randomisation, to be excluded according to prespecified rules.¹⁸ However, across trials, there is substantial variability in how this population is defined and bias may be introduced by subjectively excluding individuals from analysis.^{18 20} In addition, mITT analyses are not typically used to account for the impact of non-adherence.

PP analyses estimate the efficacy of interventions typically by excluding or censoring individuals with major protocol violations, including those who are non-adherent to their allocated intervention.^{1 6 17} Excluding participants in this way can lead to selection bias because non-adherent individuals generally differ from those who are fully adherent with respect to prognostic factors.^{21 22} Furthermore, using a PP analysis to address differential

non-adherence is likely to reduce the protection provided by randomisation, so that trial arms are not fully comparable; this potentially biases the study results in either direction.¹¹ In other words, any difference in outcomes between trial arms may no longer be due to the experimental intervention only. To obtain valid results from a PP analysis, we need to recover the protection due to randomisation, typically through a statistical method that (given certain assumptions) correctly adjusts for factors associated with both adherence and outcome (confounders).²¹

Statistical techniques that attempt to account for the impact of non-adherence and thus estimate the causal effects of experimental interventions exist. These range from simple approaches, such as including observed adherence as a covariate within a regression model, which like PP analyses is susceptible to selection bias, to more sophisticated techniques, such as instrumental variable (IV) methods and inverse-probability weighting, which allow for non-adherence while attempting to maintain the balance produced by randomisation.^{23 24} Several of these methods attempt to estimate the complier average causal effect (CACE), which is the causal effect of an intervention for individuals who would always be fully adherent regardless of assignment (known as compliers).²⁵ In other words, it is a comparison of the average outcome among those who are fully adherent in the experimental arm with the average outcome among the comparable group in the control arm who would fully adhere to the experimental intervention, if offered.

It is unclear which of the alternative methods have been applied in the setting of non-inferiority trials, to what extent, and with what results. Therefore, this systematic review aimed to identify statistical methods that can be used to account for the impact of non-adherence to interventions (thereby estimating the causal effects of experimental interventions) in randomised non-inferiority trials. Secondary aims were to quantify the use of such methods in these studies and examine their impact on non-inferiority conclusions.

METHODS

The Ovid MEDLINE database was searched for terms related to adherence, non-inferiority trials and statistical methods for handling non-adherence in the titles, abstracts and keywords of papers published up to 31 December 2020 (full search strategy is provided in the online supplemental appendix 1). Eligibility based on identifying appropriate statistical methods was assessed using a three-stage process. First, two authors independently reviewed the title and abstract of each paper. Those where the comparison was not randomised, the primary analysis was not for non-inferiority, or the analysis assessed cost-effectiveness were excluded (cost-effectiveness analyses were not of interest because the focus of this review was on estimating the efficacy of interventions). Papers not published in English were

also excluded. If the full text was unavailable, the abstract was reviewed against the eligibility criteria to ensure that key papers were not excluded. Next, an automated search of the full texts was performed in order to identify those containing the terms ‘compliance’, ‘adherence’ or ‘complier’. Finally, full-text reviews of the remaining papers were performed independently by two authors to identify (1) randomised trials with a primary analysis for non-inferiority that applied (or planned to apply, for protocol papers) statistical methods to account for the impact of non-adherence to interventions, and (2) methodology papers that described such statistical methods and included a non-inferiority trial application. Any discrepancies between reviewer pairs were discussed with a third author in order to reach a consensus. In addition, statisticians within the field were consulted in order to identify key publications, and the reference lists and citations of eligible papers searched for relevant analyses (performed by one author (MD)). Meta-analyses and systematic reviews identified were also searched for eligible non-inferiority trials. Where a trial’s published protocol and results paper were both eligible and reported the same statistical method of interest, the protocol paper was excluded to avoid double counting. Statistical analysis plans were requested for all eligible trials.

A standardised electronic form was used to extract the relevant information from each paper considered eligible. This included details of the trial characteristics (journal, year of publication, disease area or patient population, unit of randomisation, type of experimental intervention, type of primary outcome and non-inferiority margin), non-adherence to the interventions (definitions and estimated levels of non-adherence), the statistical method attempting to account for non-adherence (name of the method, estimand, estimate of effect and confidence interval (CI), conclusion regarding non-inferiority and any advantages/disadvantages of the method stated) and any other analyses applied to the same outcome (analysis population, estimand, estimate of effect and CI, and conclusion regarding non-inferiority). Data extraction was performed by one author (MD). The primary outcome was the statistical method applied (or planned to be applied) in order to account for non-adherence to the interventions. Other outcomes were the impact of applying these methods on the trial conclusions (compared with other analyses applied to the same outcome, where available) and the advantages and disadvantages of the methods where stated by the authors. The impact of applying the methods of interest was assessed using trial results papers only.

This systematic review was registered with PROSPERO and conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.²⁶ Information was largely combined using a narrative synthesis approach, that is, ‘synthesis of findings from multiple studies that relies primarily on

the use of words and text to summarise and explain the findings of the synthesis’.²⁷ All analyses were conducted using Stata V.15.1.

Patient and public involvement

Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.

RESULTS

After removing duplicate publications, our search identified 3235 papers. Of these, 934 were excluded following review of the titles and abstracts, 790 did not contain any keywords in the full texts and 1489 were excluded after full-text review, leaving 22 papers whose citations and reference lists contained a further 5 papers meeting eligibility criteria. After removing publications of the same trial reporting identical statistical methods of interest, 24 papers remained (figure 1).

