



Published in final edited form as:

*Stat Methods Med Res.* 2019 ; 28(10-11): 3318–3332. doi:10.1177/0962280218801134.

## Adaptive non-inferiority margins under observable non-constancy

Brett Hanscom<sup>1</sup>, James P Hughes<sup>1,2</sup>, Brian D Williamson<sup>2</sup>, Deborah Donnell<sup>1</sup>

<sup>1</sup>Statistical Center for HIV AIDS Research and Prevention, Fred Hutch Cancer Research Center

<sup>2</sup>Department of Biostatistics, University of Washington

### Abstract

A central assumption in the design and conduct of non-inferiority trials is that the active-control therapy will have the same degree of effectiveness in the planned non-inferiority trial as in the prior placebo-controlled trials used to define the non-inferiority margin. This is referred to as the ‘constancy’ assumption. If the constancy assumption fails, decisions based on the chosen non-inferiority margin may be incorrect, and the study runs the risk of approving an inferior product or failing to approve a beneficial product. The constancy assumption cannot be validated in a trial without a placebo arm, and it is unlikely ever to be met completely. When there are strong, observable predictors of constancy, such as dosing and adherence to the active-control product, we can specify conditions where the constancy assumption will likely fail. We propose a method for using measurable predictors of active-control effectiveness to specify non-inferiority margins targeted to the planned study population characteristics. We describe a pre-specified method, using baseline characteristics or post-baseline predictors in the active-control arm, to adapt the non-inferiority margin at the end of the study if constancy is violated. Adaptive margins can help adjust for constancy violations that will inevitably occur in real clinical trials, while maintaining pre-specified levels of type I error and power.

### Keywords

non-inferiority margin; adaptive design; non-constancy; meta-analysis; HIV prevention; PrEP

### Introduction

Non-inferiority (NI) trials are designed to determine whether a new therapy is as effective as, or at least not meaningfully worse than, an existing standard-of-care therapy. By using previously approved therapies as controls, NI trials are used to infer whether a new therapy is effective without a placebo arm, given that placebo controls are unethical when effective treatments exist. Without a placebo arm, however, strong assumptions must be made about the effectiveness of the active-control therapy in the NI trial. Specifically, it must be assumed that the active-control therapy will be as effective in the planned NI trial as it was in prior

---

**Corresponding author:** Brett Hanscom, Fred Hutch Cancer Research Center, 1100 Fairview Ave NE, Seattle, WA 98109, USA, bhanscom@fredhutch.org.

Declaration of Conflicting Interests: The authors declare that there is no conflict of interest.

placebo-controlled trials. This is referred to as the ‘constancy’ assumption. If the constancy assumption fails, inferences made from the NI trial could be invalid. For example, if the active-control is not as effective as expected, an NI trial could lead to approval of an ineffective or harmful therapy. Conversely, if the active-control is more effective than expected, an NI trial could fail to achieve approval for an effective therapy.

To mitigate the effects of non-constancy, various authors suggest modifications to the non-inferiority margin when evidence of constancy failure exists. Everson-Stewart [4] proposed a model for assessing constancy based on heterogeneous effectiveness in population subgroups, and recommended tightening the margin, or moving to a superiority design, if constancy is violated. Koopmeiners and Hobbs [9] developed a Bayesian approach to adjusting the NI margin based on inter-trial heterogeneity. Nie and Soon [12, 13] published a regression-model approach for identifying potential non-constancy, and use baseline population characteristics to define a non-inferiority margin appropriate for the enrolled study participants.

In this article we extend Nie and Soon’s idea of a regression-based non-inferiority margin to include trial-level data and post-randomization factors in the active-control arm. We then propose several methods for post-trial adaptation of the margin, and evaluate how each method influences operating characteristics such as Type-I error and power. A trial testing a novel HIV pre-exposure prophylaxis (PrEP) agent compared to an established effective daily oral pill is used for illustration.

### Defining the NI Margin

A non-inferiority margin  $\delta$  is the amount by which an experimental product  $E$  can be worse than a standard of care therapy  $C$  and still be considered clinically useful. Statistical inference in a non-inferiority trial is based on the null hypotheses that the treatment effect is equal to  $\delta$ , as opposed to superiority trials where the null hypothesis represents ‘no effect’ (e.g.,  $RR=1.0$ ). In a trial estimating  $RR_{E/C}$ , the relative risk comparing  $E$  to  $C$ , the statistical hypotheses are

$$\begin{aligned} H_0: RR_{E/C} &= \delta \\ H_1: RR_{E/C} &< \delta \end{aligned} \quad (1)$$

and provided that the observed value  $\widehat{RR}_{E/C}$  is significantly less than  $\delta$ , the new product is considered ‘non-inferior’ to  $C$ . If  $\delta = 1.1$ , for example, then  $E$  is allowed to have a 10% higher risk than  $C$  and still be considered clinically important. (Note that although the parameter ‘relative risk’ is used here for illustration, non-inferiority margins can be specified on any scale, and the methods we propose can be applied generally to any treatment-comparison parameter.)

To protect against falsely declaring non-inferiority (inflated type I error) if the active control is not as effective as in prior trials, non-inferiority margins are often chosen to be conservative (i.e., less likely to produce a statistically significant result). Defining the margin typically involves estimating the effect of an active-control therapy based on prior placebo-controlled trials, and choosing a margin that conserves some proportion  $\rho$  of the active-

control effect [5]. For example, consider a series of two arm trials comparing the active-control therapy  $C$  to placebo  $P$ , and let  $\widehat{RR}_{C/P}$  be the estimated relative risk based on a meta-analysis. The non-inferiority margin  $\delta$  can be defined, conservatively, to preserve at least  $(100 \times \rho)\%$  of the active-control benefit (on the log scale) by setting

$$\delta = \left( LCL_{95}(\widehat{RR}_{P/C}) \right)^{1-\rho} \quad (2)$$

where  $LCL_{95}$  is the lower limit of the 95% confidence interval and  $\widehat{RR}_{P/C} = \widehat{RR}_{C/P}^{-1}$ .

The lower confidence limit can be thought of as the ‘assured effect’, i.e. evidence from prior trials rules out a smaller effect, but does not assure anything larger. The assured effect is typically referred to as ‘M1’. Using the  $LCL_{95}$  as the estimated active-control effect acknowledges the uncertainty associated with the meta-analytic estimate and provides a degree of protection against non-constancy. Values for  $\rho$  can range between 0 and 1, where  $\rho = 0$  preserves none of the benefit of  $C$  but assures that the experimental therapy  $E$  is at least minimally better than placebo. Choosing  $\rho = 1$  gives a margin of 1.0 which is equivalent to requiring superiority of  $E$  over  $C$ . The value for  $\rho$  is often taken to be 0.50, yielding a margin that preserves at least 50% of the benefit provided by  $C$  (on the log relative risk scale). A margin that preserves at least some of the active-control benefit is commonly referred to as the ‘M2’ margin. If the results of a non-inferiority trial comparing  $E$  to  $C$  satisfy

$$UCL_{95}(\widehat{RR}_{E/C}) < \delta \quad (3)$$

where  $UCL_{95}$  is the upper limit of the 95% confidence interval, then the trial results are consistent with non-inferiority. Appropriate values for  $\rho$  will depend on context and should be justified for any individual trial.

### Population-specific NI margins

Non-constancy can occur for many reasons, including differences in participant characteristics, differences in dosing or adherence, changes in background supportive care, or actual declines in biological benefit of the active control (e.g., due to antibiotic resistance.) The meta-analysis-based margin  $\delta$  in (2) effectively assumes that the active-control effect in the planned trial will correspond to the average effect observed in prior trials, and ignores predictive factors that might provide a more precise specification of the margin. While some sources of non-constancy are unobservable, when observable participant characteristics are known to be modifiers of effectiveness for the active-control therapy, they can be used to specify margins appropriate to the (assumed) characteristics of the planned study population.

