

REVIEW

Open Access



Interpretation of active-control randomised trials: the case for a new analytical perspective involving averted events

David T. Dunn^{1,2*}, Oliver T. Stirrup¹, Sheena McCormack² and David V. Glidden³

Abstract

Active-control trials, where an experimental treatment is compared with an established treatment, are performed when the inclusion of a placebo control group is deemed to be unethical. For time-to-event outcomes, the primary estimand is usually the rate ratio, or the closely-related hazard ratio, comparing the experimental group with the control group. In this article we describe major problems in the interpretation of this estimand, using examples from COVID-19 vaccine and HIV pre-exposure prophylaxis trials. In particular, when the control treatment is highly effective, the rate ratio may indicate that the experimental treatment is clearly statistically inferior even when it is worthwhile from a public health perspective. We argue that it is crucially important to consider *averted* events as well as observed events in the interpretation of active-control trials. An alternative metric that incorporates this information, the averted events ratio, is proposed and exemplified. Its interpretation is simple and conceptually appealing, namely the proportion of events that would be averted by using the experimental treatment rather than the control treatment. The averted events ratio cannot be directly estimated from the active-control trial, and requires an additional assumption about either: (a) the incidence that would have been observed in a hypothetical placebo arm (the counterfactual incidence) or (b) the efficacy of the control treatment (relative to no treatment) that pertained in the active-control trial. Although estimation of these parameters is not straightforward, this must be attempted in order to draw rational inferences. To date, this method has been applied only within HIV prevention research, but has wider applicability to treatment trials and other disease areas.

Keywords Active-control, Non-inferiority, Estimand, Averted, Relative efficacy

Introduction

Active-control trials, in which an experimental treatment is compared with an established treatment, are performed when the inclusion of a placebo control group is deemed to be unethical [1]. For time-to-event

outcomes, the primary estimand is usually the rate ratio or the closely-related hazard ratio [2–5]. Here, we present examples which demonstrate that this estimand can be clinically misleading, and highlight the importance of considering *averted* events as well as observed events [6]. We propose an alternative metric which incorporates the number of averted events, thereby avoiding the limitations of the rate ratio. We introduce the problem with a hypothetical COVID-19 vaccine active-control trial.

*Correspondence:

David T. Dunn
d.dunn@ucl.ac.uk

¹ Institute for Global Health, University College London, London, UK

² MRC Clinical Trials Unit, University College London, 90 High Holborn, London WC1V 6LJ, UK

³ Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Hypothetical COVID-19 vaccine trial

The first licensed COVID-19 vaccine, BNT162b2 (BioNTech/Pfizer), was found to reduce the incidence of COVID-19 by approximately 95% [7]. Imagine that we wished to assess the clinical efficacy of a new COVID-19 vaccine shortly after the licensure of BNT162b2. Given such high clinical efficacy, we conduct a large, active-control trial with 10,000 person-years follow-up per arm, using BNT162b2 as the comparator (Table 1). In this trial we observe 20 cases of COVID-19 in the BNT162b2 arm and 80 cases in the experimental vaccine arm. The rate ratio is very high (4.00, 95% CI 2.42–6.90) – at face value, this suggests that the experimental vaccine is markedly inferior to BNT162b2, arguing strongly against its licensure.

We now consider a different perspective. The 95% efficacy of BNT162b2 indicates that there would have been 400 ($=20/(1-0.95)$) infections in each arm if *none* of the trial participants had been vaccinated. As 80 COVID-19 cases occurred in the experimental arm, this implies that the experimental vaccine averted 320 cases, and that its efficacy was 80% ($=320/400$). An efficacy of 80% comfortably exceeds the target of 50% set by the World Health Organization and the US Food and Drug Administration for the licensure for COVID-19 vaccines [8, 9]. Coincidentally, 80% is the approximate efficacy of the ChAdOx1 viral-vector vaccine (provided the prime-boost interval is ≥ 12 weeks), which is considerably cheaper than mRNA vaccines and has less stringent cold chain requirements [10]. Thus, if ChAdOx1 had been assessed against BNT162b2 in an active-control trial, the use of the rate ratio could have led to the unwarranted rejection of a viable vaccine option in resource-limited settings. Indeed, ChAdOx1 saved more lives worldwide in 2021 than any other COVID-19 vaccine [11]. An alternative and more meaningful metric than the rate ratio is the efficacy of the experimental vaccine compared with the control vaccine (“relative efficacy”) i.e. $80/95 = 0.842$. We return

to this metric in [Mathematical formulae and alternative approaches to estimating the averted events ratio](#) section.