The 24 publications, which consisted of 4 protocols, 13 results papers and 7 methodology papers, reported relevant methods on 26 occasions (2 methodology papers both contained 2 relevant analyses). Four of the analyses included in methodology papers were re-analyses of non-inferiority trials, one included a simulation study based on a non-inferiority trial and four included simulation studies not based on real trials. Fifteen of the 24 papers included (63%) were published within the last 5 years and the most common type of experimental intervention studied was drug interventions (35%) (table 1; online supplemental table A1).

Non-adherence to interventions

Non-adherence to the randomly assigned interventions was defined in the methods, statistical analysis plan or results section of most analyses (n=19, 73%). Fifteen (79%) used a binary definition of adherence, whereas 3 (16%) used a continuous measure (one was unclear). Of the 19 analyses that defined non-adherence to the interventions, 13 reported estimates of non-adherence (the remaining 6 were protocols or simulation studies). More than half reported estimates of non-adherence that were no more than 10%, though the range was wide (1.7%–51.3%) (table 2). For reasons that were not reported, two papers provided data on non-adherence to the interventions in only one arm of the trial.

Statistical methods for handling non-adherence to interventions

In total, 11 different statistical methods that attempt to account for non-adherence to interventions were identified (table 3). The most common were IV approaches (n=9, 35%), including observed adherence as a covariate within a regression model (n=3, 12%), and modelling adherence as a time-varying covariate in a time-to-event analysis (n=3, 12%). Other methods included rank preserving structural failure time models

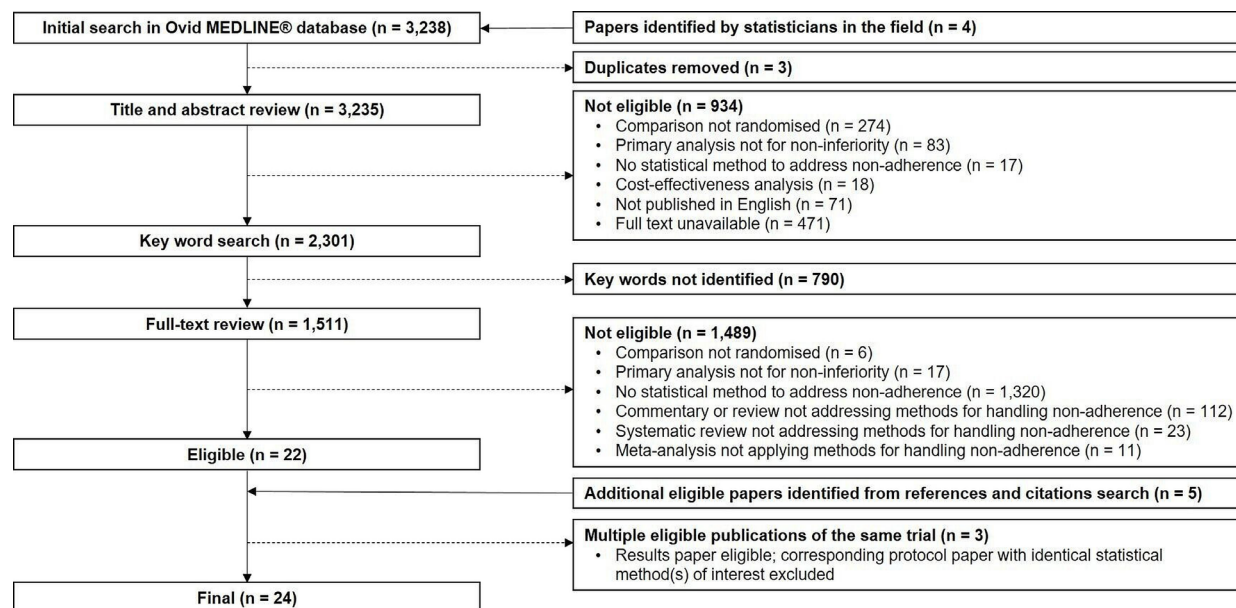


Figure 1 Flow chart showing the eligibility of papers reviewed (uploaded separately).

and G-estimation (n=2, 8%), inverse-probability-of-treatment weighting (n=2, 8%) and the tipping point approach (n=2, 8%, both in the same methodology paper). The other five techniques identified were all reported once. Further details of the methods reported more than once are provided in [table 4](#) and online supplemental table A2. The techniques identified in the 17 protocols and results papers were more commonly specified as sensitivity analyses (n=13, 76%) than primary analyses (n=3, 18%) (one was unclear).

Advantages and disadvantages of the statistical methods

The advantages and disadvantages of the methods identified (as stated by the authors) are given in [table 3](#). Advantages or disadvantages of the techniques used were stated in 8 (33%) of the 24 papers included; 6 were methodological papers and 2 were results papers. No advantages or disadvantages were stated for 5 of the 11 methods identified.

Impact of the statistical methods on non-inferiority conclusions

Twelve of the 13 results papers (92%) also included an alternative analysis of the same outcome (online supplemental table A3). All 12 performed an ITT or mITT analysis. In addition, some reported results from PP (n=6, 50%) or as-treated (AT; n=2, 17%) analyses. Non-inferiority conclusions from the alternate analyses were in agreement with those from the methods of interest on six occasions and could not be compared on six occasions (due to different measures of effect or the results not being provided in full). Five of the six analyses where the different methods were in agreement concluded non-inferiority of the experimental intervention versus the comparator. The remaining trial provided mixed findings regarding non-inferiority across the two different countries included, though the

interpretation of this study appeared inconsistent with its design (a CI approach to determining non-inferiority was stated in the methods but not used).

Statistical analysis plans

Statistical analysis plans were requested for all 17 non-inferiority trials where the protocol or results paper was included in the review, and obtained for nine of these trials.

