

# Simultaneous False-Decision Error Rates in Master Protocols with Shared Control: False Discovery Rate Perspective

JINGJING YE\*, XIAOYUN (NICOLE) LI, CHENGXING LU, AND WILLIAM WANG

---

## Abstract

Master protocol is a type of trial designs where multiple therapies and/or multiple disease populations can be investigated in the same trial. A shared control can be used for multiple therapies to gain operational efficiency and gain attraction to patients. To balance between controlling for false positive rate and having adequate power for detecting true signals, the impact of False Discovery Rate (FDR) is evaluated when multiple investigational drugs are studied in the master protocol. With the shared control group, the “random high” or “random low” in the control group can potentially impact all hypotheses testing that compare each of the test regimens and the control group in terms of probability of having at least one positive hypothesis outcome, or multiple positive outcomes. When regulatory agencies make the decision of approving or declining one or more regimens based on the master protocol design, this introduces a different type of error: simultaneous false-decision error. In this manuscript, we examine in detail the derivations and properties of the simultaneous false-decision error in the master protocol with shared control under the framework of FDR. The simultaneous false-decision error consists of two parts: simultaneous false-discovery rate (SFDR) and simultaneous false non-discovery rate (SFNR). Based on our analytical evaluation and simulations, the magnitude of SFDR and SFNR inflation is small. Therefore, the multiple error rate controls are generally adequate, further adjustment to a pre-specified level on SFDR or SFNR or reduce the alpha allocated to each individual treatment comparison to the shared control is deemed unnecessary.

KEYWORDS AND PHRASES: Shared control, False-discovery rate, Platform trial, Simultaneous decision error, Master protocol.

---

## 1. INTRODUCTION: MASTER PROTOCOLS AND MULTIPLICITY ISSUES

Master protocols have received growing interests in drug development in recent years. Master protocols, classified as basket trials, umbrella trials or platform trials, refer to a type of trial designs that test multiple therapies, either individually or in combination, and/or in multiple disease populations in parallel under a single overarching protocol, without a need to develop individual protocols for every sub-study [6]. Master protocols can potentially address the clinical development challenges of determining which therapy or combinations could get the most robust response in diseases that have heterogeneous patient populations or where multiple therapeutic approaches exist. One type of master protocols is platform trial with shared control, that compares multiple treatments simultaneously with a shared control arm. This contrasts with the traditional designs where candidate compounds are developed separately in separate studies; and each of them has its own control arm. These traditional approaches have become expensive over the years. Platform trials with a shared control arm can study multiple candidate compounds simultaneously, leading to reduction of the

total sample size [18]. One should note that although there is no uniform definition of platform trials in literature, in our paper, we use the definition based on the recommendation of American Statistical Association (ASA) Biopharmaceutical Section (BIOP) Oncology Scientific Working Group (SWG) master protocol sub-team. Platform trial is defined as a master protocol designed to incorporate design features of both the basket and umbrella trials, or with focus on the perpetual manner of a basket or/and umbrella trials [9]. Based on this definition, master protocols with shared control can be considered as platform trials with shared control.

Despite the advantages, platform trials with shared control present some unique statistical and regulatory challenges from the design perspective. Due to the potentially large number of patient cohorts, treatment regimens or sub-studies, it requires more careful assessments of multiplicity, i.e, the control of Type I error for statistical and regulatory decision making. The traditional type I error of interest, namely the family-wise type I error (FWER), is defined as the probability of claiming at least one treatment positive assuming none is active. When multiple independent two-arm trials are conducted comparing each investigational regimen to its control in traditional designs, no multiplicity adjustment is necessary if independent hypotheses are tested

\*Corresponding author: <https://orcid.org/0000-0003-0086-2228>.

[5, 13]. However, in platform trials with shared controls, multiple experimental treatment regimens are integrated into one trial. With the use of a shared control arm, test statistics are positively correlated. [8] and [1] showed that the family-wise error rate (FWER), e.g., the probability of at least one false positive finding across the arms, is smaller compared to tests in independent two-arm trials. Therefore, the traditional FWER under global null is no longer the most relevant measure in platform trial with a shared control [3].

As Fleming [7] eloquently pointed out, clinical trial results are subject to the sampling context, the phenomenon of “random high” bias can lead to overestimation or underestimation of outcomes. With the shared control group, the “random high” or “random low” in the control group can potentially result in inflated decision errors for all hypotheses that compare each of the test regimens and the control group. In regulatory decision context, Collignon et al (2020) [5] referred it as another error, simultaneous multiple false positive regulatory decisions. The simultaneous false-positive regulatory error is different from the FWER on multiple treatments because the decision is now correlated when a shared control is in place. This can be illustrated as multiple false-positive error (MFPE) as the probability of making multiple (at least two) simultaneous false positive decisions when all null is true. In the confirmatory setting, regulators are mainly concerned on type-I error and Type-II error is the sponsor’s risk. However, both are errors that would need to be recognized and would lead to wrong decisions.