Effect modifiers are often identified using *post hoc* subgroup analyses within individual trials. Although informative, subgroup analyses are generally underpowered and exploratory in nature. In addition, analyses of effectiveness based on post-randomization factors such as drug adherence do not benefit from the protection of randomization. Effect modifiers are more reliably identified by using meta-analysis regression to aggregate results across multiple studies. Meta-analysis regression can improve power and precision by virtue of the

combined, increased sample size [17]. While it is well recognized that cross-trial comparisons may be influenced by ecological bias [7], by comparing post-randomization effects across studies, rather than within studies, the meta-analysis approach can reduce the potential for confounding. By treating factors such as drug adherence as trial-level variables in a meta-regression model, post-randomization effect modifiers can be identified and estimated with reduced risk of the confounding usually associated with analysis of post-randomization subgroups.

To estimate the size and importance of potential effect modifiers, we use mixed-effects meta analysis extended to a regression model that includes study-level, fixed-effects covariates [15]. Predictive factors are divided into two categories: (1) fixed population attributes such as race and gender that are measured at baseline, and (2) dynamic features, such as dosing and drug adherence, that could change during the course of a trial and cannot be assessed until the trial is underway. The following model is used for  $RR_j$ , the relative risk comparing P to C for study  $j$ :

$$\log(RR_j) = \beta_0 + \beta_1 x_j + \beta_2 z_j + b_j + \varepsilon_j \quad (4)$$

where  $x_j$  is a set of population attributes,  $z_j$  is a set of dynamic features,  $b_j$  are the study-specific random effects with  $b_j \sim N(0, \tau^2)$ ,  $\tau^2$  represents between-study heterogeneity not explained by  $x_j$  and  $z_j$ , and  $\varepsilon_j$  are error terms with  $\varepsilon_j \sim N(0, \sigma_j^2)$  where  $\sigma_j^2$  represent within-trial sampling variability. Fixed-effect parameter estimates  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ , and  $\hat{\beta}_2$  are used to estimate effectiveness in a new study with anticipated attributes  $x^A$  and features  $z^A$ , by setting

$$\widehat{RR}_{P/C}(x^A, z^A) = \exp(\hat{\beta}_0 + \hat{\beta}_1 x^A + \hat{\beta}_2 z^A). \quad (5)$$

A population-specific NI margin is computed based on the lower 95% confidence limit of the regression model estimate (population-specific M1), preserving at least the fraction  $\rho$  of the active control benefit, by setting

$$\delta(x^A, z^A, \rho) = \left( LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)) \right)^{1-\rho}, \quad (6)$$

which represents a population specific value for M2. Recall that because the NI trial is a comparison of experimental versus active control, somewhat counterintuitively the active-control effect is formulated as the placebo-versus-control relative risk (i.e., a measure of how much worse the placebo performs as compared to the active control therapy.) The NI margin defines how much worse the experimental therapy can be as compared to the active-control therapy.

### Case study: HIV pre-exposure prophylaxis (PrEP)

HIV/AIDS remains a global pandemic with no vaccine or cure; prevention strategies are therefore desperately needed. Daily oral TDF/FTC, used as pre-exposure prophylaxis (PrEP), has been shown in multiple randomized, placebo-controlled trials to reduce the risk of HIV infection [1, 2, 6, 10, 11, 16, 18]; however the estimated benefit varies widely across

different studies. Because many people are unable to take daily oral PrEP consistently, there is strong impetus for developing long-acting products. Given the established effectiveness of TDF/FTC, it is unethical to use placebo controls, and hence active-control trials are now the most appropriate design for testing new prevention therapies.

Based on the meta-regression analysis including all prior PrEP trials, two factors are predictive of oral PrEP effectiveness: adherence and gender. Fitting model (5) to the PrEP-trial data gives the following:

$$\widehat{RR}_{PI}(sex, adherence) = \exp(0.7520 - 0.1058 \times sex - 2.276 \times adherence) \quad (7)$$

where *sex* is an indicator variable of sex at birth with 1 = *male* and 0 = *female*, and *adherence* is a measure between zero and one of the proportion of active-arm participants with detectable plasma levels of PrEP.

Figure 1 shows a scatterplot of trial-level results as a function of adherence, as well as the fitted regression line for men, along with confidence bounds. For a planned study in men, the lower confidence limit represents M1 and would be used as the basis for computing an NI margin, depending on expected adherence. The regression line in Figure 1 drops below 1.0 at 0.3, suggesting a threshold effect whereby PrEP provides little observable protective benefit in a population where adherence is below 30%. A similar fitted line and confidence bound could be generated for a study in women, or in a study with a mix of men and women.

A range of margins – derived from the fitted model – are shown in Table 1 based on different levels of assumed adherence in the study population. Low adherence requires lower (stricter) margins and higher adherence allows larger margins. When projected adherence is lower than approximately 40%, the value M1 is below 1.0, indicating that there is not strong evidence for effectiveness, and that a superiority trial is recommended. (A non-inferiority trial with a margin of 1.0 is equivalent to a superiority trial.) This suggests that the meta-regression model may be used not only to set population-specific NI margins, but also to determine under what circumstances an NI trial is appropriate.

### Type I error and power under non-constancy

Even if a population-specific approach is used to select the margin, the observed values of the effect modifiers in the study population may not match the values used in the planning phase. If the observed values are substantially different from the planning phase, the predicted efficacy of the active control will be different than planned, and the constancy assumption will not hold. With the pre-planned NI margin the trial runs the risk of declaring support for a product that doesn't work (type I error), or failing to support a product that does work (type II error).

To illustrate, consider a trial that is designed under the assumed values  $x = x^A$  and  $z = z^A$ , with corresponding NI margin  $\delta(x^A, z^A, \rho)$  as in (6). The statistical hypotheses are

$$\begin{aligned} H_0: RR_{E/C} &= \delta(x^A, z^A, \rho) \\ H_1: RR_{E/C} &< \delta(x^A, z^A, \rho) \end{aligned} \quad (8)$$

where  $\delta()$  will typically be larger than 1.0. Re-expressing these hypotheses in terms of incidence rates helps illustrate the impact of non-constancy. The relative effectiveness  $RR_{E/C}$  can be written as  $\lambda_E/\lambda_C^A$ , where  $\lambda_E$  is the incidence rate in  $E$ , and  $\lambda_C^A$  is the incidence rate in  $C$  when  $x = x^A$  and  $z = z^A$ . The null hypothesis in (8) can then be written as

$$H_0: \lambda_E = \lambda_E^0 = \lambda_C^A \delta(x^A, z^A, \rho), \quad (9)$$

where  $\lambda_E^0$  is defined to be the highest allowable incidence under  $E$  that would be considered clinically acceptable.

The alternative hypothesis used to compute sample size and power is often based on the desire to reject  $H_0$  if  $E$  is equivalent to  $C$ , i.e., if  $H_1: RR_{E/C} = 1.0$ . If there is reason to expect that  $E$  may outperform  $C$  then the study may be powered to establish a somewhat stronger effect, say  $H_1: RR_{E/C} = \xi$ , where  $\xi \in [0, 1]$ . The alternative hypothesis can then be expressed in terms of incidence rates as

$$H_1: \lambda_E = \lambda_E^1 = \lambda_C^A \xi. \quad (10)$$

Note that since an NI trial is designed to rule out the margin, not equality, the trial is powered to detect an effect size (ratio of the alternative to the null) equal to  $\xi/\delta(x^A, z^A, \rho)$ .

If the true study population characteristics are  $x^T$  and  $z^T$  such that  $RR_{C/P}(x^T, z^T) \neq \delta(x^A, z^A, \rho)$  then the planned margin  $\delta(x^A, z^A, \rho)$  will be incorrect and the study will not retain desired levels of type I error and power. By expressing  $RR_{E/C}$  in terms of incidence rates, and using the fact that  $\lambda_C^T = \lambda_P RR_{C/P}(x^T, z^T)$ , where  $\lambda_P$  is incidence in a hypothetical placebo arm, type I error and power can be expressed as a function of  $RR_{C/P}(x^T, z^T)$ :

$$\begin{aligned} P[\text{Type I error}] &= P\left[ UCL_{95}\left(\frac{\lambda_E}{\lambda_P RR_{C/P}(x^T, z^T)}\right) < \delta(x^A, z^A, \rho) \mid \lambda_E = \lambda_E^0 \right] \\ \text{Power} &= P\left[ UCL_{95}\left(\frac{\lambda_E}{\lambda_P RR_{C/P}(x^T, z^T)}\right) < \delta(x^A, z^A, \rho) \mid \lambda_E = \lambda_E^1 \right]. \end{aligned} \quad (11)$$

Error rates for an example trial are plotted in Figure 2 as a function of the (hypothetical) percent risk reduction provided by  $C$  (versus placebo) in the true study population, or  $(1 - RR_{C/P}(x^T, z^T)) \times 100$ . If  $C$  is more effective than planned,  $E$  will not look as good in comparison, and the type II error rate will be high (i.e., the trial will have low power.) If  $C$  is less effective than planned,  $E$  will appear better in comparison to  $C$ , and the chances of a false-positive result will increase. Even fairly small deviations from constancy can have a substantial impact on the probability of making a false conclusion. For instance if  $C$  is 60% effective instead of the planned 50%, the type II error probability increases from 10% to

48% (i.e., power drops from 90% to 52%). Likewise, if  $C$  is only 40% effective, the type I error probability increases from 2.5% to 16%. To reduce these error probabilities we propose an adaptive NI margin approach based on the pre-specified meta-regression model used to select the study NI margin, and observed values of the effect modifiers in the NI study population.

