APPENDIX

The Ovid MEDLINE® database was searched for terms related to treatment adherence, statistical methods for handling non-adherence, and non-inferiority trials in the titles, abstracts and key words of articles published up to 31st December 2020. The full search strategy was as follows:

1. ("intention to treat" or "intention-to-treat" or "ITT" or "intent-to-treat" or "intent to treat").ti,ab,kw.
2. ("as treated" or "as-treated").ti,ab,kw.
3. ("per protocol" or "per-protocol").ti,ab,kw.
4. ("complier average causal effect" or "CACE" or "local average treatment effect").ti,ab,kw.
5. ("non-compliance" or "noncompliance" or "compliance" or "non-adherence" or "nonadherence" or "adherence").ti,ab,kw.
6. ("instrumental variable" or "instrumental variables" or "IV analysis" or "IV analyses").ti,ab,kw.
7. ("non-inferiority" or "noninferiority" or "non-inferior" or "noninferior").ti,ab,kw.
8. 1 or 2 or 3 or 4 or 5 or 6
9. 7 and 8

Table A1. Characteristics of eligible analyses (n=26)

Characteristics	**n (%)**
Journal (n=24)	
The New England Journal of Medicine	4 (17)
Statistics in Medicine	3 (13)
BJOG	2 (8)
Journal of Biopharmaceutical Statistics	2 (8)
Trials	2 (8)
Addiction	1 (4)
American Journal of Kidney Disease	1 (4)
BMC Psychiatry	1 (4)
BMJ Open	1 (4)
Clinical Gastroenterology and Hepatology	1 (4)
Clinical Pediatrics	1 (4)
Food and Nutrition Bulletin	1 (4)
The Journal of the American Medical Association (JAMA)	1 (4)
JAMA Network Open	1 (4)
PLOS Medicine	1 (4)
Wellcome Open Research	1 (4)

Table A2. Explanation of formulae included in Table 4

Method	Formula	Explanation of formula
IV approaches	With two randomised groups (*Z*; *Z*=0 for those in the CON group, *Z*=1 for those in the EXP group), a binary measure of the intervention actually received (*X*; *X*=1 when EXP received, *X*=0 when CON received), and a continuous outcome (*Y*), the conventional IV estimator of CACE is: $$IV = \frac{E[Y\|Z=1] - E[Y\|Z=0]}{\Pr(X=1\|Z=1) - \Pr(X=1\|Z=0)}$$	The conventional IV estimator is equivalent to the ITT effect of *Z* on *Y* (i.e. the ITT effect of randomised group on the outcome) divided by the ITT effect of *Z* on *X* (i.e. the ITT effect of randomised group on the intervention actually received). In other words, the ITT effect of randomised group on the outcome is inflated according to the level of adherence with the randomly assigned interventions in the ITT population. For instance, if the ITT effect of randomised group on the outcome was estimated to be +2% as a risk difference, 95% of those in the CON group received CON, 5% of those in the CON group received EXP, 85% of those in the EXP group received EXP, and 15% of those in the EXP group received CON, the conventional IV estimator of CACE would be: $$IV = \frac{2}{0.85 - 0.05} = 2.5\%$$
Adherence modelled as a time-varying covariate in a time-to-event analysis	An extension of the Cox proportional hazards 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 hazard ratio for the effect of receiving EXP versus receiving CON.	The Cox proportional hazards model is expressed in terms of the hazard function, which is the probability of an outcome occurring at a particular point in time, given that it has not already occurred. The hazard function consists of the baseline hazard (the value of the hazard function when all of the covariates in the model are equal to zero) and the coefficients, which estimate the effects of the covariates. The baseline hazard is able to vary over time and it is assumed that each covariate has a multiplicative effect on the baseline hazard that is constant over time (the proportional hazards assumption). These multiplicative effects are known as hazard ratios. The standard formulation of the Cox model fixes the values of the covariates to be constant over time, but this extension of the method allows the values of the covariates to vary over time. The approach described includes the intervention actually received as the only covariate in the model, the value of which is allowed to vary over time for each participant. If a participant was to

