

Title: A Bayesian averted infection framework for PrEP trials with low number of HIV infections: application to the results of the DISCOVER trial

Prof. David V. Glidden*, PhD

Dr. Oliver T. Stirrup, PhD

Prof. David T. Dunn*, PhD

Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA [DVG]

Institute for Global Health, University College London, London, UK [OTS]

MRC Clinical Trials Unit at University College London, London, UK [DTD]

* Full professor

Corresponding Author: David V. Glidden, Ph.D., Department of Epidemiology and Biostatistics; University of California, San Francisco, 550 16th St., San Francisco, CA 94158
Phone: +1 (415) 476-1111, Ext. 13603; Fax: +1 (415) 514-8150

email: david.glidden@ucsf.edu

Short title: Averted infections in DISCOVER

Summary:

Trials of candidate agents for HIV pre-exposure prophylaxis (PrEP) may randomise between a new agent (nPrEP) and oral co-formulated emtricitabine plus tenofovir disoproxil fumarate (F/TDF). This design presents unique challenges in design and interpretation. First with two active arms, HIV incidence may be low. Second, F/TDF effectiveness varies across populations; thus, similar HIV incidence between arms could be consistent with a wide range of effectiveness for the nPrEP. We propose a two-part approach to trial results. First, we use Bayesian methods to incorporate assumptions about the background trial HIV incidence in the absence of PrEP, possibly augmented by external data. Based on this, we estimate and compare the number of averted (or prevented) HIV infections in each of the two trial arms, calculating the averted infections ratio (AIR). We apply these methods to a recently completed trial of tenofovir alafenamide with emtricitabine (F/TAF) for PrEP. Our framework demonstrates that leveraging external information to estimate averted infections and the AIR enhances the efficiency and interpretation of active-controlled PrEP trials.

1 The DISCOVER Trial Design

There is an urgent need for innovative trial approaches which can ethically and rigorously assess new HIV prevention methods.¹⁻⁴ Randomized active controlled trials, with an oral emtricitabine plus tenofovir disoproxil fumarate (F/TDF) control group, have been the preferred design for assessing new pre-exposure prophylaxis (nPrEP) agents. To date, active controlled nPrEP trials have been large and expensive, with power calculations requiring observation of at least 100 incident HIV infections.⁵⁻⁷ Observing this number of infections is a particular challenge in active controlled trials because at least one, and possibly both, arms may be receiving effective prevention. There has been robust discussion about designing cost-effective trials, identifying rigorous sources of evidence, and modifying regulatory standards for nPrEP. These dilemmas are well illustrated by the design, results, and interpretation of the DISCOVER study.

The DISCOVER study was a randomized double-blind double-dummy active-controlled trial comparing co-formulated emtricitabine and tenofovir alafenamide (F/TAF; Descovy®) vs. F/TDF for PrEP in men (MSM) and transgender women (TW) who have sex with men.⁵ The study met its pre-specified margin for non-inferiority (NI) of F/TAF compared to F/TDF.⁸ Based on DISCOVER, the US Food and Drug Administration approved F/TAF for PrEP in MSM and TW.⁹

In this paper, we present a reanalysis of the DISCOVER trial data using an alternative framework. Our approach compares the numbers of prevented infections on each arm which are calculated based on explicit assumptions about the counterfactual background HIV incidence,¹⁰ and allows incorporation of expert opinion and/or external data through Bayesian inference. Our analyses suggest that F/TDF and F/TAF have more similar effectiveness than might be apparent.

2 DISCOVER Results

Table 1 summarises the results of DISCOVER. There were 22 total HIV infections: 5 suspected baseline infections and 17 post-baseline infections. By PrEP group, there were 11 F/TDF vs. 6 F/TAF HIV post-baseline infections. The F/TAF:F/TDF relative rate (RR) was 0.55 with a 95% confidence interval (CI) of 0.22–1.48 (Table 1). The CI upper bound was within a pre-specified NI margin⁵ of <1.62 and met the criterion for NI, providing affirmative evidence that at least 50% of the control (F/TDF) effectiveness was preserved by the investigational treatment (F/TAF). The NI margin was derived based on the results of completed placebo-controlled trials of F/TDF and a set of assumptions¹¹ (further details in Supplementary Appendix, page S1). However, a comparison of the RR to a fixed margin has important

limitations; it does not address the strength of evidence for non-inferiority or estimate the effectiveness of F/TAF.

We examined the robustness of the NI conclusion (using the 1.62 margin) by hypothetical addition or subtraction of HIV infections from the F/TAF arm (Table 2). The trial would have failed to demonstrate NI with the addition of a single additional HIV infection in the F/TAF arm. Conversely, subtracting three infections from the F/TAF arm would have led to the conclusion of superiority of F/TAF. The qualitatively different conclusions supported by small changes to number of infections, present a challenge for interpretation of the strength of DISCOVER's evidence.

The FDA briefing document alludes to the low number of HIV infections: “... the similarity between F/TAF and F/TDF can mean either that both drugs were effective or neither drug was effective because the population was not at substantial risk.”¹² This distinction is crucial: if the former is true then the conclusion of NI is justified, whereas if the latter is true then the trial had no possibility of generating information about the relative effectiveness of the two drugs. The briefing document further outlined why the former hypothesis was more plausible, citing the high effectiveness of F/TDF in previous placebo-controlled trials of MSM and an apparent high underlying risk of HIV infection in the trial population based on self-reported condomless anal sex and high STI rates.