DISCUSSION

To the best of our knowledge, this is the first systematic review undertaken to both identify statistical methods that adjust for the impact of non-adherence to interventions in randomised non-inferiority trials and also identify the frequency and consequences of their use. We found that few papers reported such methods (less than 2% of those reaching full-text review). This may be partly due to unfamiliarity with such techniques among trialists and statisticians as a result of the long lead time for statistical methodology to make its way into routine practice. The most common techniques identified were IV approaches, including observed adherence as a covariate within a regression model, and modelling adherence as a time-varying covariate in a time-to-event analysis. Overall, the number of trials implementing relevant statistical methods was too small to draw firm inferences about their impacts on non-inferiority conclusions. In six analyses where the results from methods of interest could be compared directly with those from an alternative analysis, conclusions regarding non-inferiority were consistent across the different approaches.

Almost half of the methods identified focus on estimating CACE (also known as the local average treatment effect (LATE)). This is the average effect of the

Table 1 Characteristics of eligible analyses (n=26)

Characteristics	n (%)
Type of publication (n=24)	
Results	13 (54)
Methodology	7 (29)
Protocol	4 (17)
Year of publication (n=24)	
2006–2010	5 (21)
2011–2015	4 (17)
2016–2020	15 (63)
Disease area or patient population	
Mental health	4 (15)
Appendicitis	2 (8)
Cancer	2 (8)
Respiratory infection/disease	2 (8)
Ulcerative colitis	2 (8)
Anaemia	1 (4)
Ear infection	1 (4)
General surgery patients	1 (4)
Heart disease	1 (4)
HIV	1 (4)
Individuals receiving life-sustaining therapies	1 (4)
Renal disease	1 (4)
Smoking cessation	1 (4)
Throat infection	1 (4)
Urinary incontinence	1 (4)
Simulation study	4 (15)
Unit of randomisation	
Individual	19 (73)
Cluster	3 (12)
Simulation study	4 (15)
Type of experimental intervention	
Drug	9 (35)
Method of treatment delivery	3 (12)
Additional patient examination	2 (8)
Nutritional	2 (8)
Surgical	1 (4)
Simulation study	4 (15)
Other	5 (19)
Type of outcome	
Binary	13 (50)
Continuous	8 (31)
Time to event	4 (15)
Count	1 (4)
Composite outcome	3 (12)

experimental intervention within the subpopulation of compliers.²⁵ We argue that this is the natural estimation focus when attempting to account for non-adherence to interventions in the context of non-inferiority trials. This is because we want to be confident that there is non-inferiority among those who would comply with

either intervention. By contrast, including participants who would not fully adhere to both interventions may bias estimation towards non-inferiority (in a similar way that, in the context of non-inferiority, ITT analyses may be biased towards non-inferiority under non-adherence). For similar reasons, we believe that the CACE is preferable to the population average treatment effect (ATE). Lastly, we note that when adjusting for observed adherence within a regression model or modelling adherence as a time-varying covariate in a time-to-event analysis, the target estimand is unclear.

The infrequent use of statistical methods for handling non-adherence seen in the current review has also been observed more generally in randomised controlled trials (RCTs). A review of 100 RCTs randomly selected from those published in 4 high-impact journals during 2008 found only 1 that attempted to account for non-adherence to interventions using a causal inference framework (in which inverse-probability-of-censoring weighting was applied).⁶ More recently, Mostazir *et al* conducted a review of statistical approaches for handling non-adherence to interventions in RCTs published between 1991 and 2015, which identified 88 analyses incorporating 9 different methods.²⁸ IV methods were among the most common and accounted for almost one in four applications of suitable techniques. However, some of the other methods identified (including CACE analyses using maximum-likelihood estimation and adjusted treatment received models) were not captured in the current review focusing on non-inferiority trials. Similarly, we did not identify all 12 approaches included in a recent review of methodological papers containing statistical techniques for handling non-adherence to interventions in the context of time-to-event outcomes.²⁹ This suggests that other relevant methods are available but either they are not suitable for comparing active interventions, as is often required in non-inferiority trials, or they may not have been applied within these studies. The three aforementioned reviews did not focus specifically on non-inferiority trials.

It is perhaps not surprising that IV approaches were the most common method identified in the current review, given that their assumptions are well suited to many double-blind trials, they can be applied across a range of trial designs, and they are relatively simple to implement in standard statistical software.³⁰ IV methods use randomisation as the instrument in order to account for unmeasured confounders of the outcome and intervention received (ie, adherence). Their main assumptions are: (1) randomisation affects the outcome only through its influence on the intervention received (the exclusion restriction), (2) randomisation does not share common causes with the outcome (the exchangeability assumption), (3) randomisation causes some participants to receive their assigned intervention (the relevance assumption) and, in order to estimate CACE, (4) there are no participants who would always receive the opposite of their random allocation (the monotonicity assumption).^{23 31} In individually randomised trials, the exclusion restriction and



monotonicity assumptions are typically satisfied by effective double blinding and/or use of objective outcomes, and the exchangeability assumption is usually valid since randomisation is expected to produce trial arms that are balanced with respect to prognostic factors.