Non-inferiority and effect preservation

Active-control trials are often designed and analysed within a non-inferiority framework [2, 5, 12]. A key aspect of non-inferiority trials is the non-inferiority margin, which is pre-specified in the trial protocol, although most trials fail to report a justification for the selected margin [13]. The concept of “preservation of effect” for defining the non-inferiority margin, as recommended in regulatory guidelines, is not widely applied [14]. The underlying idea is that the experimental treatment should demonstrate efficacy greater than a specified fraction of the efficacy of the control treatment. Two key pieces of information are required to use this approach: the efficacy of the control treatment as inferred from previous placebo-controlled trials, and the *fraction* of this effect to be preserved. Conventionally, this fraction has been set at 50%, although it has been argued that higher, more conservative values should be used [3, 15]. Also, for non-continuous outcomes, the *scale* for assessing effect preservation needs to be selected. The standard approach for time-to-event outcomes is to use a log-incidence scale, driven by statistical modelling considerations [3, 4]. However, this scale is arbitrary and inference based upon it lacks clear interpretability, as discussed in the next section.

HIV pre-exposure prophylaxis trial

HIV pre-exposure prophylaxis is the use of antiretroviral drugs to prevent the acquisition of HIV infection rather than to prevent disease in those already infected with the virus. The first regimen to be approved was the two-drug combination TDF-FTC, which confers very high protection ($> 95\%$) if taken as indicated [16]. DISCOVER was an active-control non-inferiority trial that assessed another two-drug combination, TAF-FTC, against TDF-FTC [17]. Analysis was performed on a log-incidence scale, with the aim of preserving 50% of the effect of TDF-FTC; non-inferiority would be concluded if the upper 97.5% confidence limit for the rate ratio (TAF-FTC versus TDF-FTC) was less than 1.62 [17].

The trial was expected to generate approximately 72 endpoints per arm, but the observed HIV incidence was much lower, with only 11 and 6 incident HIV infections in the TDF-FTC and TAF-FTC arms, respectively (Table 2) [17]. The observed upper 97.5% confidence limit for the rate ratio was 1.48, slightly lower than non-inferiority margin of 1.62, allowing non-inferiority to be concluded. However, this conclusion is very unstable – for

Table 1 Results from a hypothetical active-control COVID-19 vaccine trial

	BNT162b2	Experimental vaccine
Person-years follow-up per arm	10,000	10,000
Efficacy	95%	80%
Observed COVID-19 cases	20	80
COVID-19 cases if subjects had not been vaccinated	400	400
Averted COVID-19 cases	380	320

Values in bold are known or directly observed, other values are inferred

Table 2 DISCOVER trial: primary outcome analysis

Original analysis on rate ratio scale					
Group	No. subjects	PYFU	Incident HIV infections	Incidence rate (per 100 PYFU)	Rate ratio (95% CI)
TDF/FTC	2693	4386	11	0.251	REF
TAF/FTC	2694	4370	6	0.137	0.55 (0.20, 1.48)
Re-analysis using an averted infections framework					
Group	PYFU	Observed infections	Predicted infections ^a	Averted infections	Averted infections ratio (95% CI)
TDF/FTC	4386	11	90.4	79.4	REF
TAF/FTC	4370	6	90.0	84.0	1.06 (0.96–1.17)

^a Applying counterfactual placebo incidence rate of 2.06 per 100 PYFU (lower 2.5% credibility limit)

example, adding a *single* additional event to the TAF-FTC arm (from 6 to 7) increases the upper 97.5% confidence limit to 1.65 i.e. above the non-inferiority margin. This inferential instability vis-a-vis the observed data is known as a high “fragility index” [18], although the relevance of this concept has been challenged by other researchers [19].

Glidden and colleagues re-analysed the DISCOVER data using an averted events (infections) framework, based on the counterfactual placebo HIV incidence rate [20]. Using a Bayesian approach that synthesised data on baseline HIV infections and incident sexual transmitted infections, the posterior mean for the counterfactual placebo incidence was estimated to be 4.51 (95% credible interval [CrI] 2.06-7.36) per 100 PYFU. Applying (pessimistically) the lower bound estimate of 2.06 per 100 PYFU gives approximately 90 predicted events in each group, had they received placebo (Table 2). If this value is accurate, both regimens averted substantial numbers of

infections: an estimated 79.4 in the TDF-FTC group and 84.0 in the TAF-FTC group (Fig. 1).