3 Averted Infections and the Averted Infections Ratio

The conventional approach¹¹ to NI trials is based on comparing the numbers of diagnosed HIV infections between study arms. We applied the framework of averted (or prevented) infections to the DISCOVER results, which shifts focus to estimation and comparison of the unobserved numbers of HIV infections that were prevented by the study drug in each arm.¹⁰ We contend evidence of effectiveness of PrEP product accumulates when it prevents infections that would otherwise happen. If an overwhelming effective preventative is used in a high risk population, then there will be few or no events. However, proof of prevention is abundant, but unseen, since there are a large number of averted infections. A 100% effective preventative is delivered in a population without risk, there will be zero events. However there are also no infection averted and no evidence of effectiveness is provided. The averted infection scale is particularly useful for trials with few HIV infections because it makes the distinction between the two scenarios.

Estimation and comparison of the number of infections averted by F/TDF and by F/TAF requires specification of the HIV rate that would have occurred in the trial in the absence of PrEP; we termed this the “background HIV incidence rate”. Initially, we considered the rate of 1.44 per 100 PY, the value assumed in the design of DISCOVER for the F/TDF arm (hence, this estimate was conservatively low). With this rate, in the absence of PrEP, approximately

126 HIV infections would have been expected in the trial population, 63 in each arm. Far fewer HIV infections were observed in each arm of DISCOVER: 52 (=63–11) fewer on F/TDF and 57 (=63–6) fewer on F/TAF (Fig 1). Under this background rate, there were many more averted infections (n=109) than observed infections (n=17) in DISCOVER.

Our preferred metric of effect preservation is the averted infections ratio¹⁰ (AIR) between the arms, $57/52 = 1.10$, 95% CI: 0.94–1.27. In other words, we estimate that in DISCOVER F/TAF prevented from 27% more to 6% fewer infections than did F/TDF. Hence, F/TAF preserved at least 94% of the effect of F/TDF – far above 50% effect preservation. The AIR estimate of 1.10 is based on an assumed background HIV incidence rate of 1.44 per 100 PY. Fig 2 shows a graphical exploration of the effect of varying this assumption for our interpretation of the DISCOVER data, displaying the number of averted infections, the AIR and the lower limit of the 95% CI for the AIR according to assumed background HIV incidence. The horizontal line at 0.5 demarcates the region for which F/TAF averted at least 50% of infections — a measure of 50% effect preservation and thus evidence of NI. If we assume that the background HIV rate is at least 0.5 per 100 PY, a very low rate for a study in any reasonable setting, then we have confidence in NI. Confidence grows rapidly with higher assumed background incidence.

The AIR can also be derived in terms of the assumed effectiveness for F/TDF (relative to background incidence).¹⁰ The DISCOVER protocol⁵ (see the Supplementary Appendix, p. S1) derived a working estimate of F/TDF effectiveness of 62% based on previous trials. Under this assumption, the estimated number of averted infections would be approximately 18 (F/TDF) and 23 (F/TAF) for an AIR of 1.28 with 95% CI: 0.71 to 1.49 — still far above 50% effect preservation.

In Table 2, we demonstrate that small changes to the data would have had a large effect on whether NI, or even superiority, was supported by conventional methods of analysis. The results using the AIR were considerably more resilient. The addition of one HIV infection to the F/TAF arm caused the RR confidence limit to fall outside the NI boundary. However, such a change would only move the AIR from 1.10 (95% CI: 0.94 to 1.27) to 1.08 (95% CI: 0.92 to 1.26). Subtracting three infections from the F/TAF arm implied the superiority of F/TAF (as well as NI) using the RR. Notably, the AIR could also conclude superiority of F/TAF even though the point estimate (1.15) and the 95% CI (1.02 to 1.31) would not be considerably altered. This reflects the fact that when the number of observed infections is small, yet the plausible number of averted infections is large, the between-arm ratios of the latter will be much more stable than the former.

4 Incorporation of Counterfactual Placebo Evidence – A Bayesian Approach

External evidence suggests that the background HIV incidence in DISCOVER is likely to have been well above 1 per 100 PY. Other MSM/TW PrEP trials with similar inclusion criteria

reported background HIV incidence ranging from 3.9 to 9.0 per hundred PY.¹³⁻¹⁵ DISCOVER participants had high incidence of sexually transmitted infections (STIs) -- the rectal gonorrhoea rate was 21.0 per 100 person⁸, suggesting sexual practices that facilitate HIV transmission. Within a month of enrolment, five suspected baseline infections were diagnosed in DISCOVER. Assuming an eclipse period of 2 weeks among the participants initiating PrEP at enrolment (n=4498), this would correspond to 173 PY of follow-up, implying a pre-PrEP incidence rate of 2.9 per 100 PY (95% CI: 0.9 to 6.7). Note, 17% of DISCOVER participants enrolled into the study directly from daily TDF/FTC; they do not count as PrEP “initiators” and are therefore not included in the background incidence calculation.