### Adaptive NI Margins

Although an NI margin must be pre-specified in order to plan an NI trial, the margin used for planning may not always be the appropriate gauge to judge whether the experimental product is truly effective. As demonstrated in the previous section, using the NI margin based on planned characteristics, rather than observed, can be detrimental to the operating characteristics of a trial, and fail to ensure that the NI trial conclusions are valid.

To make certain that an NI trial will only support a therapy that meets a pre-specified level of effectiveness, we propose adapting the planned margin using equation (5) together with observed study-population characteristics. The adaptive margin is based on the idea of simply inserting the observed values  $x^T$  and  $z^T$  into equation (6) to compute the adapted margin

$$\delta^a = \left( LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)) \right)^{1-\rho}. \quad (12)$$

We define a more general notation that encompasses multiple approaches to adapting the margin. Let  $\Delta$  be the relative risk defining the amount of benefit an experimental therapy is required to provide over a (hypothetical, unobserved) placebo. The adaptive M2 margin, expressed in terms of  $\Delta$ , is

$$\delta^a = \Delta \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)), \quad (13)$$

and  $(1 - \rho) \times 100$  represents the required percent risk reduction over placebo.  $\Delta$  can also be thought of as the ratio M2/M1. The general approach will use meta-regression to compute M1 for the observed study population, and then use a pre-specified method to compute and generate the adapted M2 margin based on equation (13).

There are two general strategies for specifying  $\Delta$ : the first pre-specifies the desired percent risk reduction relative to placebo (fixed  $\Delta$ ); and the second pre-specifies the proportion of (observed) active-control benefit that must be preserved by the experimental therapy (fixed  $\rho$ ).

For method one, a fixed level of benefit  $\Delta$  is chosen, which could be either (a) the amount of benefit over placebo required by the planned margin, or (b) an investigator-defined minimal clinically important difference (MCID). For example, if the planned margin is the meta-regression-based margin  $\delta(x^A, z^A, \rho)$  described in (6), the planned value for  $\Delta$  is

$$\begin{aligned}
 \Delta_{\text{Plan}} &= \delta(x^A, z^A, \rho) \times LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A))^{-1} \\
 &= \left( LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)) \right)^{1-\rho} \times LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A))^{-1} \\
 &= \left( LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)) \right)^{-\rho},
 \end{aligned} \tag{14}$$

which uses the planned values  $x^A$  and  $z^A$  in relation (13). If  $x^T = x^A$  and  $z^T = z^A$  then  $\delta^a = \delta(x^A, z^A, \rho)$ . This strategy may be used by investigators who wish to keep constant the required benefit over placebo, regardless of enrolled-participant characteristics.

An NI trial might also be planned using a fixed MCID, based on investigator consensus and/or expert opinion. For example, it might be determined that regardless of study population characteristics, it is essential that the experimental product provide a reduction in risk of at least 10% ( $RR_{EP} = 0.90$ ) over what would be expected with placebo. In this case would be defined as

$$\Delta_{MCID} = RR_{MCID} = 0.90, \tag{15}$$

where  $RR_{MCID}$  is the risk reduction corresponding to the pre-specified MCID. Since relative risks increase in value as benefit decreases,  $RR_{MCID}$  is actually the maximum allowable relative risk.

The second strategy defines based on a fixed proportion  $\rho$  of the benefit provided by the active-control therapy in the *observed* study population. Substituting  $x^T$  and  $z^T$  into equation (14), yields

$$\Delta_{\text{Estimated}} = \left( LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)) \right)^{-\rho}, \tag{16}$$

which can be used in (13) to generate the adjusted margin in (12).

This second approach may be desirable when the active-control therapy has different levels of effectiveness in different populations. For example, assume that the active-control treatment has been shown to be more effective in adults than in adolescents, and that, despite plans to recruit adults, the NI study population is mostly adolescents. Because the study population has more adolescents than planned, the effectiveness of the active-control therapy is assumed to be lower than planned. Nevertheless, investigators may still be content with an experimental therapy that preserves at least 50%, say, of the benefit achievable by the active-control among adolescents. By using the planned, fixed value for  $\rho$  (0.5) and the estimated value for M1, can be adapted using equation (16). The value of M1 (i.e.,  $LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))$ ) will be lower than originally planned, giving a higher than planned value of  $\Delta_{\text{Estimated}}$  and although this corresponds to a smaller required percent risk reduction than planned (i.e., a smaller value of  $(1 - \Delta_{\text{Estimated}}) \times 100$ ), it still ensures that the new therapy preserves 50% of active-control benefit in the enrolled study population.



## Placing limits on change in NI margins

Both of the approaches defined above have undesirable properties when there are extreme changes in effect modifier characteristics from the planned trial. In the second approach using  $\Delta_{Estimated}$ , if active-control effectiveness in the NI trial is much lower than expected, the adapted NI margin fails to ensure that the experimental therapy provides any benefit over placebo. In both approaches, if active-control effectiveness is much higher than expected, the adapted margin can be arbitrarily high. Both problems can be controlled by placing limits on the margin.

If low active-control effectiveness is a concern,  $\Delta_{Min}$  may be defined by selecting the more stringent of the two choices  $\Delta_{MCID}$  or  $\Delta_{Estimated}$ , which is accomplished by choosing the minimum:

$$\begin{aligned}\Delta_{Min} &= \min(\Delta_{MCID}, \Delta_{Estimated}) \\ &= \min(RR_{MCID}, (LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)))^{-\rho})\end{aligned}\quad (17)$$

This will typically mean using  $\Delta_{Estimated}$  when the active-control effect is as planned or larger, and using  $\Delta_{MCID}$  when the effect of the active-control effect is estimated to be relatively small. Using this method prevents the level of required effectiveness from diminishing too far in a study population where the active-control therapy is thought to be not working well, for example due to low adherence.

Investigators or regulators may also wish to impose an upper limit on the NI margin, but allow adaptation of the margin below the maximum level. In this case  $\Delta_{Max}$  may be defined by considering  $\Delta_{Estimated}$  in combination with a maximum value for the NI margin,  $\delta^{Max}$ . The value of  $\Delta_{Cap}$  corresponding to the desired margin is

$$\Delta_{Max} = \delta^{Max} / (LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))), \quad (18)$$

which may then be combined with  $\Delta_{Estimated}$  to place a cap on  $\delta$  by setting

$$\Delta_{Cap} = \min(\Delta_{Max}, \Delta_{Estimated}). \quad (19)$$

This strategy will typically mean using  $\Delta_{Estimated}$  when the active-control effect is as planned or smaller, and using  $\Delta_{Max}$  when the effect of the active control is estimated to be relatively large. Setting a maximum prevents the margin from increasing to a point where the experimental therapy is allowed to be substantially worse than the active control. Although this technique will effectively require the proportion of preserved active-control benefit  $\rho$  to increase as active-control effectiveness increases, investigators and/or regulators may feel more comfortable placing a cap on the absolute magnitude of the NI margin.