An alternative metric to the ratio of the observed events is the ratio of *averted* events (averted events ratio [AER]) between the groups (84.0/79.4=1.06, 95% CrI 0.96–1.17). In other words, TAF-FTC prevented an estimated 6% more infections than TDF-FTC, with a plausible range from 4% fewer infections to 17% more infections. With the AER, conclusions about non-inferiority are made on the basis of the lower confidence limit; thus, we can conclude that TAF-FTC preserved at least 96% of the effect of TDF-FTC, emphatically demonstrating non-inferiority. In this framework, adding one extra event to the TAF-FTC arm (decreasing the predicted number of averted infections from 84.0 to 83.0) has no material effect on the averted infections ratio (1.05, 95% CrI 0.95–1.16). When both treatments are highly effective, as here, the AER is much more stable than the rate ratio.

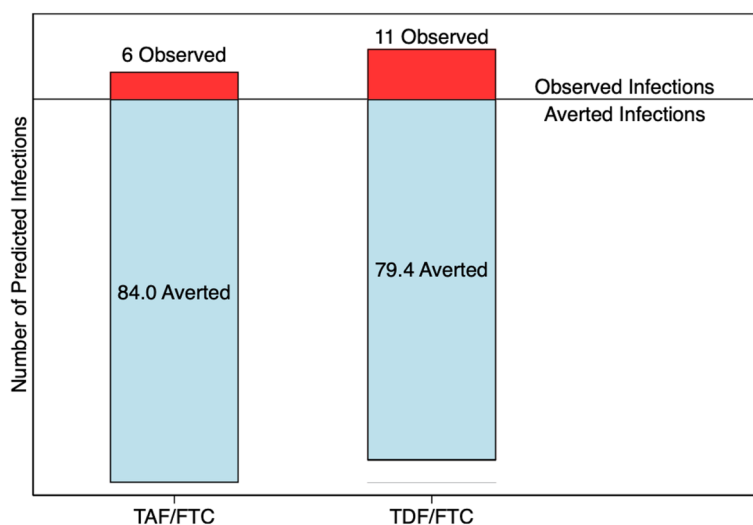


Fig. 1 Averted and observed infections in the DISCOVER trial. Assuming counterfactual placebo incidence = 2.06 per 100 PYFU

Alternative estimation approach

Formalising the argument in the previous section, let λ_E and λ_C denote the observed incidence rates in the experimental and control arms, and let λ_P denote the counterfactual placebo incidence. The AER is calculated by

$$\Psi = \frac{\lambda_P - \lambda_E}{\lambda_P - \lambda_C} \tag{1}$$

Now let $\theta_{CP} = (\lambda_P - \lambda_C)/\lambda_P = 1 - \lambda_C/\lambda_P$ denote the efficacy of the control treatment (relative to no treatment) and let $\beta_{EC} = \lambda_E/\lambda_C$ denote the rate ratio (or hazard ratio) observed in the active-control trial. Dividing all terms in Eq. (1) by λ_P ,

$$\begin{aligned} \Psi &= \frac{1 - \lambda_E/\lambda_P}{1 - \lambda_C/\lambda_P} \\ &= \frac{1 - \beta_{EC}\lambda_C/\lambda_P}{\theta_{CP}} \\ &= \frac{1 - \beta_{EC}(1 - \theta_{CP})}{\theta_{CP}} \end{aligned} \tag{2}$$

This formulation reveals that the AER can alternatively be estimated via the counterfactual effectiveness of the active-control treatment, rather than the counterfactual placebo incidence [21]. The choice of which formulation to use depends on the disease context. Because HIV incidence changes relatively gradually in a given population, estimation of the counterfactual placebo incidence may be feasible in HIV prevention research. In contrast, the incidence of SARS-CoV-2 (and thus COVID-19) has fluctuated in a largely unpredictable manner, implying the likely need to perform estimation via the counterfactual vaccine efficacy.