We used a Bayesian approach to generate inferences regarding the hypothetical background HIV incidence rate among the DISCOVER participants had they not been receiving PrEP, based on both the observed background HIV incidence and the incidence of rectal gonorrhoea during the study period. For the latter, we used data from a systematic review by Mullick and Murray¹⁶ that evaluated the correlation between HIV and rectal gonorrhoea incidence rates among MSM not using PrEP. The linear regression formula from this study yields an estimated background HIV incidence of 6.6 per 100 per year for the DISCOVER participants. We built on this approach, directly using the raw data from the systematic review in our analysis. Additionally, the HIV incidence inferred from the baseline prevalent infections in DISCOVER was included in the model, assuming this was consistent with the background incidence during the trial. We applied a Bayesian analysis to estimating the AIR using a weakly informative prior (to restrict values to a plausible range) for background HIV incidence, combined with the trial results augmented with data on rectal gonorrhoea and baseline infection. We used Stan¹⁷ software to estimate the posterior mean and associated 95% credible interval (CrI) for the AIR and other parameters of interest.¹⁸ The complete details of our methods have been provided in the Supplementary Appendix (pp. S2-S3).

Fig 3 shows the posterior density for the background HIV incidence ϵ in the DISCOVER trial based on this analysis, posterior mean of 4.5 (95% CrI 2.0–7.3) per 100 person years. Fig 4 displays the posterior density for the associated AIR with posterior mean 1.03 (95% CrI 0.98–1.11). The posterior is a flexible tool for describing uncertainty, for example allowing us to calculate that the probability that the AIR lies between 0.95 to 1.05 (effectiveness preservation within +/- 5%) equals 80%. While the posterior probability of superiority (AIR > 1) is 88%, the probability that the effect of F/TAF was greater than 10% higher (AIR > 1.1) was 3%. Hence, the analysis strongly suggested that the posterior effectiveness of F/TDF (93%, 95% CrI 85 to 97) and F/TAF (96%, 95% CrI 91 to 99) were very similar and that any potential superiority of F/TAF is modest.

5 Implications for DISCOVER and beyond

Based on HIV incidence, the evidence for F/TAF's non-inferiority from DISCOVER might appear to be weak; the trial observed just 15% (22/144) of its planned endpoints. While the trial met the pre-specified margins for non-inferiority, conclusions are sensitive to relatively small changes to the number of infections in each arm. However, our Bayesian analysis estimates an 80% posterior probability that the ratio of averted infections lies between 0.95 and 1.05 (effectiveness preservation within +/- 5%); hence, we are confident not just that F/TAF is non-inferior to F/TDF, but that the effectiveness of F/TDF and F/TAF (relative to the background HIV incidence) are highly similar.

Nevertheless, evaluation of the clinical and public health utility of F/TAF must incorporate many issues beyond effectiveness, including safety, cost, and access. Because DISCOVER only enrolled MSM/TW, clinical data on F/TAF effectiveness for PrEP is lacking in many key populations, including cisgender women, transgender men, heterosexuals men, and people who inject drugs.¹⁹ Although F/TDF is associated with small decreases in renal and bone health, F/TAF is associated with small increases in weight and cholesterol. However, these small changes may not be clinically significant.²⁰⁻²² From a societal perspective, generic F/TDF will be available in 2020, bringing discounted prices and making it unlikely that F/TAF will be cost-effective.²²

Beyond its specific results, the DISCOVER trial is highly instructive for future active controlled PrEP trials. A nPrEP agent is most effectively assessed in a trial enrolling people at substantial risk of HIV infection. If there are high levels of adherence to efficacious agents, the number of observed infections will be low, but the number of averted infections will be high. This fact underpins the rationale for our proposed framework for analysis which places the emphasis on the number of averted infections. The inference on averted infections requires explicit assumptions about the background HIV rate and/or control arm effectiveness.

Our method contrasts with the fixed margin NI method¹¹ which was used to power the DISCOVER and HPTN083 trials. This approach develops a pre-specified margin which is compared to the confidence interval for the RR. The fixed margin is derived from two components: assumptions about control group effectiveness and a minimum standard for effect preservation. The conflation of assumed control effectiveness and the set standard for NI in the fixed margin method makes it difficult to compare evidence for effectiveness and NI across studies. Note, the DISCOVER results would not meet the NI margin of 1.23 set for the HPTN083 study (*cf.* 1.62 for DISCOVER). The difference in margins in this instance does not arise from differences in standards for effect preservation, but is due to HPTN083 assuming lower F/TDF adherence than DISCOVER (reflecting the former's participant recruitment strategy). The development of these NI margins is contrasted in the Supplementary Appendix

(Tables S1 and S2, pp S4-S5). Another problem with the fixed margin approach is that it is entirely pre-specified. If data arise during the trial (e.g., higher than assumed adherence) which contradict the assumed control arm effectiveness, the NI margin can be excessively conservative.²³ There is an incentive to default to conservative assumptions in the design phase; this conservatism costs power and requires larger sample sizes.²⁴

Our strategy by contrast combines inference and decision on NI based on a pre-specified standard (e.g. AIR ≥ 0.50) with assumptions on the background HIV incidence and/or the control arm effectiveness. The uncertainty in these parameters is handled transparently and formally through sensitivity analysis (e.g., Fig 2) Bayesian analysis. The latter approach can incorporate data which inform these key parameters. Sources can include HIV infections during screening, on-trial STI rates and/or drug levels. One might choose a sceptical prior distribution on the background HIV incidence as a way to require a higher standard of evidence. An advantage of this approach is dealing with key assumptions in a way which is flexible and transparent while enforcing a pre-specified standard for effect preservation. Our hope is that the incorporation of external information can allow for smaller trial sizes than those determined by the fixed margin approach. This is an area of active investigation.