Figure 3 illustrates how varying levels of effectiveness in the active-control arm leads to different adaptive NI margins using  $\Delta_{Estimated}$  and  $\Delta_{Min}$ . If effectiveness is higher than planned,  $\delta^a$  will shift to the right relative to  $\delta$ , thereby relaxing the NI margin. Since  $\Delta_{Estimated}$  is smaller than  $\Delta_{MCID}$  in this case,  $\Delta_{Min} = \Delta_{Estimated}$  and the adapted margin is

same for both approaches. If active-control effectiveness is somewhat lower than planned the adapted margin preserving proportional benefit will become correspondingly more stringent. If we require the margin to preserve an MCID, this reduces  $\delta^a$  to 1.0, effectively requiring the experimental intervention to be superior to the active control.

The bottom row in Figure 3 shows a scenario where active-control effectiveness is so low that there is no assured effect (i.e.,  $LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)) = 1.0$ ) and  $Estimated = 1.0$ . This implies a requirement of superiority when using  $Estimated$  but a requirement of ‘super superiority’ ( $\delta^a < 1.0$ ) when using  $Min = MCID$ . Super superiority simply means that the experimental product must not only be superior to the active control, but that the  $UCL_{95}(\widehat{RR}_{E/C})$  must rule out small benefits. For example, if  $\delta^a = 0.90$ ,  $UCL_{95}(\widehat{RR}_{E/C})$  must be less than 0.90, thus assuring the experimental intervention reduces risk by an additional 10% compared to the active control. Note that whenever  $\delta^a$  is fixed in advance (e.g., when using  $Plan$  or  $MCID$ ), super superiority may be required.

### Adapting the Statistical Hypotheses

Once the adapted margin  $\delta^a$  has been specified, this margin becomes the null hypothesis for statistical inference. Although the nominal value of the null hypothesis will have changed from the planning stage, the adapted null still corresponds to the pre-planned amount of benefit that the experimental therapy is required to provide relative to a hypothetical placebo. When active-control effectiveness depends on observable effect modifiers, and the null hypothesis is expressed in relation to the active-control therapy, the nominal value of the null hypothesis must change to ensure that the experimental treatment produces the required level of effectiveness over placebo. Note that if a trial is planned based on (6), the analysis margin (and hence the null hypothesis) will be the same as the planning margin if the constancy assumption is met, i.e., if  $x^T = x^A$  and  $z^T = z^A$ .

Just as the nominal value of the null hypothesis can change under an adaptive NI margin strategy, so too can the nominal value of the alternative hypothesis. Although the alternative hypothesis is typically expressed in relation to the active control, it is the alternative hypothesis in relation to the NI margin that determines power and sample size. We define the effect size  $\Omega$  as the ratio  $\xi/\delta$ , which represents the ratio of the null and alternative hypotheses. For example, if the alternative hypothesis is  $\xi = 0.90$  and the NI margin  $\delta$  is 1.1, the effect size used to compute power and sample size is  $\Omega = \xi/\delta = 0.82$ .

Because active-control effectiveness – and hence the NI margin – is a moving target under non-constancy, it is useful to anchor the alternative hypothesis to the hypothetical placebo arm, just as we did for the null. Let  $\Omega_{Plan}$  be defined as the target effect size of the experimental treatment over placebo, defined as

$$\Omega_{Plan} = \xi / LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)). \quad (20)$$

The effect size  $\Omega_{Plan}$  can be thought of as the target benefit of the experimental product over placebo, as opposed to the target benefit over the active-control therapy, and this target remains fixed even when the active-control effect changes as a result of non-constancy. The

planned alternative hypothesis for computing power can be expressed as a function of  $\Omega_{Plan}$  as

$$\xi = \Omega_{Plan} \times LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)), \quad (21)$$

and once the values for  $x^T$  and  $z^T$  have been observed we can adjust the nominal value of  $\xi$  to reflect the pre-planned target effect. Using the observed study population characteristics gives the following adaptive alternative hypothesis:

$$\xi^a = \Omega_{Plan} \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)). \quad (22)$$

Although the nominal value of  $\xi$  will have changed, by fixing  $\Omega$  we preserve the initial, planned target effect of the experimental treatment over placebo. The adapted value  $\xi^a$  can now be used to compute power under the new hypotheses, as discussed in the next section.

Investigators may not want to change the alternative hypothesis, even when faced with non-constancy. For example, a common non-inferiority alternative hypothesis is  $\xi = RR_{E/C} = 1.0$ , meaning that trial is designed to establish non-inferiority in the situation where experimental and active-control treatments are equally effective. This is a natural choice, however it will have important effects on power, as we will see in the next section.

### Updating Type-I Error and Power

In the context of potential non-constancy we think of the type-I error rate as the probability of declaring non-inferiority when the true  $RR_{E/C}$  is just at the point of being unacceptable, i.e., when  $RR_{E/C}$  is equal to an NI margin appropriate for the study population. In other words, the null hypothesis of interest is the adjusted null hypothesis reflecting enrolled study participants. If a fixed NI margin is used that is too high (too permissive), type-I error will be too high, and vice-versa for a margin that is too low. Adjusting the NI margin to reflect the observed study population removes the mismatch between the desired null hypothesis and fixed NI margin, and thereby prevents inflation and reduction in Type-I error. A fundamental assumption is that the adjusted NI margin is a reasonable estimate of the point at which a new therapy would be unacceptable. The validity of this assumption will depend on the quality of the trials used to construct the meta-regression model.

Statistical power will depend on the ratio of the adjusted alternative and null hypotheses, i.e., the adjusted effect size. Provided that this ratio does not change, power will be not be affected. For example, if the NI margin is planned using the meta-analysis regression in (6), the planned effect size can be written as the ratio of (21) and (6) as follows,

$$\begin{aligned} \frac{\xi}{\delta} &= \frac{\Omega_{Plan} \times LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A))}{(LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)))^{1-\rho}} \\ &= \Omega_{Plan} \times (LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)))^\rho. \end{aligned} \quad (23)$$

If  $\delta^a$  is computed using the pre-specified value  $\Omega_{Plan}$ , the adjusted effect size does not change:

$$\begin{aligned}\frac{\xi^a}{\delta^a} &= \frac{\Omega_{Plan} \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))}{\Delta_{Plan} \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))} \\ &= \frac{\Omega_{Plan}}{(LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)))^{-\rho}} \\ &= \Omega_{Plan} \times (LCL_{95}(\widehat{RR}_{P/C}(x^A, z^A)))^\rho.\end{aligned}\quad (24)$$

In other words, if the null and alternative hypotheses are adjusted by using the pre-specified values for  $\Omega_{Plan}$  and  $\Delta_{Plan}$ , the effect size ratio remains the same, and there is no loss or gain in power. However, if  $\xi$  is allowed to vary depending on observed population characteristics, the effect size will no longer remain constant. Using estimated effectiveness to define  $\xi$  as in (16), the adjusted effect becomes

$$\begin{aligned}\frac{\xi^a}{\delta^a} &= \frac{\Omega_{Plan} \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))}{\Delta_{Estimated} \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))} \\ &= \Omega_{Plan} \times (LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T)))^\rho,\end{aligned}\quad (25)$$

which is a function of observed population characteristics. This means that on the one hand, if the active-control therapy is observed to be more effective than planned, the effect size, and consequently the power, will be lower than planned. On the other hand, if the active-control therapy is estimated to be worse than planned, the effect size will be larger than planned and power will be greater than planned.

Similarly, if the null hypothesis is adjusted but the alternative hypothesis remains constant at, for example,  $\xi = 1.0$ , the effect size, and hence power, will vary depending on the adjusted null, following the equation

$$\frac{\xi}{\delta^a} = \frac{1}{\Delta \times LCL_{95}(\widehat{RR}_{P/C}(x^T, z^T))}, \quad (26)$$

where  $\Delta$  could be  $Plan$ ,  $Adjusted$ ,  $MCID$ ,  $Min$ , or  $Cap$ . In this case the effect size is just the inverse of the non-inferiority margin. Stronger observed effects in the active-control arm (i.e., larger values for placebo-versus-control relative risk) will imply larger effect sizes (smaller absolute value of the relative risk) and hence higher power. Conversely, weaker than expected effects in the active-control arm would yield smaller effect sizes and reduced power. This is intuitive, since if the two interventions are equally effective, it is more difficult to establish non-inferiority using a narrower margin.

### Case Study: HIV PrEP

Returning to the example of HIV PrEP, assume a trial is planned to evaluate a new long-acting therapy in men, and adherence to oral PrEP in the active-control arm is projected to

be 60%. The planned NI margin can be taken from Table 1 as  $\delta = 1.23$ . Assuming a planned alternative hypothesis  $\xi = 0.80$ , the effect size  $\Omega$  would be  $0.80/1.23 = 0.52$ , and the study would achieve 90% power with a sample size of 231 new HIV infections.