When these assumptions hold, it is relatively straightforward to show that if we regress the intervention received (ie, adherence) on randomisation, and then use this model to predict each participant's adherence, these predictions are orthogonal (independent) of all adherence–outcome confounders. Therefore, if in a second step we regress the outcome on these predictions, we get an unconfounded estimate of the effect of adherence on outcome. It follows that, in contrast to techniques that involve inverse-probability weighting, when the above four IV assumptions hold, IV methods enable us to estimate CACE even in the presence of unmeasured confounding (although inclusion of measured confounders can improve precision).^{32 33} While IV methods may thus appear a panacea, as usual in statistics, there are no free lunches: a lack of precision and statistical power is often a challenge with IV techniques and methods used to adjust for non-adherence more generally.^{5 30 34 35}

The two-stage least-squares (2SLS) regression approach sketched in the previous paragraph can be applied when the intervention is not all or nothing. Suppose that a non-inferiority trial is conducted to assess whether prescribing one dose of a medication per week is non-inferior to prescribing two doses per week over the course of 4 weeks. For each participant, the monotonicity assumption requires that the potential number of doses taken would be lower if the participant was randomly assigned to receive one dose per week than if they were randomised to receive two doses per week. Assuming there are no covariates and the monotonicity assumption holds, it can be shown that the 2SLS estimator converges toward a weighted average of the causal effects of one unit increases in the intervention among compliers (individuals whose intervention intensity is affected by randomisation (the instrument)).^{36 37} This is because the implicit effect of the 2SLS analysis is that values of the outcome at which there are more compliers get given greater weight.

A limitation of IV methods is that when interventions are administered at multiple timepoints, standard approaches are susceptible to time-varying confounding and selection bias.²¹ These biases occur when previous values of a covariate predict the current intervention received and the current value of the covariate predicts outcome.³⁸ If the time-varying confounders are themselves affected by previous intervention received, so-called G-methods, such as inverse-probability weighting or G-estimation, are required to allow for the feedback loop occurring between the intervention received and confounders over time.^{21 24 39} G-methods were seldom reported in the current review, perhaps because they can be more complex to implement than alternative approaches and also rely on assumptions which may be vulnerable to violations. When considering whether to apply an IV approach or a G-method, statisticians might consider whether the exclusion restriction and monotonicity assumptions are realistic given the context of the trial, and whether randomisation is a sufficiently strong instrument. Where outcomes are collected at multiple timepoints, inverse-probability weighting may be a more attractive approach if data on potential confounders are also collected throughout follow-up.

In order to estimate the effect of the experimental intervention in the absence of (full) protection by randomisation, additional assumptions must be made. Most of these are, by their nature, inherently untestable. Each of the statistical methods identified in the current review make slightly different assumptions in order to estimate the effects of interventions under full adherence, and hence each has a different method of estimation; both assumptions and estimation methods have associated advantages and disadvantages. Crucially, all of the methods require reliable information regarding adherence to the randomly assigned interventions, which is often challenging to measure, particularly for long-term therapies.⁴⁰

Despite these limitations, it is our view that the methods identified have an important role in non-inferiority trials with non-adherence to interventions and should be applied as sensitivity analyses

Table 2 Estimates of non-adherence to interventions reported in methodology and results papers, combined across trial arms unless reported (n=13)

Estimate of non-adherence	Binary measure of adherence (n=11)	Continuous* measure of adherence (n=2)	Binary or continuous* measure of adherence (n=13)
≤5%	4 (36)	0 (0)	4 (31)
6%–10%	4 (36)	0 (0)	4 (31)
11%–25%	0 (0)	1 (50)	1 (8)
26%–50%	2 (18)†	1 (50)	3 (23)
>50%	1 (9)†	0 (0)	1 (8)

Data presented as n (%).

*Mean level of non-adherence.

†Two papers provided an estimate of non-adherence in only one arm of the trial.

Table 3 Statistical methods that were identified as attempting to account for non-adherence to interventions

Method (estimand*)	Brief description	n (%)	Advantages†	Disadvantages†
IV approaches (CACE)	CACE estimated using the conventional IV estimator, the generalised method of moments IV estimator or 2SLS regression	9 (35)	<ol style="list-style-type: none"> 1. Straightforward to compute.³⁰ 2. Preserves the balance in patient characteristics from randomisation.³⁰ 3. Although the validity of the IV estimator depends on several key assumptions, these assumptions are likely to hold in most double-blinded studies.³⁰ 4. Correctly adjusts for missing outcome data, assuming that these are MAR.⁴¹ 5. Can account for unknown confounders.¹¹ 6. Recent methods using doubly robust procedures have been developed to boost power when using IV estimation.¹¹ 	<ol style="list-style-type: none"> 1. Accurate data on compliance behaviour must also be available.³⁰ 2. The IV method ... does increase the sample size requirements of the study as the expected proportion of non-compliers increases.³⁰ 3. Requires the 'exclusion restriction' to be fulfilled (ie, treatment allocation only influences the outcome through the treatment and not through any other pathways). This assumption is unverifiable and we are only likely to be confident that it holds in a double-blinded study.¹¹
Adjustment for observed adherence (estimand unclear)	Observed adherence included as a covariate within a regression model	3 (12)	None stated	None stated
Adherence modelled as a time-varying covariate in a time-to-event analysis (estimand unclear)	Attributes the time at risk between observations and the outcomes occurring during the same period to the concurrent value of adherence	3 (12)	None stated	<ol style="list-style-type: none"> 1. Results ... (from the) time-dependent analysis could be the consequence of a selection bias.⁴²
Tipping point approach (estimates the probability of reversing the trial conclusions under a range of assumptions about the outcome data following non-adherence)	Outcome data following non-adherence is treated as missing. Assesses how sensitive the trial results are to the values of these missing outcomes	2 (8)‡	<ol style="list-style-type: none"> 1. A model or mechanism for the missing outcomes does not have to be assumed.⁴³ 2. All randomised individuals are included in the analysis.⁴³ 	<ol style="list-style-type: none"> 1. Require(s) assumptions that are often difficult or not possible to verify.⁴³
Rank preserving structural failure time model and G-estimation (ATE)	Untreated survival times (those that would occur if no EXP were received) are assumed to be a function of both time on/off EXP and the effect of EXP compared with CON. The value for the effect of EXP that results in equal untreated survival times in the randomised groups is identified (via G-estimation) and used to calculate adjusted survival times that would have been observed had no switching occurred	2 (8)	<ol style="list-style-type: none"> 1. Takes compliance history into account.⁴⁴ 2. Maintains the original randomised group.⁴⁵ 3. G-estimation can be performed for other types of outcomes, such as continuous, binary or count responses, using structural nested mean models.⁴⁵ 	<ol style="list-style-type: none"> 1. Can only be used under a specific set of assumptions.⁴⁴ 2. The rank preserving structural failure time model ... incorporates a strong non-interaction assumption with respect to the treatment effect.⁴⁵
Inverse-probability-of-treatment weighting (ATE)	A pseudo-population is created by re-weighting participants' outcomes according to the probability of adherence at each visit, given previous values of the intervention received and confounders. The causal effect of EXP is estimated by performing an unadjusted analysis in the pseudo-population, which is equivalent to a weighted analysis in the original cohort	2 (8)	<ol style="list-style-type: none"> 1. Ensure(s) that the reweighted arms are similar and comparable.¹¹ 2. Sensitivity analysis methods are available to address unobserved confounding and covariate measurement errors.¹¹ 	<ol style="list-style-type: none"> 1. It eliminates bias if all confounders can be appropriately adjusted for, but in general this will not be possible.¹¹