This strategy, combining Bayesian inference with the AIR, will be particularly advantageous for active controlled trials with low event rates, which can be strongly informative if there is confidence that they have been well-conducted in cohorts at substantial risk for HIV in the absence of PrEP.

Contributions

This paper arose through discussions between DTD and DVG. OTS did the Bayesian analysis. DVG wrote the first draft of the paper and edited subsequent revisions, and DTD, and OTS contributed substantially to these.

Declaration of interests

DTD reports personal fees from ViiV Healthcare and Gilead Sciences, outside of the submitted work. DVG has accepted fees from Gilead Sciences and Merck.

Acknowledgments

Authors were supported by US National Institutes of Health grant R01AI143357 (to DVG). DTD was supported by UK Medical Research Council grant MR_UU_12023/23. We thank Sheena McCormack and Julia Marcus for their insights and critical reading of the manuscript.

Citations

1. Sugarman J, Celum CL, Donnell D, Mayer KH. Ethical considerations for new HIV prevention trials. *Lancet HIV*. 2019; 6:e489-91.
2. Janes H, Donnell D, Gilbert PB, et al. Taking stock of the present and looking ahead: envisioning challenges in the design of future HIV prevention efficacy trials. *Lancet HIV*. 2019; 6:e475-e482
3. Donnell D. Current and future challenges in trial design for pre-exposure prophylaxis in HIV prevention. *Stat Comm Inf Dis*. 2019;11(1).
4. Glidden DV, Mehrotra ML, Dunn DT, Geng EH. Mosaic effectiveness: measuring the impact of novel PrEP methods. *Lancet HIV*. 2019; 6:e800-6.
5. Safety and efficacy of emtricitabine and tenofovir alafenamide (F/TAF) Fixed-dose combination once daily for pre-exposure prophylaxis in men and transgender women who have sex with men and are at risk of HIV-1 infection (DISCOVER). <https://clinicaltrials.gov/ct2/show/NCT02842086> Retrieved on 30 April 2020.
6. Safety and efficacy study of injectable cabotegravir compared to daily oral tenofovir disoproxil fumarate/Emtricitabine (TDF/FTC), for pre-exposure prophylaxis in HIV-uninfected cisgender men and transgender women who have sex with men <https://clinicaltrials.gov/ct2/show/NCT02720094> Retrieved on 30 April 2020.
7. Evaluating the safety and efficacy of long-acting injectable cabotegravir compared to daily oral TDF/FTC for pre-exposure prophylaxis in HIV-uninfected women. <https://clinicaltrials.gov/ct2/show/NCT03164564> Retrieved on 30 April 2020.
8. Hare CB, Coll J, Ruane P, et al. The phase 3 DISCOVER study: Daily F/TAF or F/TDF for HIV pre-exposure prophylaxis. In Congress of Retroviruses and Opportunistic Infections, Seattle, Feb 2019 . <http://www.croiwebcasts.org/p/2019croi/104>, Retrieved on 30 April 2020.
9. Voelker R. PrEP drug Is approved for some patients but not for others. *JAMA*. 2019; 322:1644
10. 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:e329-34.

11. Fleming, T R, Odem-Davis, K, Rothmann, MD, Li Shen Y. Some essential considerations in the design and conduct of non-inferiority trials. *Clin Trials*. 2011; 8:432–39.
12. DESCovy® for HIV pre-exposure prophylaxis: antimicrobial drugs advisory committee meeting briefing document. <https://www.fda.gov/media/129609/download>. Retrieved on 30 April 2020
13. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med*. 2010; 363:2587-99.
14. McCormack S, Dunn DT, Desai M, et al. 2016. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* 2016; 387:53-60.
15. Molina JM, Capitant C, Spire B, et al. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. *N Engl J Med*. 2015; 373:2237-46.
16. 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: 214-7.
17. Carpenter B, Lee D, Brubaker MA, Riddell A, Gelman A, Goodrich B, Guo J, Hoffman M, Betancourt M, Li P. Stan: A Probabilistic Programming Language. *J Stat Softw*. 2017; 76:1–32.
18. Spiegelhalter, D.J., Abrams, K.R. and Myles, J.P., 2004. *Bayesian approaches to clinical trials and health-care evaluation* (Vol. 13). John Wiley & Sons.
19. Goldstein RH, Walensky RP. Where were the women? Gender parity in clinical trials. *N Engl J Med*. 2015; doi: 10.1056/NEJMp1913547
20. Gupta SK, Post FA, Arribas JR, et al. Renal safety of tenofovir alafenamide vs. tenofovir disoproxil fumarate: A pooled analysis of 26 clinical trials. *AIDS*. 2019; 33:1455-65.
21. Walensky RP, Horn TH, Paltiel AD. The Epi-TAF for tenofovir disoproxil fumarate? *Clin Infect Dis*. 2015; 62:915-8.
22. Hill A, Hughes SL, Gotham D, Pozniak AL. Tenofovir alafenamide versus tenofovir disoproxil fumarate: is there a true difference in efficacy and safety? *J Virus Erad*. 2018;4:72-9

23. Hanscom B, Hughes JP, Williamson BD, Donnell D. Adaptive non-inferiority margins under observable non-constancy. *Stat Methods Med Res.* 2019;28:3318-32.
24. 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 Comm Inf Dis.* 2019;11(1).