Table 2 illustrates how the four ways of defining  $\delta^a$  would influence the adapted margin  $\delta^a$  at the end of a trial, depending on observed active-control arm adherence, and also how power is affected by the choice of alternatives. In approach one (first three rows of Table 2),  $\delta^a$  is maintained at the planned level ( $\delta_{Plan} = 1.23/1.5 = 0.82$ ). If adherence is higher than planned (70%) or lower than planned (50%) the margin changes substantially to 1.54 or 0.95, respectively. Although  $\delta_{Plan}$  is constant, these margins preserve very different proportions ( $\rho$ ) of the active control benefit as compared to the planned level; equivalent to  $\rho = 0.32$  and  $\rho = 1.33$  respectively. (Note that the value  $\rho = 1.33$  corresponds to the requirement that the benefit of the experimental agent be 33% larger than the assured effect of the active control, and is equivalent to requiring super superiority.) If instead the target benefit  $\Omega$  over placebo is held constant at 0.53 (middle four columns of Table 2), the effect size will not change and the target of 90% power will be achieved regardless of observed effectiveness. And finally, if the alternative hypothesis is fixed (last four columns of Table 2), higher than planned active-control effectiveness leads to too much power (nearly 100%) and lower than planned effectiveness results in low power (26%).

If instead the investigators wish to adapt  $\delta^a$  to maintain proportional benefit ( $\rho = \text{Estimated}$  and  $\rho = 0.5$ ) as shown in the middle three rows of Table 2, changes to the margin will be less dramatic, with higher and lower than planned effectiveness leading to margins of 1.37 and 1.08 respectively. If the target benefit  $\Omega$  is held constant at 0.53, the adapted alternative will change by the same amount as when preserving planned benefit, and the effect size will no longer remain stable. Higher than planned effectiveness leads to a smaller effect size (RR=0.73) while lower than planned effectiveness leads to a larger effect size (RR=0.57), with corresponding reductions and increases in power. If, on the other hand, a fixed alternative is selected ( $\xi = 0.80$ ), the effect size responds to changes in effectiveness much the same way it does when using  $\delta_{Plan}$ , only the changes in power are less extreme (power becomes 98% and 63% for higher and lower effectiveness, respectively).

When a minimum benefit requirement ( $\delta_{MCID} = 0.90$ ) is also imposed on the adapted margin, as shown in the third set of results in Table 2, the results match those of previous example when effectiveness is higher than planned, but differ when effectiveness is lower than planned. For lower than planned active-control effectiveness, we have

$$\Delta = \min(\Delta_{Estimated}, \Delta_{MCID}) = \min(0.93, 0.90) = 0.90, \quad (27)$$

which reduces the margin to 1.05 and increases the required proportion  $\rho$  to 0.69. The effect size for the fixed  $\Omega$  scenario is slightly reduced, although power is similar, and the effect size is also reduced for the fixed  $\xi$  scenario, with power correspondingly dropping from 63% to 54%.

If a maximum NI margin  $\delta^{Max}$  is imposed, as shown in the final section of Table 2,  $\delta^a$  is constrained at 1.23 even when active-control effectiveness is higher than planned. The

proportion of benefit preserved increases to 0.67, and power drops dramatically under the adapted alternative hypothesis.

### Dynamic features and sampling variation

It will not always be possible to measure dynamic features in the entire study cohort. Lab-based drug adherence assessment, for example, requires costly collection and testing of samples. Typically it will be sufficient to generate an unbiased estimate of adherence using a random subset of participants, at a random set of time points. To compute the adapted margin, a sample-based estimate  $\hat{z}^T$  may be substituted into (13) in place of  $z^T$ . The degree to which sampling variation may influence the margin depends on both the estimated regression coefficients in (5) and the sampling distribution of  $z^T$ . Potential bias in  $\delta(x^T, z^T, \rho)$  introduced by sampling can be quantified by

$$\delta(x^T, z^T, \rho) - \int \delta(x^T, \hat{z}^T, \rho) f(\hat{z}^T) d\hat{z}^T \quad (28)$$

where  $f()$  is the sampling distribution of  $\hat{z}^T$  and  $\delta()$  is defined as in equation (6). The true value of  $z^T$  will generally not be known, but sample sizes for  $\hat{z}^T$  should be chosen to provide precise estimates of  $z^T$  and minimize potential bias.

## Discussion

In regulatory settings, NI margins must be set in advance. However, specifying non-inferiority margins is an imprecise and subjective process, and the validity of these margins relies heavily on the assumption of constancy. A pre-specified adaptive margin approach, included as a secondary or sensitivity analysis in a trial, could have considerable credibility where the trial data suggest non-constancy. Once the trial is complete, the pre-specified regression model and pre-specified adaptive method can be used to update the end-of-study margin according to observed effect modifiers in the study population, preserving planned levels of Type-I error and assuring pre-planned levels of benefit over placebo.

We have proposed two different approaches for defining an adaptive NI margin, and the rationale for each choice is slightly different. Decisions to limit the potential change in the margin need to be specified in advance. Determining the most appropriate strategy for adapting the margin will depend on the goals of the trial, the factors that influence non-constancy, and the investigators' perspective.

The International Council for Harmonization draft E9(R1) guidelines recommend defining study estimands with respect to events that occur after randomization [8]. Our proposed approach involves two estimands, (1) the relative effectiveness of the experimental therapy compared to the active-control, and (2) the effectiveness of the active-control therapy compared to a hypothetical placebo arm. We assume that estimand (1) will be estimated using the intent-to-treat (ITT), or 'treatment policy' strategy, which recognizes that post-randomization events, such as non-adherence, may directly influence the estimates. Estimand (2) is also constructed based on ITT estimates from prior placebo-controlled trials, and represents an average effect given observed study-population behavior. Combining these

estimates allows investigators to infer whether the experimental treatment provides sufficient clinical benefit as compared to what would have been experienced under placebo.

A key element of our approach is that NI-margin adaptation is not based on observed effectiveness data from within the trial. Meta-regression parameters are estimated prior to starting the trial, and depend entirely on external efficacy data. The end-of-study M1 margin depends only on the pre-specified model for active-control arm effectiveness and observed effect modifiers at baseline and post-randomization in only the active-control arm of the NI trial. The NI margin (M2) will depend on M1 and a pre-defined amount of preserved benefit. It may be reasonable to update the meta-regression with data from external trials that conclude during the conduct of the NI trial, but the decision to update the model and the decision about which adaptive approach will be selected should be explicitly pre-specified.

Just as historical trials may show that participant characteristics can influence active-control efficacy, experimental-therapy efficacy may similarly depend on characteristics of participants in the experimental arm. In the context of an NI trial there is no historical information regarding effect modification in the experimental arm, but exploratory analyses may be possible. For example, a secondary objective of the trial might be to assess whether key baseline and post-randomization factors also modify experimental-arm effectiveness; such assessments might include subgroup analyses or tests for interaction.

Our meta-analysis regression method is an important extension of Nie and Soon's approach [13] in that it allows for the inclusion of post-randomization dynamic features, which in some settings may be the most influential effect modifiers. In trials where the active-control arm medication is controlled by the participant, the importance of medication adherence likely outweighs any known effect modifier that could be measured at baseline.

Rohmel and Kieser [14] address the idea of "variable margins" which have been proposed as way to construct more reasonable NI margins for binary end-point trials when failure rates in the active-control arm are substantially different than expected. The variable-margin approach allows the end-of-study margin to depend on observed failure rates, and when these rates are lower than a preset threshold, the margin switches from a difference-in-proportions scale to the odds-ratio scale. Although changing the scale provides some flexibility in the face of non-constancy, unlike the meta-analysis approach it does not address the question of how to define an appropriately sized margin based on observed levels of active-control effectiveness.

The choice of endpoint-assessment scale is nevertheless important. Although our development of the adaptive NI margin approach uses the relative risk scale, the same approach can be used for risk differences, differences in means, or any outcome scale. For example, by applying the log transformation to (6), so that the conserved proportion  $\rho$  becomes a multiplier on a risk difference instead of an exponent on a risk ratio. The required benefit  $\beta$  and target benefit  $\Omega$  would then become additive differences instead of multipliers. In applications where event rates are fairly high (a threshold 20% is often used in the variable-margin approach) the risk-difference scale may be more intuitive and useful.