		switch from receiving CON to EXP, their time at risk and outcome status prior to the switch would be attributed to receiving CON, whereas their time at risk and outcome status following the switch would be attributed to receiving EXP (and vice versa). This allows us to estimate the multiplicative effect of receiving EXP compared with receiving CON on the hazard of the outcome.
Rank preserving structural failure time model and G-estimation	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_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 in 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.	Rank preserving structural failure time models are used to account for treatment switching in randomised trials with survival outcomes. The method assumes that each participant's counterfactual event time (i.e. the time to the event which would be observed if they received no EXP) is equal to the amount of time that they do not receive EXP, plus the amount of time that they receive EXP multiplied by the effect of receiving EXP. Due to randomisation, we expect the counterfactual event times to be balanced between the randomised groups (similar to the way that we expect the distribution of participant characteristics, such as age and sex, to be balanced between the randomised groups). The effect of receiving EXP is estimated using each participant's observed times on and off EXP and then trying different values for the effect of receiving EXP until balance is achieved on the counterfactual event times between the randomised groups. This iterative process is known as G-estimation. The effect of receiving EXP versus CON is expressed as the relative amount by which average event times are increased by EXP (known as the acceleration factor). Therefore, an acceleration factor >1 represents longer event times (on average) when receiving EXP and an acceleration factor <1 represents shorter event times (on average) when receiving EXP. The acceleration factor is used to calculate adjusted event times that would have been observed had switching not occurred.

IV = instrumental variable; CON = control; EXP = experimental intervention; CACE = complier average causal effect; ITT = intention-to-treat.

Table A3. Comparison of non-inferiority conclusions from results papers applying statistical methods of interest versus an alternate approach

Reference	Method of interest	Was method of interest applied in primary or sensitivity analysis?	Alternate analyses performed	Comparison of non-inferiority conclusions
Geldsetzer et al. 2018[1]	IV approach	Sensitivity	ITT (primary analysis, all randomised patients)	Different measures of effect
Huang et al. 2018[2]	IV approach	Sensitivity. Applied to PP analysis (patients for whom the trial intervention was achieved at all time points).	ITT (primary analysis, not defined)	Consistent – both concluded non-inferiority
Beaver et al. 2017[3]	IV approach	Primary	ITT (not defined), PP (those who had their first post-randomisation appointment in line with the randomisation), and AT (not defined)	Consistent – all concluded non-inferiority
Kitchener et al. 2006[4]	IV approach	Sensitivity	ITT (primary analysis, not defined)	Consistent – both concluded non-inferiority
Halpern et al. 2020[5]	IV approach	Sensitivity	mITT (primary analysis, those who were randomised, were not subsequently found to be ineligible, and did not withdraw consent)	Different measures of effect
Butler et al. 2019[6]	IV approach	Primary	mITT (those who had undergone randomisation and had available outcome data, regardless of protocol deviations or the intervention they received)	Consistent – both concluded non-inferiority
Bilimoria et al. 2016[7]	IV approach	Sensitivity	ITT (primary analysis, not defined), PP (limited to adherent programs), and AT (which assessed actual exposure to policy change)	Different measures of effect

Reference	Method of interest	Was method of interest applied in primary or sensitivity analysis?	Alternate analyses performed	Comparison of non-inferiority conclusions
Rimoin et al. 2011[8]	Adjustment for observed adherence	Not stated. Applied to ITT (all randomised patients meeting the inclusion and exclusion criteria) and PP analyses (those who were adherent and returned for the follow-up visit).	Unadjusted ITT and unadjusted PP	Consistent – all adjusted analyses resulted in the same conclusions as the corresponding unadjusted analyses
Hahn et al. 2019[9]	Adherence modelled as a time-varying covariate in a time-to-event analysis	Sensitivity	ITT (primary analysis, all randomised patients according to original group allocation) and PP (excluded patients who did not receive the assigned intervention)	Different measures of effect
Dignass et al. 2009[10]	Adherence modelled as a time-varying covariate in a time-to-event analysis	Sensitivity	ITT (all randomised patients who received some intervention and had at least 1 post-baseline efficacy assessment), PP1 (those who dropped out of the study censored at the time of dropout) and PP2 (excluded dropouts)	Results not provided in full
Brunori et al. 2007[11]	Adherence modelled as a time-varying covariate in a time-to-event analysis	Sensitivity	ITT (patients were considered part of the diet group even after a switch to dialysis) and PP (diet group patients were censored when switched to dialysis)	Different measures of effect
Flum et al. 2020[12]	Inverse-probability-of-treatment weighting	Sensitivity. Applied to PP analysis (those adherent to the randomisation assignment).	ITT (primary analysis, not defined)	Consistent – both concluded non-inferiority