Table 1: Primary Results of the DISCOVER Study

Study Arm	No. Randomised	Person Years	Total HIV+	Post Baseline. HIV+	Post-Baseline HIV Rate (100 PY)
F/TAF	2694	4370	7	6	0.137
F/TDF	2693	4386	15	11	0.251

Relative risk (post-baseline events) 0.55, 95% CI: 0.20, 1.48

Table 2: Sensitivity analyses showing the instability of rate ratio, but stability of the averted infections ratio (AIR), by altering the number of HIV cases in the F/TAF arm of the DISCOVER trial

Scenario	HIV+ cases		Rate ratio (95% CI)	AIR ⁺ (95% CI)	AIR in Bayesian analysis (95% CrI)
	F/TA F	F/TD F			
A	6	11	0.55 (0.20–1.48)	1.10 (0.94-1.17)	1.03 (0.98–1.11)
B	7	11	0.64 (0.25–1.65)	1.08 (0.92-1.26)	1.02 (0.97–1.10)
C	3	11	0.27 (0.08–0.98)	1.15 (1.02-1.31)	1.05 (1.00–1.14)

Scenario A. Observed results, non-inferiority shown (since 1.48<1.62)

Scenario B. One additional event on F/TAF, non-inferiority not shown (since 1.65>1.62)

Scenario C. Three fewer events on F/TAF, superiority shown (since 0.98<1.00)

+ Assuming a background HIV incidence of 1.44 per 100 PY

The AIR unequivocally demonstrates non-inferiority under each of these scenarios since the lower confidence limit of the AIR indicates preservation of effect higher than 90%.

Figure Legends

Fig 1: Averted versus observed infections assuming background incidence of 1.44 per 100 person years in the DISCOVER trial.

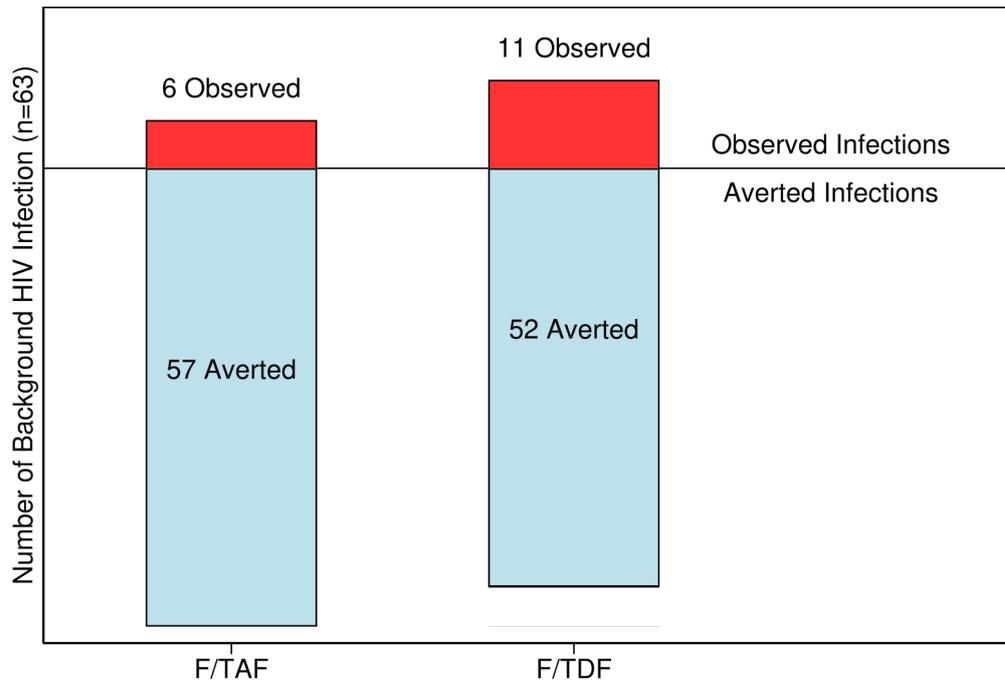


Fig 2: Averted infections ratio calculated with lower 95% confidence limit calculated from the on-study infections in the DISCOVER data by assumed background HIV incidence rate.