Similarly, when comparing the mean value of continuous measures, defining the margin in terms of absolute differences may often be appropriate.

Meta-analysis based adaptive NI margins have important limitations. First, meta-analysis results are only as good as the trials upon which they are based, and it is not always the case that multiple, high-quality trials are available. Particularly for meta-analysis regression, multiple trials with accurate measures of important effect modifiers are necessary to achieve a reliable regression model. While some fields of study are rich with existing, high-quality trial data, others are not.

Critical effect modifiers such as adherence can be measured in very different ways, yielding different results; for example, self-reported drug adherence has been shown to be consistently higher than lab-based adherence measures [10, 18]. It is therefore essential that assessment of effect modifiers in the new trial be consistent with the prior studies used to construct the margin. If measurement is not consistent, regression-based estimates of active-control effectiveness are unlikely to be accurate, in which case the adapted NI margin will not result in the desired trial characteristics.

In addition, if measurement methods are not clearly pre-specified, the opportunity could arise for inappropriate manipulation of the margin. For example, if it is known that higher measured adherence corresponds to higher effectiveness, one need only choose an inflated adherence measure, such as self-report, to increase the estimated active-control effect and relax the NI margin, thus making drug approval more likely. Trial integrity will depend on the use of carefully pre-specified procedures that incorporate independently and objectively measured effect modifiers.

Relying on between-trial effect-modifier estimates will not always prevent confounding. Unmeasured differences in study populations can introduce ecological bias [7], such as might occur with gender in the PrEP example. If drug adherence were not included in model (7), it would likely appear that oral PrEP is much less effective in women than in men. This result, however, would be due primarily to the fact that in several large PrEP trials in women, adherence was very low, whereas in most trials that include men, adherence was moderate or high (Figure 1). It is therefore important to use caution when using trial-level data to evaluate effect modifiers, and if post-randomization factors are not included in the model, consider using individual participant data as proposed by Hua et al. [7].

Sampling error may also introduce bias to an adapted NI margin. In situations where it is difficult to obtain sufficiently precise estimates of  $z^T$ , the potential for bias may be higher. We would suggest assessing the potential for bias by evaluating (28) using a variety of reasonable values for  $x^T$  and  $f()$ . Ideally, the sampling plan would sample randomly among participants across study time to account for temporal trends.

Even if effect modifiers could be measured and modelled perfectly, unmeasured effect modifiers always may exist. If non-constancy results from factors that cannot be measured or have otherwise not been included in the model, the adaptive margin cannot make appropriate corrections. In the study of HIV prevention, for instance, it is not possible to measure sexual



exposure to HIV. If exposure to HIV is substantially less than expected based on prior efficacy trials, active-control effectiveness may be lower than is predicted by the model [3].

In the presented results we assumed a fixed study design, adapting the NI margin at the conclusion of the trial. In future work we will investigate the possibility of adapting NI margins based on interim analyses in group sequential trials, and appropriate ways to update sample sizes as a result of interim updates to the margin and hypotheses.

Meta-analysis regression methods offer a way to define NI margins appropriate to a specific study population, and to adapt end-of-study margins to observed characteristics of study populations. In the presence of known, measurable effect modifiers, these methods can substantially reduce the undesirable consequences of violating the assumption of constancy.

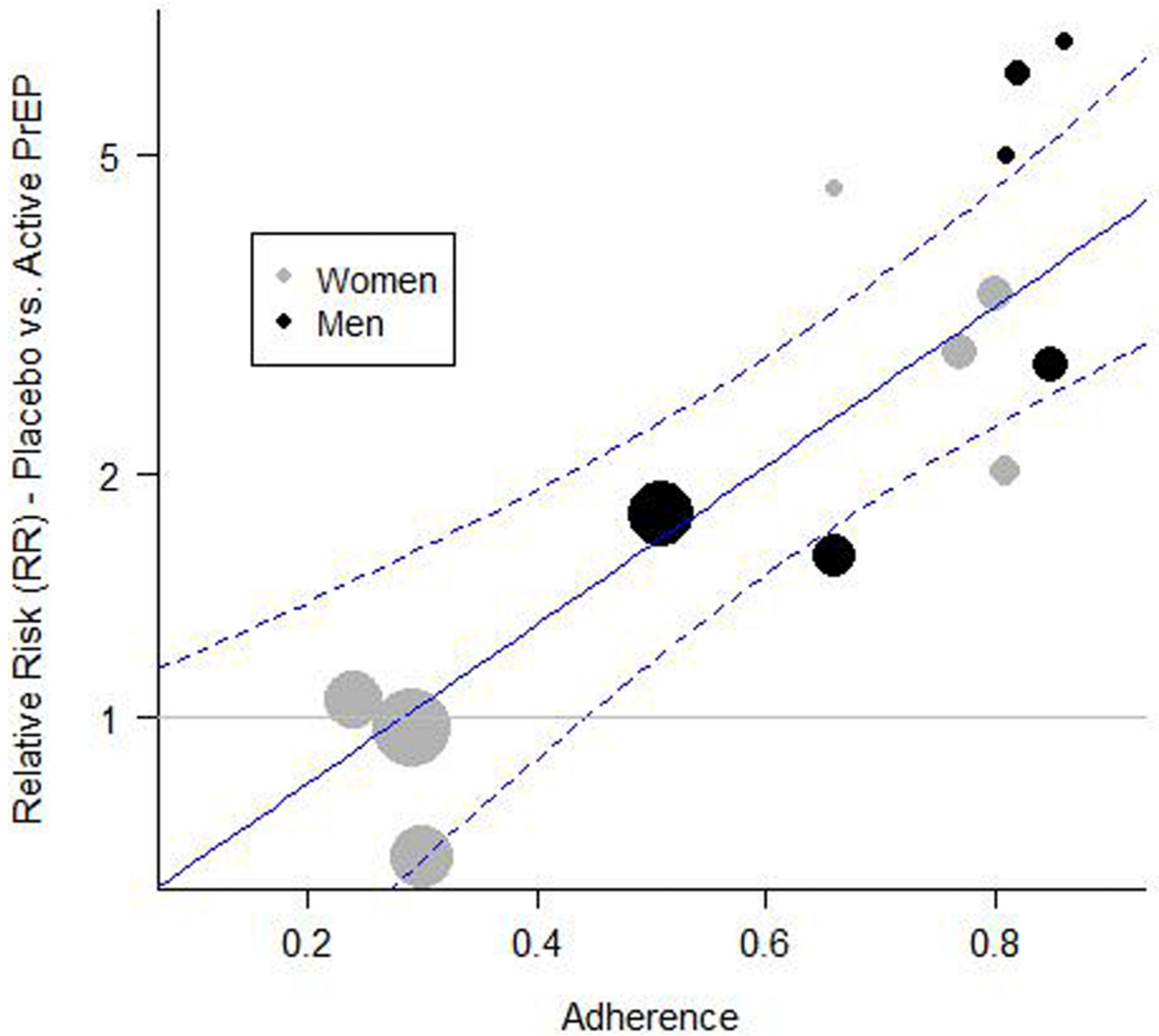
## Acknowledgement of Funding:

This work was supported by the HIV Prevention Trials Network (HPTN) and NIH grant: NIAID 5 UM1 AI068617-10.