Continued

Table 3 Continued

Method (estimand*)	Brief description	n (%)	Advantages†	Disadvantages‡
Structural mean models (CACE)	Baseline variables that predict adherence differentially in each arm of the trial and are also conditionally independent of outcome are identified. Enables the estimation of two distinct causal parameters, from which a contrast can be made	1 (4)	1. Straightforward to implement using standard statistical software ³⁵	1. This paper highlights the increase in variance experienced when fitting these models, something that can only be reduced when the models include strong predictors of adherence and outcome. ³⁵ 2. Extension of the approach to handle non-linear outcome variables is also required ³⁵
CACE analysis using propensity score approach (CACE)	A propensity score is developed in the EXP group in order to predict the probability that those in the CON group would have been fully adherent if assigned to EXP. CON group outcomes are re-weighted using these probabilities and compared with the outcomes of those who were fully adherent in the EXP group	1 (4)	None stated	None stated
CACE analysis using a mixture modelling approach (CACE)	A mixture model is used to identify those in the CON group that are likely to have been fully adherent had they been assigned to EXP. Outcomes in this subgroup are compared with outcomes of those who were fully adherent in the EXP group	1 (4)	None stated	None stated
Test statistic based on the OR in CRTs (estimand unclear)	A test statistic for assessing non-inferiority based on the OR in CRTs under the Dirichlet multinomial model	1 (4)	None stated	None stated
CACE analysis (CACE)	Exact method not stated	1 (4)	None stated	None stated

*Intercurrent event component of the estimand.

†As stated by the authors.

‡Both applications of the tipping point approach were reported in the same methodology paper.

AIE; average treatment effect in the population; CACE; complier average causal effect; CON; control; CRT; cluster randomised trial; EXP; experimental intervention; IV; instrumental variable; MAR, missing at random; OR; odds ratio; 2SLS; two-stage least squares.

Table 4 Details of the statistical methods reported on more than one occasion

Method	Description	Key assumptions	Advantages	Disadvantages
IV approaches ^{23 31}	<p>With two randomised groups (Z; Z=0 for those in the CON group and Z=1 for those in the EXP group), a binary measure of the intervention received (X; X=1 when EXP received and X=0 when CON received), and a continuous outcome (Y), the conventional IV estimator of CACE is:</p> $IV = \frac{E[Y Z=1] - E[Y Z=0]}{\Pr(X=1 Z=1) - \Pr(X=1 Z=0)}$ <p>Alternatively, 2SLS regression can be used in the presence of confounding by baseline covariates (C). In the first stage, X is regressed on Z and C. In the second stage, Y is regressed on both the predicted value of X obtained from the first stage and C. The coefficient for the effect of X on Y in the second stage provides an estimate of CACE</p>	<p>The instrument (randomisation):</p> <ol style="list-style-type: none"> Affects the outcome only through the intervention received (the exclusion restriction). Does not share common causes with the outcome (the exchangeability assumption). Causes some participants to receive their assigned intervention (the relevance assumption). <p>Additional assumption required to estimate CACE:</p> <ol style="list-style-type: none"> There are no participants who would always receive the opposite of their random allocation (the monotonicity assumption) 	<ol style="list-style-type: none"> Preserves randomised comparison.³⁰ Assumptions well suited to double-blinded trials (typically assumptions 1 and 4 are satisfied by effective double blinding and use of objective outcomes, and assumption 2 valid due to randomisation).³² Under assumptions 1–4, able to estimate CACE even when unmeasured confounding is present.³² Inclusion of confounders in 2SLS regression can improve precision.³³ Can be extended to allow for partial adherence, binary or time-to-event outcomes, and clustering, though additional assumptions may be required.^{19 36 46–48} 	<ol style="list-style-type: none"> Requires untestable assumptions (1 and 4) that may be violated. When adherence is binary, the exclusion restriction implies no effect of EXP in those who are non-adherent. Only appropriate when crossovers occur and cannot be used when non-trial interventions are received.⁵ The sample size required to maintain statistical power to detect non-inferiority increases as the quantity of non-adherence increases.⁵ Simple approaches described involve a single measure of the intervention received (eg, >80% of sessions attended versus <80%, or the proportion of sessions attended) and are susceptible to time-varying confounding (where predictors of both adherence and outcome vary over time)
Adjustment for observed adherence	<p>Observed adherence at a fixed timepoint (eg, whether surgery was performed) or multiple timepoints (eg, whether medication was taken adequately between follow-up visits) included as a covariate within a regression model</p>	<ol style="list-style-type: none"> Individuals in the EXP and CON groups with the same level of observed adherence are comparable. Within each trial arm, the effect of adherence on the outcome is the same. The functional form of adherence is correctly specified. Other model assumptions are not violated 	<ol style="list-style-type: none"> Relatively straightforward to implement 	<ol style="list-style-type: none"> Susceptible to selection bias and should be avoided. Different factors may lead to non-adherence in the two groups, meaning those in the EXP group with an observed level of adherence may differ from those in the CON group with the same level of adherence. Also, some of those in the CON group that were non-adherent may have been fully adherent if assigned to the EXP group (and vice versa).³³ Does not account for time-varying confounding

Continued

Table 4 Continued

Method	Description	Key assumptions	Advantages	Disadvantages
Adherence modelled as a time-varying covariate in a time-to-event analysis ⁴⁹	An extension of the Cox PH model that allows the intervention received to vary over time. The model takes the form: $\lambda_i(t) = \lambda_0(t) \exp(\beta X_i(t))$ where $\lambda_i(t)$ is the hazard function at time t , $\lambda_0(t)$ is the baseline hazard and $X_i(t)$ takes the value 0 while a participant receives CON and 1 while they receive EXP. β is the log HR for the effect of receiving EXP vs receiving CON.	<ol style="list-style-type: none"> 1. Only the current value of the intervention (at time t) affects the hazard. 2. The effect of receiving EXP is the same for all participants regardless of when it is received. 3. Other assumptions of the Cox PH model (including uninformative censoring) are not violated 	<ol style="list-style-type: none"> 1. Allows for time-varying confounding that is not influenced by previous intervention received. 2. Can be extended to allow for more flexible measures of the intervention received, such as cumulative exposure to the intervention. 3. Non-trial interventions may be incorporated 	<ol style="list-style-type: none"> 1. Assumptions 1 and 2 are difficult to verify and may be violated 2. Susceptible to selection bias if switching is related to prognostic factors⁴⁹ 3. Does not account for time-varying confounding that is influenced by previous intervention received
Tipping point approach ⁴³	Outcome data following non-adherence is treated as missing. A range of assumptions about these outcomes are explored to assess how sensitive the trial results are to the missing values	Assumptions made about the values of missing outcomes following non-adherence, for example, all missing outcomes are (1) failures, (2) successes in the CON group and failures in the EXP group (worst case scenario) or (3) failures in the CON group and successes in the EXP group (best case scenario)	<ol style="list-style-type: none"> 1. All randomised individuals are included in the analysis.⁴³ 2. A model or mechanism for the missing outcomes does not have to be assumed⁴³ 	<ol style="list-style-type: none"> 1. While a range of assumptions about the missing values can be explored, these assumptions are often difficult or not possible to verify⁴³
Rank preserving structural failure time model and G-estimation ^{49,50}	Let T_i denote the observed survival time for the i th participant and U_i their survival time that would have been observed if they received no EXP. T_i is assumed to be a function of time on (T_{OFF_i}) and time off (T_{OFF_i}) EXP and T_i related to U_i by the causal model: $U_i = T_{OFF_i} + e^{-\psi_0} T_{ON_i}$ e^{ψ_0} is the amount by which expected survival times are increased by EXP (the acceleration factor). Due to randomisation, U_i is assumed to be independent of trial arm. Untreated event times are predicted for all participants and the value of ψ_0 that results in equal untreated survival times u_i the randomised groups identified (using G-estimation). This value of ψ_0 is used to calculate adjusted survival times that would have been observed had switching not occurred	<ol style="list-style-type: none"> 1. If no participants received any EXP, the average survival time in the two groups would be equal (due to randomisation). 2. The effect of receiving EXP is the same for all participants regardless of when it is received. 3. If participant i experiences the event of interest before participant j when both are treated, then participant i would also experience the event before participant j when both are untreated (rank preserving) 	<ol style="list-style-type: none"> 1. Preserves randomised comparison.⁴⁵ 2. Takes adherence history into account.⁴⁴ 3. Allows for time-varying confounding, including when these confounders are affected by previous intervention received. 4. Does not require information on potential confounders (only randomised group, observed event times and intervention history)⁵¹ 	<ol style="list-style-type: none"> 1. Assumptions 2 and 3 are difficult to verify and may be violated.⁴⁹ 2. Additional assumptions required when the CON group receive an active intervention.⁵⁰ 3. Only appropriate when crossovers occur and cannot be used when non-trial interventions are received.⁵¹ 4. G-estimation may not work well if the number of participants or events is small.⁵⁰ 5. Computationally intensive⁵²