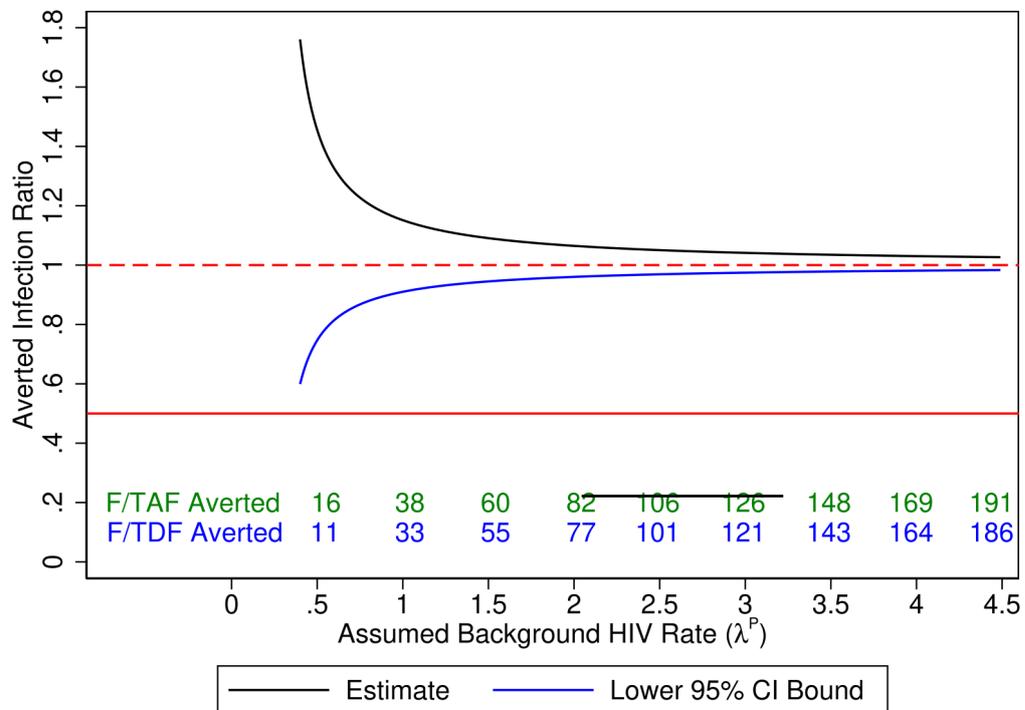


Fig 3: Estimated posterior density for the background HIV incidence in DISCOVER. The shaded area shows the 95% credibility interval, and the vertical line is the posterior mean.

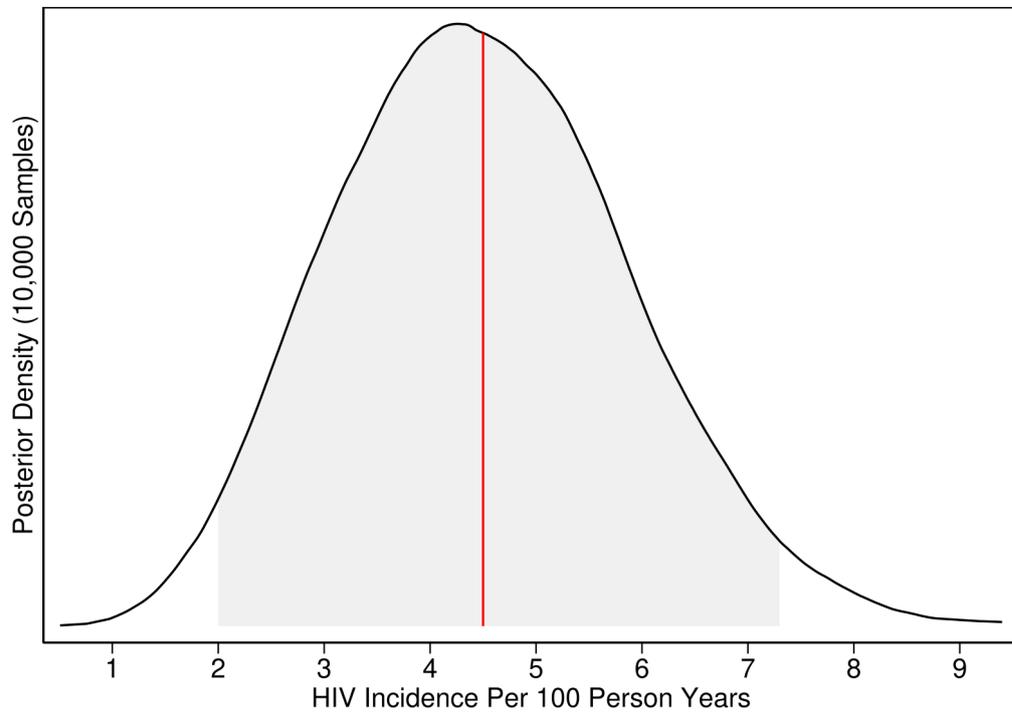
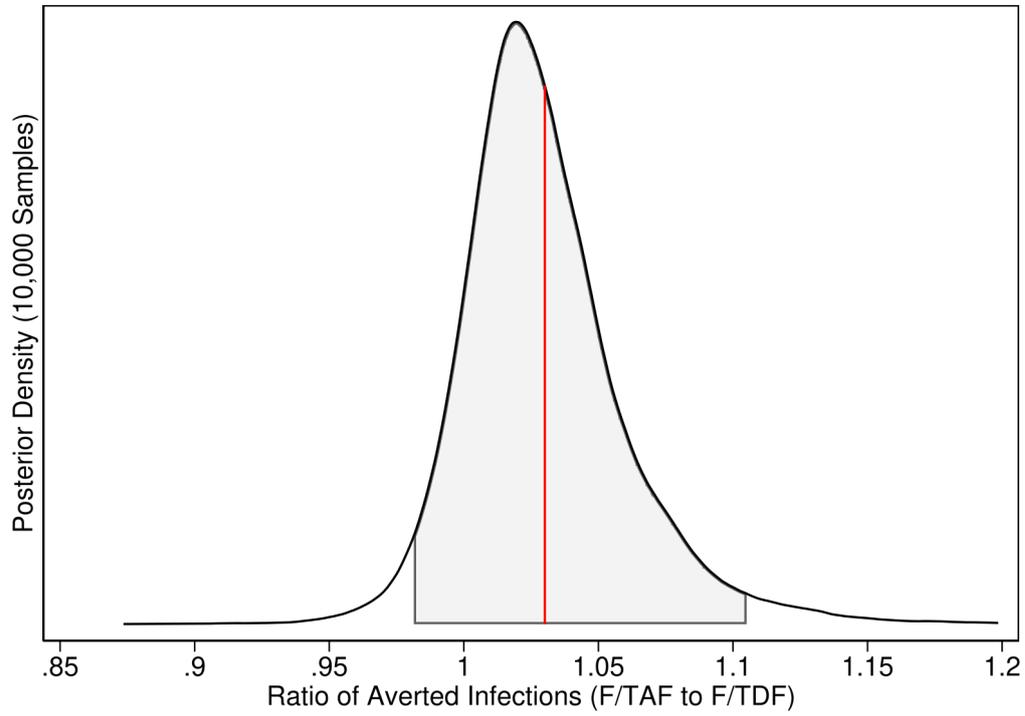


Fig 4: Estimated posterior density for the averted infection ratio in DISCOVER. The shaded area shows the 95% credibility interval, and the vertical line is the posterior mean.



Supplement

Derivation of the DISCOVER non-inferiority margin

Non-inferiority in DISCOVER was pre-specified using the paradigm known as the fixed-margin approach or “95/95”.¹¹ The approach derived an assumed effectiveness for the control arm (F/TDF) based on a meta-analysis of completed placebo controlled trials of F/TDF in MSM/TW. To correct for uncertainty in this extrapolation, it applies an ad hoc penalty by using the lower 95% CI from the meta-analysis. Using the penalty yields a margin with a type I error rate which is typically well below 0.05.¹ This approach, known as discounting, is not substantively supported or transparent to clinical colleagues, who rely on trial experts for a meaningful definition of non-inferiority. Further, data collected during the trial (including sexually transmitted infection rate, self-reported sexual practices or measures of product adherence) could contradict the level of effectiveness assumed by the “95 assumption” discounting. However, the 95/95 method fixes the NI margin in the design stage and has no way to incorporate this critical information.

Using this, the assumed HIV incidence rate ratio between F/TDF and a counterfactual placebo group was calculated in the DISCOVER protocol as 2.64 (equivalent to 62% F/TDF effectiveness); see supplementary material for calculation and explanation. The non-inferiority margin is defined to be half-way between the assumed value and null value (1.00) on the log scale. That gives $\sqrt{(2 \cdot 64)} = 1.62$. If the CI for the comparison of F/TAF to F/TDF lies within the margin, then at least 50% of the assumed (62%) effectiveness of F/TDF is considered preserved by F/TAF. The 50% effect preservation criterion is entirely arbitrary. It may be a sensible standard but the scale (log rate ratio) is not readily interpretable. We have adopted a 50% preservation standard using the AIR scale which is more clinically relevant. The AIR standard has particular advantages in scenarios where effective product use sharply reduces the numbers of HIV infections in both arms.

Bayesian Methods for Constructing a Posterior Distribution of the Averted Infection Ratio

A natural choice to model the marginal distribution for the (underlying/unobserved) incidence rates (IR) of rectal gonorrhoea and of HIV in different populations is the gamma distribution. The gamma distribution is restricted to positive values, and allows considerable flexibility of

shape with only two parameters. We used Moran's bivariate gamma distribution to jointly describe the IR of rectal gonorrhoea and of HIV within any given population.² This distribution can be defined by first transforming random variables (Z_1 and Z_2) from the standard bivariate normal distribution with correlation ρ into a copula $C(\Phi(z_1), \Phi(z_2))$, where Φ is the standard normal cumulative distribution function, and secondly using the inverse cumulative distribution functions of univariate gamma distributions, denoted $F()$ and $G()$, respectively. This ($W_1=F^{-1}(\Phi(Z_1)), W_2=G^{-1}(\Phi(Z_2))$) yields W_1 and W_2 which follow a joint distribution function with marginal distribution which follow the gamma. This joint distribution provides some intuition regarding the strength of association between variables indicated by the ρ parameter because ρ is related to the correlation coefficient in the Normal copula.