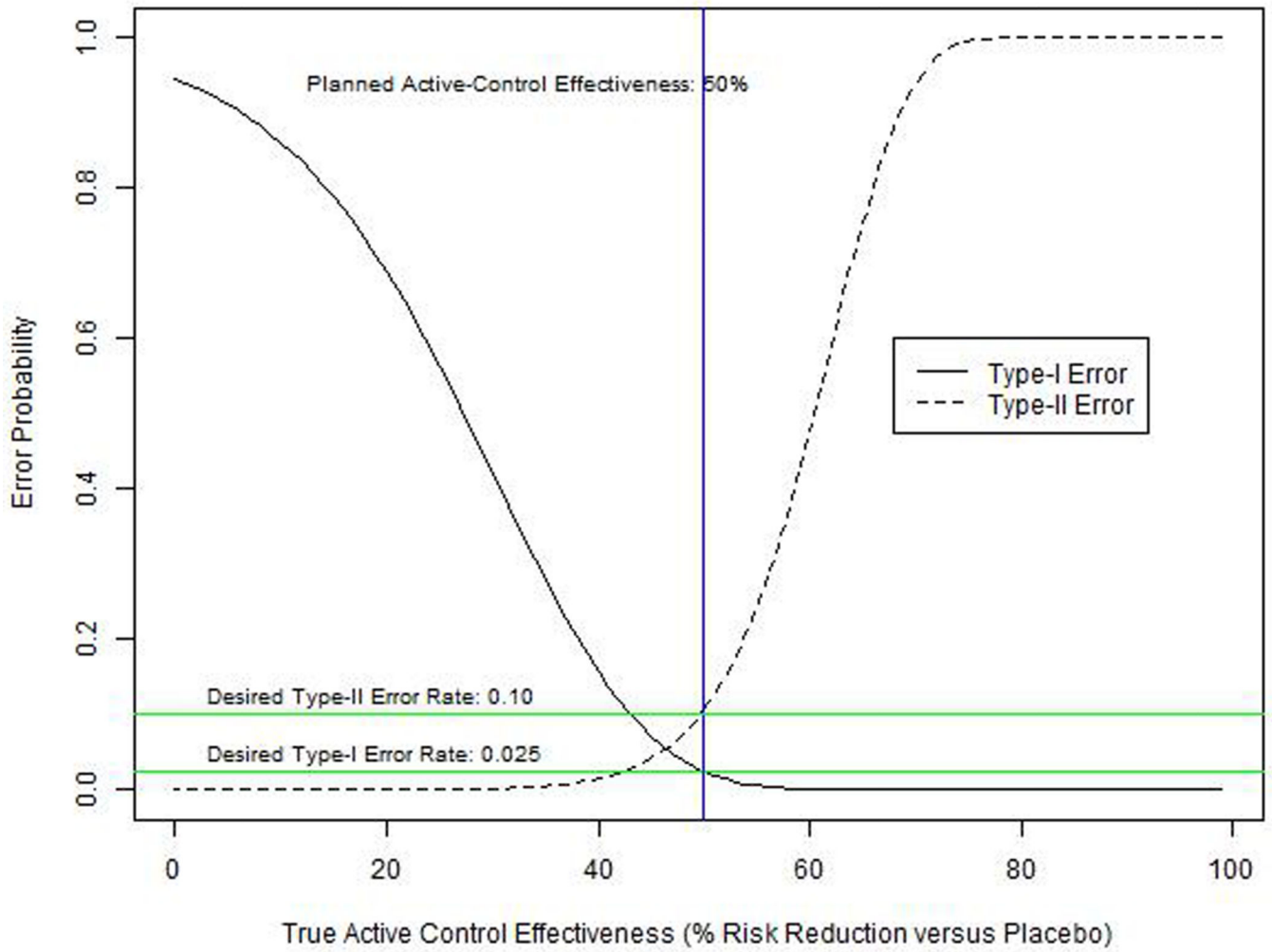
## References

- [1]. Baeten JM, Donnell D and Ndase P (2012) Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med* 367: 399–410. [PubMed: 22784037]
- [2]. Choopanya K, Martin M and Suntharasamai P (2013) Antiretroviral prophylaxis for HIV infection in injecting drug users in bangkok, thailand (the bangkok tenofovir study): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* 381: 2083–90. [PubMed: 23769234]
- [3]. Coley RY and Brown ER (2016) Estimating effectiveness in HIV prevention trials with bayesian hierarchical compound poisson frailty mode. *Statistics in Medicine* 35: 2609–2634. [PubMed: 26869051]
- [4]. Everson-Stewart S (2010) Non-Inferiority Clinical Trials: BioCreep and a Flexible Margin Approach for Addressing Non-constancy. PhD Thesis, University of Washington.
- [5]. FDA (2010) Guidance for industry: Non-inferiority trials. Technical report, FDA.
- [6]. Grant R, Lama J, Anderson P and McMahan V (2010) Preexposure chemorophylaxis for HIV prevention in men who have sex with men. *N End J Med* 363: 2587–99.
- [7]. Hua H, Burke DL, Crowther MJ, Ensor J, Smith CT and Riley RD (2016) One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. *Statistics in Medicine* 36: 772–789. [PubMed: 27910122]
- [8]. ICH (2017) ICH harmonized guideline: Estimands and sensitivity analysis in clinical trials, E9(R1). Technical report, International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use.
- [9]. Koopmeiners JS and Hobbs BP (2016) Detecting and accounting for violations of the constancy assumption in non-inferiority clinical trials. *Statistics Methods in Medical Research* 0(1): 1.
- [10]. Marrazzo J, Ramjee G and Richardson B (2015) Tenofovir-based preexposure prophylaxis for HIV infection among african women. *N End J Med* 372: 509–18.
- [11]. Molina L, Capitant C, Spire B, Pialoux G and Cotte L (2015) On-depand preexposure prophylaxis in men at high risk for HIV-1 infection. *N End J Med* 373: 2237–46.
- [12]. Nie L and Soon G (2010) An adaptive noninferiority margin and sample size adjustment in covariate-adjustment regression model approach to nininferiority clinical trials. *Model Assisted Statistics and Applications* 5: 169–177.
- [13]. Nie L and Soon G (2010) A covariate-adjustment regression model approach to noninferiority margin definition. *Statistics in Medicine* 29: 1107–1113. [PubMed: 20209669]

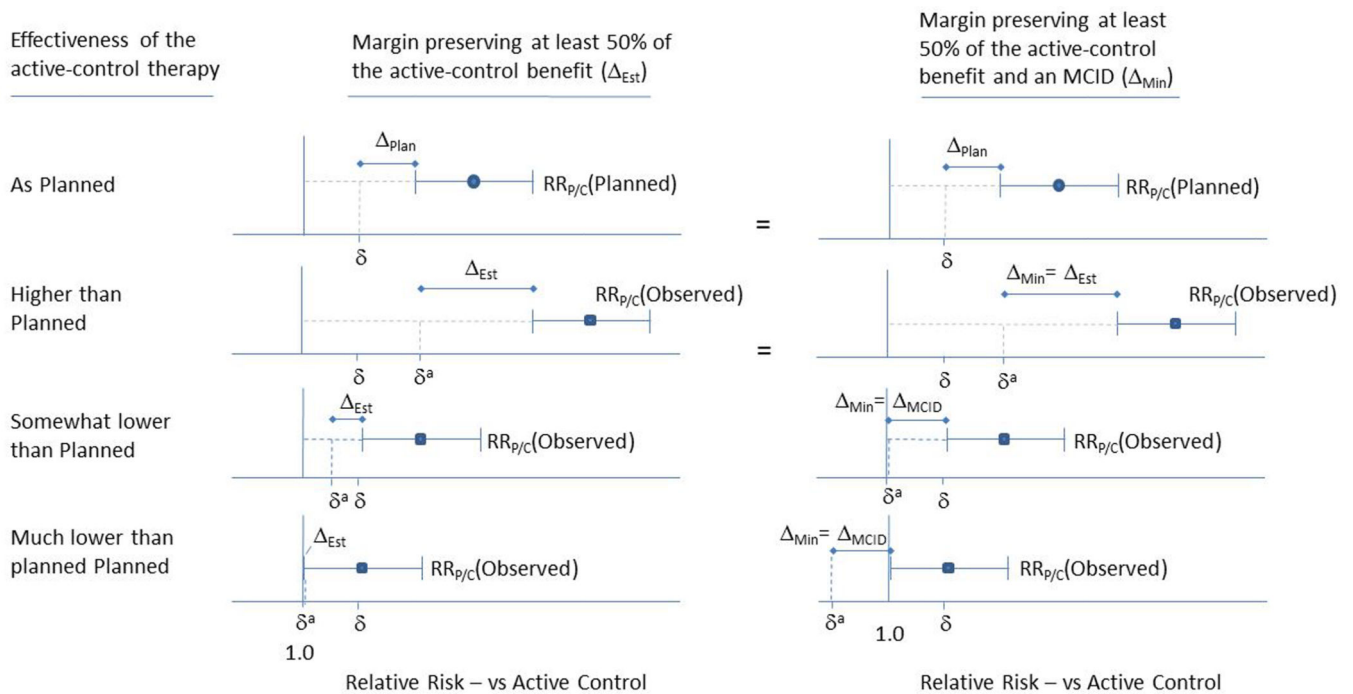
- [14]. Rohmel J and Kieser M (2012) Investigations on non-inferiority - the Food and Drug Administration draft guidance on treatments for nosocomial pneumonia as a case for exact tests for binomial proportions. *Statistics in Medicine* 32(14): 2335–2348. [PubMed: 22991269]
- [15]. Sutton A, Abrams K, Jones D, Sheldon T and Song F (2000) *Methods for meta-analysis in medical research*. John Wiley and Sons.
- [16]. Thigpen M, Kebaabetswe P and Paxton L (2012) Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in botswana. *N Engl J Med* 367: 423–34. [PubMed: 22784038]
- [17]. Thompson S and Higgins J (2002) How should meta-regression analysis be undertaken and interpreted? *Statistics in Medicine* 21(11): 1559–1573. [PubMed: 12111920]
- [18]. Van Damme L, Corneli A and Ahmed K (2012) Preexposure prophylaxis for HIV infection among african women. *N End J Med* 367: 411–22.