Continued

Table 4 Continued

Method	Description	Key assumptions	Advantages	Disadvantages
Inverse-probability-of-treatment weighting ^{24,52-54}	Confounding is accounted for by re-weighting participants' outcomes. Typically, logistic regression is used to predict probabilities of adherence at each visit given previous values of the intervention received and confounders. The inverse of these probabilities (the weights) are used to create a pseudo-population in which time-varying confounders are not associated with the intervention. The causal effect of EXP is estimated by performing an unadjusted analysis in the pseudo-population, which is equivalent to a weighted analysis in the original cohort.	<ol style="list-style-type: none"> 1. The absence of unmeasured confounding (exchangeability). 2. Participants have a non-zero probability of receiving each intervention (positivity). 3. The observed outcome equals the counterfactual outcome of the intervention actually received (consistency). 4. No misspecification of the model used to estimate the weights 	<ol style="list-style-type: none"> 1. Preserves randomised comparison.¹¹ 2. Takes adherence history into account. 3. Allows for time-varying confounding, including when these confounders are affected by previous intervention received. 4. Can be extended to allow for non-trial interventions, but may result in large weights and estimates very sensitive to model specification⁵³ 	<ol style="list-style-type: none"> 1. Eliminates bias if all confounders can be appropriately adjusted for, but in general this will not be possible.¹¹ 2. Cannot be used if covariates perfectly predict adherence.⁵² 3. Unstable in the presence of extreme weights⁵²

CACE, complier average causal effect; CON, control; EXP, experimental intervention; HR, hazard ratio; IV, instrumental variable; PH, proportional hazards; 2SLS, two-stage least squares.

alongside other techniques (such as ITT, PP and AT analyses). Given that agreement between ITT and PP analyses cannot guarantee unbiased conclusions in these studies, those with non-trivial non-adherence should assess the sensitivity of trial results to different assumptions in order to guard against falsely claiming non-inferiority and accepting a worse intervention. Careful consideration needs to be given to the assumptions that are most plausible given the trial context and planned design, before selecting the appropriate statistical method to adjust for potential non-adherence. Relevant data needed to implement the chosen technique should then be collected as fully as possible. Clearly, the best approach to reducing the potential biases introduced by non-adherence to the interventions is to design trials that minimise such non-adherence. Future work should compare the performance of the methods identified under different non-adherence scenarios in non-inferiority trials to facilitate understanding of when they might be applied appropriately.

Strengths and limitations

This is the first systematic review (protocol published on PROSPERO) to identify statistical methods that attempt to account for the impact of non-adherence to interventions in randomised non-inferiority trials, quantify the use of such methods in these studies and examine their impact on non-inferiority conclusions. The review included publications from any year, journal or disease area and involved two authors agreeing the eligibility of each paper identified in the search. However, it has some limitations. First, the search for eligible papers had to be restricted to those containing terms related to adherence or statistical methods for handling non-adherence in the titles, abstracts and keywords. Publications applying suitable methods in sensitivity analyses may not have referred to the techniques within these fields and, if so, would not have been captured by the search. A wide range of search terms were used to try and mitigate this problem and, therefore, a large number of papers were reviewed. Second, only one database was searched and papers not published in English were excluded meaning it is possible that some eligible papers may not have been captured by the search. However, most major non-inferiority trials are likely to be published in English within one of the MEDLINE journals and, therefore, should have been captured. Third, one author performed the data extraction from eligible publications, though other authors were consulted where necessary. Finally, while statistical analysis plans were requested, these could not be obtained for eight of the trials included. For these studies, we cannot be sure that the details provided in the publications reviewed are accurate accounts of the planned analyses.

CONCLUSION

In non-inferiority trials with non-adherence to interventions, ITT and PP analyses are often performed but may result in biased estimates of efficacy and, therefore, agreement between these approaches does not guarantee that conclusions regarding non-inferiority are unbiased. Statistical methods that attempt to account for the impact of non-adherence and thereby estimate the causal effects of interventions are available, but their use in non-inferiority trials remains extremely infrequent. It is our view that the methods identified should be applied more widely within sensitivity analyses of non-inferiority trials. In particular, those with non-trivial non-adherence should assess the sensitivity of trial results to different assumptions in order to guard against falsely claiming non-inferiority and accepting a worse intervention.

Contributors MD devised the review, reviewed the eligibility of papers identified, carried out data extraction and analysis, wrote the initial manuscript, and is responsible for the overall content as the guarantor. KF, JRC, JAT and DE commented on the review protocol, reviewed the eligibility of papers identified and assisted with writing the final manuscript. All authors provided final approval of the version to be published.

Funding JRC is supported by the Medical Research Council (MRC) Clinical Trials Unit (grant numbers: MC_UU_12023/21 and MC_UU_12023/29). JAT is jointly funded by the UK MRC and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union (grant number: MR/R010161/1).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethical approval was not required because this review used publicly accessible documents.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. No additional data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) 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, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Matthew Dodd <http://orcid.org/0000-0002-6207-6604>

Katherine Fielding <http://orcid.org/0000-0002-6524-3754>

James R Carpenter <http://orcid.org/0000-0003-3890-6206>

Jennifer A Thompson <http://orcid.org/0000-0002-3068-3952>

Diana Elbourne <http://orcid.org/0000-0003-3044-4545>

REFERENCES

- Piaggio G, Elbourne DR, Altman DG, *et al.* Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152–60.
- Suda KJ, Hurley AM, McKibbin T, *et al.* Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy* 2011;31:833–9.
- Bikdeli B, Welsh JW, Akram Y, *et al.* Noninferiority designed cardiovascular trials in Highest-Impact journals. *Circulation* 2019;140:379–89.
- Piaggio G, Elbourne DR, Pocock SJ, *et al.* Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *Jama* 2012;308:2594–604.
- Mo Y, Lim C, Watson JA, *et al.* Non-adherence in non-inferiority trials: pitfalls and recommendations. *BMJ* 2020;370:m2215.
- Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. *Trials* 2012;13:84.
- DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med Care* 2004;42:200–9.
- DiMatteo MR, Giordani PJ, Lepper HS, *et al.* Patient adherence and medical treatment outcomes: a meta-analysis. *Med Care* 2002;40:794–811.
- Simpson SH, Eurich DT, Majumdar SR, *et al.* A meta-analysis of the association between adherence to drug therapy and mortality. *BMJ* 2006;333:15.
- 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.
- Mo Y, Lim C, Mukaka M, *et al.* 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.
- Committee for Proprietary Medicinal Products. Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol* 2001;52:223–8.
- Rehal S, Morris TP, Fielding K, *et al.* Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open* 2016;6:e012594.
- Matilde Sanchez M, Chen X. Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Stat Med* 2006;25:1169–81.
- Brittain E, Lin D. A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Stat Med* 2005;24:1–10.
- Clarke PS, Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. *Biostatistics* 2010;11:756–70.
- Anon. ICH Harmonised tripartite guideline. statistical principles for clinical trials. International Conference on harmonisation E9 expert Working group. *Stat Med* 1999;18:1905–42.
- Gupta S. Intention-to-treat concept: a review. *Perspect Clin Res* 2011;2:109–12.
- Nagelkerke N, Fidler V, Bernsen R, *et al.* Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med* 2000;19:1849–64.
- Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ* 2010;340:c2697.
- Hernán MA, Robins JM. Per-Protocol analyses of pragmatic trials. *N Engl J Med* 2017;377:1391–8.
- DeMets DL. Statistical issues in interpreting clinical trials. *J Intern Med* 2004;255:529–37.
- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996;91:444–55.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550–60.
- LHY Y, 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.
- Page MJ, Moher D, Bossuyt PM, *et al.* PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;372:n160.
- Popay J, Roberts H, Sowden A, *et al.* *Guidance on the conduct of narrative synthesis in systematic reviews: a product of the ESRC methods programme (version 1)*. Lancaster, UK: University of Lancaster, 2006.
- Mostazir M, Taylor RS, Henley W, *et al.* An overview of statistical methods for handling nonadherence to intervention protocol in randomized control trials: a methodological review. *J Clin Epidemiol* 2019;108:121–31.
- 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 Making* 2019;39:910–25.
- 30 Kim MY. Using the instrumental variables estimator to analyze noninferiority trials with noncompliance. *J Biopharm Stat* 2010;20:745–58.
- 31 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.
- 32 Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. *Emerg Themes Epidemiol* 2018;15:1.
- 33 Stuart EA, Perry DF, Le H-N, *et al.* Estimating intervention effects of prevention programs: accounting for noncompliance. *Prev Sci* 2008;9:288–98.
- 34 Davies NM, Smith GD, Windmeijer F, *et al.* Issues in the reporting and conduct of instrumental variable studies. *Epidemiology* 2013;24:363–9.
- 35 Gillespie D, Farewell D, Barrett-Lee P, *et al.* The use of randomisation-based efficacy estimators in non-inferiority trials. *Trials* 2017;18:117.
- 36 Angrist JD, Imbens GW. Two-Stage least squares estimation of average causal effects in models with variable treatment intensity. *J Am Stat Assoc* 1995;90:431–42.
- 37 Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014;33:2297–340.
- 38 Daniel RM, Cousens SN, De Stavola BL, *et al.* Methods for dealing with time-dependent confounding. *Stat Med* 2013;32:1584–618.
- 39 Hernán MA, Cole SR, Margolick J, *et al.* Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 2005;14:477–91.
- 40 Lam WY, Fresco P. Medication adherence measures: an overview. *Biomed Res Int* 2015;2015:1–12.
- 41 Kitchener HC, Dunn G, Lawton V, *et al.* Laparoscopic versus open colposuspension--results of a prospective randomised controlled trial. *BJOG* 2006;113:1007–13.
- 42 Brunori G, Viola BF, Parrinello G, *et al.* Efficacy and safety of a very-low-protein diet when postponing dialysis in the elderly: a prospective randomized multicenter controlled study. *Am J Kidney Dis* 2007;49:569–80.
- 43 Kim M, Wang C, Xue X. Assessing the influence of treatment nonadherence on noninferiority trials using the tipping point approach. *Stat Med* 2019;38:650–9.
- 44 Wu Y, Zhao L, Hou Y, *et al.* Correcting for non-compliance in randomized non-inferiority trials with active and placebo control using structural models. *Stat Med* 2015;34:950–65.
- 45 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;29:2107–16.
- 46 Palmer TM, Thompson JR, Tobin MD, *et al.* Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008;37:1161–8.
- 47 Tchetgen Tchetgen EJ, Walter S, Vansteelandt S, *et al.* Instrumental variable estimation in a survival context. *Epidemiology* 2015;26:402–10.
- 48 Agbla SC, De Stavola B, DiazOrdaz K. Estimating cluster-level local average treatment effects in cluster randomised trials with non-adherence. *Stat Methods Med Res* 2020;29:911–33.
- 49 Morden JP, Lambert PC, Latimer N, *et al.* Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol* 2011;11:4.
- 50 Latimer NR, Henshall C, Siebert U, *et al.* Treatment switching: statistical and decision-making challenges and approaches. *Int J Technol Assess Health Care* 2016;32:160–6.
- 51 Allison A, White IR, Bond S. rpsftm: an R package for RANK preserving structural failure time models. *R J* 2017;9:342–53.
- 52 Mansournia MA, Etminan M, Danaei G, *et al.* Handling time varying confounding in observational research. *BMJ* 2017;359:j4587.
- 53 Ellis AR, Brookhart MA. Approaches to inverse-probability-of-treatment--weighted estimation with concurrent treatments. *J Clin Epidemiol* 2013;66:S51–6.
- 54 Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;34:3661–79.