We specify priors for two parameters of each marginal gamma distribution (e.g. for the incidence rates of rectal gonorrhoea and HIV, respectively) and for the association parameter ρ . We set a uniform prior over $\{-1, 1\}$ for ρ . For the marginal gamma parameters we specified relatively informative priors, as the small number of studies available means that we cannot rely on the data alone to restrict the upper possible range of IR. Our prior distribution for the mean HIV incidence rate across different populations implies a plausible prior 95% range of 0.005 to 5.7 per 100 person-years.

We specified an exponential (rate=1) distribution for the shape parameter for both rectal gonorrhoea and HIV and exponential distributions with rates of 50 and 100, for the scale parameter respectively (equivalent to an average IR of 2.0 per 100 person-years and 1.0 per 100 person-years). Analyses were performed on the on the person-year scale, and then rescaled for reporting.

For each of the studies cited by Mullick and Murray,¹⁶ we extracted the raw data for n rectal gonorrhoea and HIV cases and person-years of follow-up (PYFU). For some studies, either the n cases or total PYFU had to be indirectly inferred from the reported incidence rate. The raw data were considered to follow a constant-hazard time-to-event model, dependent on the study- and disease-specific IR in each case (unobserved and treated as a latent variable). For the DISCOVER trial, we extracted the n rectal gonorrhoea cases from the study report, and inferred the relevant total PYFU from the reported incidence rate (this does not match the PYFU for HIV incidence, as follow-up in each patient is stopped at first observation of rectal gonorrhoea). Our statistical model included latent variables for both the rectal gonorrhoea IR in DISCOVER and the HIV IR that would have been observed without PrEP, we obtain a conditional distribution for the latter based on the observed rectal gonorrhoea data.

The HIV incidence rates for the on TDF and on TAF arms in the DISCOVER trial used independent prior exponential distribution for the IR with rate parameter 100 (on the /person-

year scale). The AIR was then sampled over the posterior distribution of all relevant model parameters.

All Bayesian analyses are conducted using the Stan probabilistic programming language with 10000 samples from the posterior distribution. Individual model parameters are summarised with posterior expectations and 95% credibility intervals.

Additional References

1. Yu B, Yang H, Sabin A. A note on the determination of non-inferiority margins with application in oncology clinical trials. *Contemp Clin Trials Commun.* 2019; 16:100454
2. Moran PAP. Statistical inference with bivariate gamma distributions. *Biometrika* 1969; 56:627–34.

Supplementary Table 1: DISCOVER Margin Derivation

Study Name*	F/TDF		Placebo		Placebo v. F/TDF HIV Incidence Log Rate Ratio (SE)
	Person Years	HIV+ No (Rate)**	Person Years	HIV+ No (Rate)**	
Ipergay ¹⁴	220	2 (0.9)	212	14 (6.6)	2.0 (0.76)
PROUD ¹³	243	3 (1.2)	222	20 (9.0)	2.0 (0.62)
iPrEx ¹²	1328	23 (2.6)	1322	56 (5.5)	0.90 (0.25)
Pooled+		1.6		7.0	1.62 (0.34)

*Rate per 100 person years

** iPrEx results taken from among the subset of iPrEx participants reporting recently condomless receptive anal sex at screening

+ Pooled values from an equal-weights meta-analysis. Results under rate column are rates per 100 person year.

The 95% confidence interval for the incidence rate ratio for an equal weighted meta analysis is $\exp(1.62-1.96*0.34) = 2.64$ to $\exp(1.62+1.96*0.34)=9.84$.

The NI margin by 95/95 is the square-root of the 95% confidence interval lower bound, $\sqrt{(2.64)} = 1.62$.

The non-inferiority margin is based on the abstracted results of 3 completed F/TDF PrEP trials with non-PrEP controls. Note: (1) PROUD did not use a matching placebo (2) the DISCOVER protocol misstates the person years, and hence, the HIV incidence for iPrEx. The non-inferiority margin is unaffected, but the pooled HIV incidence is slightly changed from the DISCOVER protocol.

Supplementary Table 2: HPTN 083 Margin Derivation

Study Name*	F/TDF		Placebo		Placebo v. F/TDF HIV Incidence Log Rate Ratio (SE)
	Person Years	HIV+ No (Rate)**	Person Years	HIV+ No (Rate)**	
Ipergay ¹⁴	220	2 (0.9)	212	14 (6.6)	2.0 (0.76)
PROUD ¹³	243	3 (1.2)	222	20 (9.0)	2.0 (0.62)
iPrEx ¹²	2124	48 (2.3)	2113	83 (3.9)	0.58 (0.20)
Pooled+					0.80 (0.19)

*Rate per 100 person years

** iPrEx results taken from all participants

+ Pooled values from an weighted by variance meta-analysis. Results under rate column are rates per 100 person year.

The 95% confidence interval for the incidence rate ratio for the meta-analysis is IRR from 1.52 to 3.23. The confidence interval gives the discounted IRR based on the assumptions about how previous trials inform the expected effectiveness of F/TDF in the current trial. This is how the margins for DISCOVER and HPTN 083 differ.

The NI margin by 95/95 is the square-root of the 95% confidence interval lower bound, $\sqrt{(1.52)} = 1.23$. The translation of the discounted IRR into a margin is how the 50% effect preservation is maintained and it is calculated identically, taking the square of the discounted IRR value, in the two protocols.