**Figure 1.** PrEP effectiveness plotted against trial-level adherence, as measured by estimated proportion of active-arm participants with detectable plasma TDF, for all randomized trials of oral PrEP versus placebo where an objective adherence measure was available. Circle sizes are proportional to the number of incident HIV infections during the trial. The fitted regression line and 95% confidence bounds (dashed lines) are shown for men.



**Figure 2.** Type I and type II error probabilities according to the true level of effectiveness (% risk reduction vs. placebo) in the active-control arm. Rates are for a hypothetical NI trial with NI-margin=1.3, effect size versus active-control=0.7, sample size = 110 events, and planned active-control effectiveness=50%.



**Figure 3.** Adaptive non-inferiority margins preserving proportional (50%) benefit ( $\Delta_{Est}$ , Column 1), and preserving at least the MCID ( $\Delta_{Min} = \Delta_{Min}(\Delta_{Est}, MCID)$ , Column 2), depending on the planned and observed effectiveness of the active control therapy. Point estimates for active control effectiveness (quantified by the relative risk comparing placebo to active control,  $RR_{p/c}$ ), and 95% confidence interval bars are derived from a meta-regression model. When the observed level of active-control effectiveness is as planned or higher,  $\delta$  is the same for both methods. (We assume that  $\Delta_{Plan} \leq MCID$ , i.e., that a trial would not be planned with a  $\Delta_{Plan}$  that was less substantial than the MCID.) When the observed level of active-control effectiveness is less than planned, the adapted margin  $\delta_a$  preserving the MCID is less than (more restrictive) the adapted margin  $\delta$  preserving proportional benefit.

**Table 1.**

Predicted oral-PrEP effectiveness in men (based on the lower confidence limits in Figure 1) for different assumed adherence, and suggested NI margins that preserve at least 50% of the benefit ( $\rho = 0.5$ ).

Adherence	Effectiveness (M1)	NI Margin (M2)
0.40	<1.0	1.0
0.50	1.17	1.08
0.60	1.50	1.23
0.70	1.89	1.37
0.80	2.30	1.52

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2.**

Planned and adaptive hypotheses, effect sizes, and power for varying levels of estimated active-control effectiveness in an example trial comparing and experimental HIV PrEP agent to an active control (oral HIV PrEP). Four methods of computing the required benefit over placebo ( $\Omega$ ) are shown, including (1)  $\Omega_{Plan}$  which is defined to preserve 50% of the active-control benefit at the planned level of effectiveness, (2)  $\Omega_{Est}$  which preserves 50% of the estimated active-control benefit at the observed level of effectiveness, (3)  $\Omega_{Min}$  which preserves both 50% of the estimated benefit at the observed level of effectiveness and an MCID (defined here as 0.90), and (4)  $\Omega_{Max}$  which preserves 50% of the estimated benefit at the observed level of effectiveness and places a maximum on the NI margin. Also shown are two methods for specifying the alternative hypothesis ( $\xi$ ), the first fixing  $\Omega$  based on the pre-planned alternative  $\xi=0.80$ , and the second method holding  $\xi$  constant at 0.80. Bolded values are fixed by design and determine the adaptive margins, hypotheses, and effect sizes. All values are relative risks except  $\rho$  and power. The pre-planned sample size is 231 HIV-infection events.

	Active-control effectiveness $\dagger$	Assured active-control benefit - $LCL_{95}(RR_{p/c})$ $\ddagger$	Required benefit over Placebo - $\Omega$	Adaptive Null / NI margin - $\delta^a$	Proportion of Benefit Preserved - $\rho$	Fixed target benefit over placebo				Fixed target benefit over active control		
						Fixed/planned target benefit over placebo - $\Omega$	Adaptive* Alternative - $\xi^a$	Effect size $\xi^a/\delta^a$	Power	Effective target benefit over placebo - $\Omega$	Fixed alternative - $\xi$	Effect size $\xi/\delta^a$
Preserve planned benefit = $\Omega_{Plan}$	Higher than planned	1.89	<b>0.82</b>	1.54	0.32	<b>0.53</b>	1.00	0.65	0.90	0.42	<b>0.80</b>	0.52
	As planned	1.50	0.82	1.23	<b>0.50</b>	0.53	<b>0.80</b>	0.65	0.90	0.53	<b>0.80</b>	0.65
	Lower than planned	1.17	<b>0.82</b>	$0.95\phi$	<b>1.33</b>	<b>0.53</b>	0.62	0.65	0.90	0.69	<b>0.80</b>	0.84
Preserve proportional benefit = $\Omega_{Est}$	Higher than planned	1.89	0.73	1.37	<b>0.50</b>	<b>0.53</b>	1.00	0.73	0.66	0.42	<b>0.80</b>	0.58
	As planned	1.50	0.82	1.23	<b>0.50</b>	0.53	<b>0.80</b>	0.65	0.90	0.53	<b>0.80</b>	0.65
	Lower than planned	1.17	0.93	1.08	<b>0.50</b>	<b>0.53</b>	0.62	0.57	0.99	0.69	<b>0.80</b>	0.74
Preserve proportional benefit and MCID = $\min(\Omega_{MCID}, \Omega_{Est}), MCID=0.90$	Higher than planned	1.89	0.73	1.37	<b>0.50</b>	<b>0.53</b>	1.00	0.73	0.66	0.42	<b>0.80</b>	0.58
	As planned	1.50	0.82	1.23	<b>0.50</b>	0.53	<b>0.80</b>	0.65	0.90	0.53	<b>0.80</b>	0.65
	Lower than planned	1.17	<b>0.90</b>	1.05	0.69	<b>0.53</b>	0.62	0.59	0.98	0.69	<b>0.80</b>	0.76
Preserve proportional benefit and limit the maximum margin = $\max(\min(1.23/LCL_{95}(RR_{p/c}), \Omega_{Est}))$	Higher than planned	1.89	0.65	<b>1.23</b>	0.67	<b>0.53</b>	1.00	0.82	0.34	0.42	<b>0.80</b>	0.65
	As planned	1.50	0.82	<b>1.23</b>	0.50	0.53	<b>0.80</b>	0.65	0.90	0.53	<b>0.80</b>	0.65
	Lower than planned	1.17	<b>0.93</b>	1.08	0.50	<b>0.53</b>	0.62	0.57	0.99	0.69	<b>0.80</b>	0.74

$\dagger$ Planned effectiveness is based on 60% adherence, higher than planned is based on 70% adherence, and lower than planned is based on 50% adherence.

<sup>‡</sup>The "assured benefit" is the Lower Confidence Limit(LCL) of the 95% confidence interval surrounding the relative risk (RR) of HIV infection comparing placebo to active-control (oral PrEP), as estimated by the meta-regression model for oral PrEP effectiveness as a function of drug adherence and sex.

<sup>\*</sup>When adherence is as planned, the alternative is also as planned (fixed at 0.80).

<sup>ϕ</sup>A margin less than one indicates that super superiority is required. In this example, in order to maintain the pre-planned benefit over placebo, the experimental therapy must be at least 5% better than the active control.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript