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

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.

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 (term non-adherence or non-compliance) 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.\textsuperscript{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 healthcare 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.\textsuperscript{5} In the setting of non-inferiority trials, the intention-to-treat (ITT) and per-protocol (PP) populations have been advocated and are commonly used.\textsuperscript{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).\textsuperscript{14–16}

Standard ITT analyses typically include all participants in their randomised groups irrespective of the intervention actually received.\textsuperscript{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.\textsuperscript{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).\textsuperscript{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.\textsuperscript{11}

Modified ITT (mITT) analyses are commonly used to address some of the limitations with standard ITT methods.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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).\textsuperscript{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.\textsuperscript{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).\textsuperscript{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. 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’. 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, 79%). 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...
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

---

Figure 1 Flow chart showing the eligibility of papers reviewed (uploaded separately).
We argue that this is the natural estimation of the causal effect of the experimental intervention within the subpopulation of compliers. 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 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). 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. 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. 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. 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).

Table 1 Characteristics of eligible analyses (n=26)

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

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 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 consideration.

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.30 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

<table>
<thead>
<tr>
<th>Estimate of non-adherence</th>
<th>Binary measure of adherence (n=11)</th>
<th>Continuous* measure of adherence (n=2)</th>
<th>Binary or continuous* measure of adherence (n=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤5%</td>
<td>4 (36)</td>
<td>0 (0)</td>
<td>4 (31)</td>
</tr>
<tr>
<td>6%–10%</td>
<td>4 (36)</td>
<td>0 (0)</td>
<td>4 (31)</td>
</tr>
<tr>
<td>11%–25%</td>
<td>0 (0)</td>
<td>1 (50)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>26%–50%</td>
<td>2 (18)†</td>
<td>1 (50)</td>
<td>3 (23)</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>1 (9)†</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method (estimand*)</th>
<th>Brief description</th>
<th>n (%)</th>
<th>Advantages†</th>
<th>Disadvantages†</th>
</tr>
</thead>
</table>
| IV approaches (CACE)                                    | CACE estimated using the conventional IV estimator, the generalised method of moments IV estimator or 2SLS regression | 9 (35)| 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 estimation11 | 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 study11 |
| 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 | 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)‡ | 1. A model or mechanism for the missing outcomes does not have to be assumed.43  
2. All randomised individuals are included in the analysis43 | 1. Require(s) assumptions that are often difficult or not possible to verify43 |
| 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) | 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 models45 | 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 effect45 |
| 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) | 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 errors11 | 1. It eliminates bias if all confounders can be appropriately adjusted for, but in general this will not be possible11 |

Continued
<table>
<thead>
<tr>
<th>Method (estimand*)</th>
<th>Brief description</th>
<th>n (%)</th>
<th>Advantages†</th>
<th>Disadvantages†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural mean models (CACE)</td>
<td>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</td>
<td>1 (4)</td>
<td>1. Straightforward to implement using standard statistical software&lt;sup&gt;25&lt;/sup&gt;</td>
<td>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.&lt;sup&gt;26&lt;/sup&gt; 2. Extension of the approach to handle non-linear outcome variables is also required&lt;sup&gt;25&lt;/sup&gt;</td>
</tr>
<tr>
<td>CACE analysis using propensity score approach (CACE)</td>
<td>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</td>
<td>1 (4)</td>
<td>None stated</td>
<td>None stated</td>
</tr>
<tr>
<td>CACE analysis using a mixture modelling approach (CACE)</td>
<td>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</td>
<td>1 (4)</td>
<td>None stated</td>
<td>None stated</td>
</tr>
<tr>
<td>Test statistic based on the OR in CRTs (estimand unclear)</td>
<td>A test statistic for assessing non-inferiority based on the OR in CRTs under the Dirichlet multinomial model</td>
<td>1 (4)</td>
<td>None stated</td>
<td>None stated</td>
</tr>
<tr>
<td>CACE analysis (CACE)</td>
<td>Exact method not stated</td>
<td>1 (4)</td>
<td>None stated</td>
<td>None stated</td>
</tr>
</tbody>
</table>

*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.

ATE, average treatment effect in the population; CACE, compiler 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

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Key assumptions</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV approaches(^{33,31})</td>
<td>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: [ IV = \frac{E[Y</td>
<td>Z=1] - E[Y</td>
<td>Z=0]}{Pr(X=1</td>
<td>Z=1) - Pr(X=1</td>
</tr>
<tr>
<td>Adjustment for observed adherence</td>
<td>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.</td>
<td>1. Individuals in the EXP and CON groups with the same level of observed adherence are comparable. 2. Within each trial arm, the effect of adherence on the outcome is the same. 3. The functional form of adherence is correctly specified. 4. Other model assumptions are not violated</td>
<td>1. Relatively straightforward to implement</td>
<td>1. 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).(^{31}) 2. Does not account for time-varying confounding</td>
</tr>
</tbody>
</table>
### Table 4 Continued

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Key assumptions</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| 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 $\lambda_0(t)$ is the hazard function at time $t$, $\lambda_i(t)$ is the baseline hazard and $X_i(t)$ takes the value 0 while a participant receives CON and 1 while they receive EXP, $\beta$ is the log HR for the effect of receiving EXP vs receiving CON. | 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 | 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 | 1. Assumptions 1 and 2 are difficult to verify and may be violated  
2. Susceptible to selection bias if switching is related to prognostic factors[^49]  
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) | 1. All randomised individuals are included in the analysis.  
2. A model or mechanism for the missing outcomes does not have to be assumed[^43] | 1. While a range of assumptions about the missing values can be explored, these assumptions are often difficult or not possible to verify[^43] |
| Rank preserving structural failure time model and G-estimation[^39][^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 $T_{ON}$ and time off $T_{OFF}$ EXP and $T_i$ related to $U_i$ by the causal model: $U_i = T_{OFF} + e^{-\psi_0} T_{ON}$, $e^{\psi_0}$ 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 $\psi_0$ that results in equal untreated survival times in the randomised groups identified (using G-estimation). This value of $\psi_0$ is used to calculate adjusted survival times that would have been observed had switching not occurred | 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) | 1. Preserves randomised comparison.  
2. Takes adherence history into account.  
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) | 1. Assumptions 2 and 3 are difficult to verify and may be violated.  
2. Additional assumptions required when the CON group receive an active intervention.  
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[^52]  
5. Computationally intensive[^52] |

[^49]: 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). 1. All randomised individuals are included in the analysis. 2. A model or mechanism for the missing outcomes does not have to be assumed[^43] 1. While a range of assumptions about the missing values can be explored, these assumptions are often difficult or not possible to verify[^43] 1. Assumptions 2 and 3 are difficult to verify and may be violated. 2. Additional assumptions required when the CON group receive an active intervention. 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[^52] 5. Computationally intensive[^52]
<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Key assumptions</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse-probability-of-treatment weighting</td>
<td>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.</td>
<td>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.</td>
<td>1. Preserves randomised comparison. 11. 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 specification.</td>
<td>1. Eliminates bias if all confounders can be appropriately adjusted for, but in general this will not be possible. 11. Cannot be used if covariates perfectly predict adherence. 3. Unstable in the presence of extreme weights.</td>
</tr>
</tbody>
</table>

CACE, complier average causal effect; CON, control; EXP, experimental intervention; HR, hazard ratio; IV, instrumental variable; PH, proportional hazards; 2SLS, two-stage least squares.
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_120323/21 and MC_UU_120239/29). 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 EDCTP programme supported by the European Union (grant number: MR/RO10161/1).

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).

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