Specification of either counterfactual parameter is challenging and requires subject-matter knowledge [6]. A sensitivity analyses of how point estimates and confidence intervals for the AER vary over the range of plausible values is highly informative. Figure 2 depicts such an analysis for the DISCOVER trial, and reveals several important points. First, the lower the value of the counterfactual parameter, the slightly higher the point estimate of the AER. (Conversely, when the experimental treatment is less effective than the control treatment, the AER is less than one.) Second, confidence intervals are considerably narrower at higher value of the counterfactual parameter; thus, for conservative inference low values should be assumed. Third, the confidence intervals are considerably narrower when imputing the counterfactual placebo incidence rather than the counterfactual efficacy of TDF-FTC, favouring the use of the former approach if feasible. Finally, we note that in addition to exploring how the AER varies over a range of values of the counterfactual parameter, we may wish

to integrate over this parameter to obtain the unconditional distribution of the AER. Bayesian inference provides a natural framework for this problem [20, 22].

Relative efficacy

Note that Eq. (2) is simply $\Psi = \theta_{EP}/\theta_{CP}$ i.e. the efficacy of the experimental treatment compared with the efficacy of the control treatment. This expression may be particularly appealing to vaccinologists, for whom vaccine efficacy is a natural metric. Thus the relative vaccine efficacy estimate of 0.842 in the "Hypothetical COVID-19 vaccine trial" section can be interpreted as the experimental COVID-19 vaccine averting 84.2% of the COVID-19 cases that would otherwise be averted by BNT162b2.

The term relative efficacy or relative effectiveness has been widely used in influenza vaccine research – a recent review paper identified 63 articles that reported this term, either in the comparison of different vaccines, doses of the same vaccine, or vaccination schedules [23]. However, in this context relative vaccine efficacy/effectiveness has been defined as:

$$(\lambda_C - \lambda_E)/\lambda_C = 1 - \beta_{EC} \tag{3}$$

This is interpreted as the proportionate reduction in influenza cases if using the experimental vaccine rather than the control vaccine (without relation to a hypothetical placebo group). While this is a meaningful metric, it is fundamentally different to the way that we have defined relative effectiveness.

A recent modelling paper acknowledged limitations in relative effectiveness, as defined in Eq. (3) [24]. For a fixed value of relative effectiveness, the number of untoward events (hospitalisations) averted was shown to be a function of the absolute efficacy of the control vaccine. These values were computed from the *difference*, rather than the ratio, between the efficacies of the experimental (enhanced) and control vaccines, an equally valid approach. In line with our conclusions, the authors stated: “We showed that relative vaccine efficacy is difficult to interpret when reported without contextual information and on its own is a potentially insufficient metric to measure and compare the benefits of enhanced influenza vaccines” [24].

Relative efficacy is also referred to in FDA guidance on COVID-19 vaccines: “For non-inferiority comparison to a COVID-19 vaccine already proven to be effective, the statistical success criterion should be that the lower bound of the confidence interval around the primary *relative efficacy* point estimate is >-10%” [9]. The guidance document does not explicitly define relative efficacy, but a recent paper on the design of non-inferiority trials for COVID-19 vaccines assumed the definition in Eq. (3)

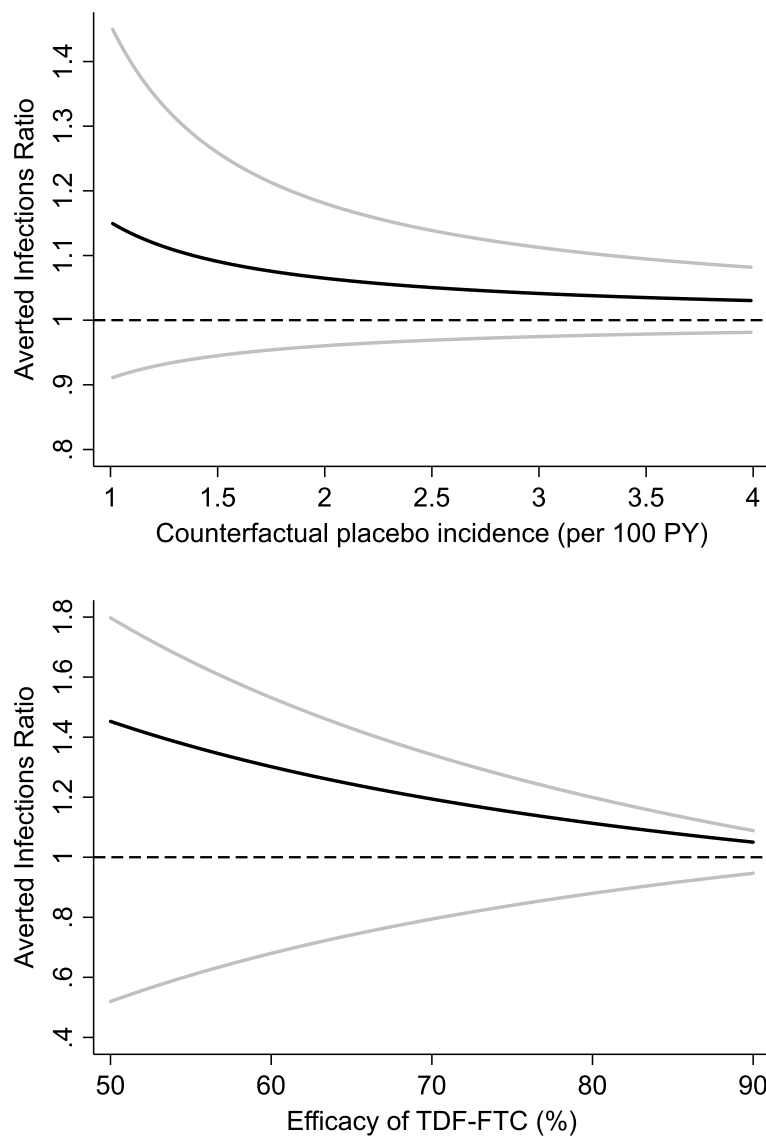


Fig. 2 Sensitivity analyses of DISCOVER data. Top panel: Varying counterfactual placebo incidence. Lower panel: Varying counterfactual effectiveness of TDF-FTC. Black line, point estimate. Grey lines, 2.5% and 97.5% confidence limits. Note different scales on graphs. Methods for deriving confidence limits are given in references 21 and 22

[25]. Clear definition of the term is important to avoid ambiguity.

Conclusions

We have shown that the standard estimand for analysing active-control trials with time-to-event outcomes, the rate ratio based on observed events, can result in misleading clinical conclusions. Valid interpretation requires consideration of the number of averted events as well as observed events, and the AER provides an intuitive and clinically meaningful measure of the relative effectiveness of the experimental treatment. The AER framework is

particularly advantageous when the control treatment is highly effective i.e. the number of averted events greatly exceeds the number of observed events.

In the field of HIV prevention, the need to estimate the counterfactual placebo incidence is increasingly accepted and various approaches have been proposed [6, 26–28]. However, most trials continue to use the rate ratio as the primary estimand, probably due to inherent conservatism in regulatory guidance. Finally, we wish to acknowledge that our work is a development of the work of several authors going back 20 years, whose ideas warrant greater attention [3, 15, 29].

Acknowledgements

None.

Authors' contributions

The original ideas were developed jointly by DTD and DVG. DTD wrote the first draft of the paper and edited subsequent revisions. All authors commented on the paper.

Funding

DTD and SMC were supported by the UK Medical Research Council grant MC_UU_00004/03 and MC_UU_00004/07. DVG was supported by US National Institutes of Health grant R01AI143357.

Availability of data and materials

Not applicable.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

DVG has received fees from Gilead Sciences. All other authors declare no competing interests.