CACE = complier average causal effect; IV = instrumental variable; ITT = intention-to-treat; PP = per-protocol; AT = as-treated; mITT = modified intention-to-treat.

REFERENCES

1. Geldsetzer P, Francis JM, Sando D, Asmus G, Lema IA, Mboggo E, et al. Community delivery of antiretroviral drugs: A non-inferiority cluster-randomized pragmatic trial in Dar es Salaam, Tanzania. PLoS Med. 2018 Sep;15(9):e1002659
2. Huang DT, Yealy DM, Filbin MR, Brown AM, Chang CH, Doi Y, et al. Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection. N Engl J Med. 2018 Jul 19;379(3):236-49
3. Beaver K, Williamson S, Sutton C, Hollingworth W, Gardner A, Allton B, et al. Comparing hospital and telephone follow-up for patients treated for stage-I endometrial cancer (ENDCAT trial): a randomised, multicentre, non-inferiority trial. Bjog. 2017 Jan;124(1):150-60
4. Kitchener HC, Dunn G, Lawton V, Reid F, Nelson L, Smith AR. Laparoscopic versus open colposuspension--results of a prospective randomised controlled trial. Bjog. 2006 Sep;113(9):1007-13
5. Halpern SD, Small DS, Troxel AB, Cooney E, Bayes B, Chowdhury M, et al. Effect of Default Options in Advance Directives on Hospital-Free Days and Care Choices Among Seriously Ill Patients: A Randomized Clinical Trial. JAMA Netw Open. 2020 Mar 2;3(3):e201742
6. Butler CC, Gillespie D, White P, Bates J, Lowe R, Thomas-Jones E, et al. C-Reactive Protein Testing to Guide Antibiotic Prescribing for COPD Exacerbations. N Engl J Med. 2019 Jul 11;381(2):111-20
7. Bilimoria KY, Chung JW, Hedges LV, Dahlke AR, Love R, Cohen ME, et al. National Cluster-Randomized Trial of Duty-Hour Flexibility in Surgical Training. N Engl J Med. 2016 Feb 25;374(8):713-27
8. Rimoin AW, Hoff NA, Fischer Walker CL, Hamza HS, Vince A, Abdel Rahman N, et al. Treatment of streptococcal pharyngitis with once-daily amoxicillin versus intramuscular benzathine penicillin G in low-resource settings: a randomized controlled trial. Clin Pediatr (Phila). 2011 Jun;50(6):535-42
9. Hahn JY, Song YB, Oh JH, Chun WJ, Park YH, Jang WJ, et al. Effect of P2Y12 Inhibitor Monotherapy vs Dual Antiplatelet Therapy on Cardiovascular Events in Patients Undergoing Percutaneous Coronary Intervention: The SMART-CHOICE Randomized Clinical Trial. Jama. 2019 Jun 25;321(24):2428-37
10. Dignass AU, Bokemeyer B, Adamek H, Mross M, Vinter-Jensen L, Börner N, et al. Mesalamine once daily is more effective than twice daily in patients with quiescent ulcerative colitis. Clin Gastroenterol Hepatol. 2009 Jul;7(7):762-9
11. Brunori G, Viola BF, Parrinello G, De Biase V, Como G, Franco V, 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 May;49(5):569-80
12. Flum DR, Davidson GH, Monsell SE, Shapiro NI, Odom SR, Sanchez SE, et al. A Randomized Trial Comparing Antibiotics with Appendectomy for Appendicitis. N Engl J Med. 2020 Nov 12;383(20):1907-19