Received: 12 December 2022 Accepted: 9 June 2023

Published online: 26 June 2023

References

1. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med.* 2000;133(6):455–63.
2. Fleming TR. Current issues in non-inferiority trials. *Stat Med.* 2008;27(3):317–32.
3. Mielke M, Munk A, Schacht A. The assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints. *Stat Med.* 2008;27(25):5093–110.
4. Donnell D, Hughes JP, Wang L, Chen YQ, Fleming TR. Study design considerations for evaluating efficacy of systemic preexposure prophylaxis interventions. *J Acquir Immune Defic Syndr.* 2013;63:S130–43.
5. Snapinn S, Jiang Q. Preservation of effect and the regulatory approval of new treatments on the basis of non-inferiority trials. *Stat Med.* 2008;27(3):382–91.
6. Dunn DT, Glidden DV, Stirrup OT, McCormack S. The averted infections ratio: a novel measure of effectiveness of experimental HIV pre-exposure prophylaxis agents. *Lancet HIV.* 2018;5(6):e329–34.
7. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N Engl J Med.* 2020;383(27):2603–15.
8. World Health Organization. R&D Blueprint. An international randomised trial of candidate vaccines against COVID-19. 2020. www.who.int/publications/i/item/an-international-randomised-trial-of-candidate-vaccines-against-covid-19.
9. US Food and Drug Administration. Development and licensure of vaccines to prevent COVID-19. 2020. www.fda.gov/regulatory-information/search-fda-guidance-documents/development-and-licensure-vaccines-prevent-covid-19.
10. Voysey M, Costa Clemens SA, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *Lancet.* 2021;397(10277):881–91.
11. Which COVID-19 vaccine saved the most lives in 2021? *The Economist.* 13 July 2022. www.economist.com/graphic-detail/2022/07/13/which-covid-19-vaccine-saved-the-most-lives-in-2021. Accessed 16 Apr 2023.
12. Phillips PPJ, Glidden DV. Noninferiority trials. In: Piantadosi S, Meinert CL, editors. *Principles and practice of clinical trials.* Cham: Springer International Publishing; 2020. p. 1–28.
13. Rehal S, Morris TP, Fielding K, Carpenter JR, Phillips PP. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open.* 2016;6(10):e012594.
14. Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness. Guidance for industry. 2016. www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials.
15. Pigeot I, Schafer J, Rohmel J, Hauschke D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Stat Med.* 2003;22(6):883–99.
16. Anderson PL, Glidden DV, Liu A, Buchbinder S, Lama JR, Guanira JV, et al. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. *Sci Transl Med.* 2012;4(151):151ra25.
17. Mayer KH, Molina JM, Thompson MA, Anderson PL, Mounzer KC, De Wet JJ, et al. Emtricitabine and tenofovir alafenamide vs emtricitabine and tenofovir disoproxil fumarate for HIV pre-exposure prophylaxis (DISCOVER): primary results from a randomised, double-blind, multicentre, active-controlled, phase 3, non-inferiority trial. *Lancet.* 2020;396(10246):239–54.
18. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol.* 2014;67(6):622–8.
19. Potter GE. Dismantling the Fragility Index: a demonstration of statistical reasoning. *Stat Med.* 2020;39(26):3720–31.
20. Glidden DV, Stirrup OT, Dunn DT. A Bayesian averted infection framework for PrEP trials with low numbers of HIV infections: application to the results of the DISCOVER trial. *Lancet HIV.* 2020;7(11):e791–6.
21. Dunn DT, Glidden DV. The connection between the averted infections ratio and the rate ratio in active-control trials of pre-exposure prophylaxis agents. *Stat Commun Infect Dis.* 2019;11(1):20190006.
22. Dunn DT, Stirrup OT, Glidden DV. Confidence limits for the averted infections ratio estimated via the counterfactual placebo incidence rate. *Stat Commun Infect Dis.* 2021;13(1):20210002.
23. McMenamin ME, Bond HS, Sullivan SG, Cowling BJ. Estimation of relative vaccine effectiveness in influenza: a systematic review of methodology. *Epidemiology.* 2022;33(3):334–45.
24. Lewis NM, Chung JR, Uyeki TM, Grohskopf L, Ferdinands JM, Patel MM. Interpretation of relative efficacy and effectiveness for influenza vaccines. *Clin Infect Dis.* 2022;75(1):170–5.
25. Mehrotra DV, Janes HE, Fleming TR, Annunziato PW, Neuzil KM, Carpp LN, et al. Clinical endpoints for evaluating efficacy in COVID-19 vaccine trials. *Ann Intern Med.* 2021;174(2):221–8.
26. Mullick C, Murray J. Correlations between HIV infection and rectal gonorrhoea incidence in men who have sex with men: Implications for future HIV pre-exposure prophylaxis trials. *J Infect Dis.* 2020;221(2):214–7.
27. Gao F, Glidden DV, Hughes JP, Donnell DJ. Sample size calculation for active-arm trial with counterfactual incidence based on recency assay. *Stat Commun Infect Dis.* 2021;13(1):20200009.
28. Glidden DV, Das M, Dunn DT, Ebrahimi R, Zhao Y, Stirrup OT, et al. Using the adherence-efficacy relationship of emtricitabine and tenofovir disoproxil fumarate to calculate background hiv incidence: a secondary analysis of a randomized, controlled trial. *J Int AIDS Soc.* 2021;24(5):e25744.
29. Durrleman S, Chaikin P. The use of putative placebo in active control trials: two applications in a regulatory setting. *Stat Med.* 2003;22(6):941–52.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.