Given a shared control, the simultaneous false-decision error would happen in two scenarios: 1) if the shared control outcome is at ‘random low’, the statistical tests may then be significant for more than one treatment comparison, even if none of the investigational drugs are active. Therefore, the chance of simultaneous false-positive increases compared to that with independent two-arm trials [8]; 2) if the shared control outcome by chance is overoptimistic, the statistical tests may then not be able to reject the null hypothesis for more than one treatment comparison, even if the alternative hypothesis of treatment efficacy is true for all investigational drugs, therefore the simultaneous false-negative may increase compared to that with independent two-arm trials [5]. Similar to FWER, family-wise type II error is defined as probability of at least one treatment is negative when all are active. The different kind of error, multiple false negative error (MFNE) is then defined as the probability of making multiple (at least two) simultaneous false negative decisions when all alternative is true.

In recognizing the new type of error introduced because of the shared control design, the goal of this paper is to characterize and quantify such error under the framework of false discovery rate. Collignon et al (2020) [5] restricted their discussion of simultaneous decision error to the false-positive. In our paper, we expand to define the **Simultaneous False-Decision Error** both to simultaneous type-I error and simultaneous type-II error. Additionally, Collignon

et al (2020) [5] and Howard et al (2018) [8] focused the discussion on MFPE. However, in reality, some treatment arms are effective, and some are not. The platform trial with shared control can also be in exploratory phase rather than confirmatory stage. This especially became relevant during the COVID-19 pandemic. There is an urgent need to evaluate multiple drugs in one study in treating COVID-19. Platform trial with shared control has been utilized in response to the requirement as an efficient design. ACTIV-2, a study for outpatients with COVID-19, is such an example. Multiple drugs were initiated in the platform trial and patients were randomized to receive either the investigational drugs or a shared placebo in Phase II setting. Then in phase III portion, patients were randomized to receive either the investigational drug or a common active comparator [4]. The concept becomes relevant in both the exploratory and confirmatory setting to evaluate the probability that multiple ineffective treatment can be declared effective and the probability that multiple effective treatments can be declared ineffective when a shared control is used.

From a societal perspective, platform-wise (or indication-wise) error rates are of importance: what is the probability, that at least one ineffective treatment is declared effective? What is the probability that multiple ineffective treatment is declared effective in a platform trial setting? What is the expected number of ineffective treatments declared effective? Proper statistical procedures may be needed to address this to ensure scientifically valid decision making.

Alternative approach to address multiplicity is thus proposed evaluating False Discovery Rate (FDR) by [2] when multiple investigational drugs are evaluated. The control of false discoveries is particularly relevant in a multi-arm platform trial with a shared control. The false discovery rate is the expected proportion of false positives among all treatment arms declared effective [17]. For example, when conducting a phase 2 platform trial to select ‘performing’ combination therapies into confirmatory phase 3 testing, one would be interested in the more practical question: among the ones that are tested positive in the platform trial under certain pairwise type I error, how many are truly effective, and how many are ineffective. This corresponds exactly the definition of false discovery rate. By the same token, one would also be interested in the rate of true positive among the ones that were tested non-positive, this corresponds to the false negative rate or false non-discovery rate.

This manuscript expands beyond the traditional type I error control and examine specifically different errors in terms of false discovery rate (FDR) and false non-discovery rate (FNR) to enable scientific decision making for multi-arm platform trials. We examine in detail the derivations and properties of the simultaneous false-decision error in the platform with shared control under the framework of FDR. The simultaneous false-decision error consists of two parts: simultaneous false-discovery rate (SFDR) and simultaneous false non-discovery rate (SFNR). We quantify the magnitude of SFDR and SFNR inflation based on our analytical

evaluation and simulations when regulatory agencies make the decision of approving or declining an application based on a platform design. These evaluation and simulations show that the magnitude of SFDR and SFNR inflation is small. With such small magnitude, our recommendation is that at the study design stage, we need be aware of this type of error, and their impact on the conclusions of a platform trial. Overall, the multiple error rate controls are generally adequate, further adjustment to a pre-specified level on SFDR or SFNR or reduce the alpha allocated to each individual treatment comparison to the shared control is deemed unnecessary.

The manuscript is organized as follows: Section 2 discusses the concepts of false discovery rate and false non-discovery rate. Applying these rates in a multi-arm trial allows us to balance the control of false discoveries while maintaining the power of detecting true treatment effects. Section 3 defines simultaneous false-discovery error rates in the context of false discovery rate and false non-discovery rate, i.e., simultaneous false discover rate (SFDR) and simultaneous false non-discovery rate (SFNR). [17] explored the idea of FDR via simulation studies. In our manuscript, statistical derivations for these error rates will be provided in the situations with independent control (i.e., for independent test statistics) and with shared control (i.e., for correlated test statistics). Additionally, theoretical derivations on FNR will be provided. Section 4 will provide analytical and simulation results that further illustrate the statistical properties of these error rates. Section 5 discusses the real-world setting for possible design changes using SFDR and SFNR. Section 6 will provide summary and discussion.

## 2. FALSE DISCOVERY RATE (FDR) AND FALSE NON-DISCOVERY RATE (FNR)

In platform trials with shared control, each treatment-control pair would have a hypothesis to test its efficacy against the null hypothesis. In a confirmatory setting in phase III clinical trials, multiple hypotheses testing may be a concern when testing several hypotheses simultaneously.

### 2.1 Definition

Instead of controlling FWER, Soric (1989) [12] proposed the framework for quantifying the statistical significance of multiple hypothesis tests based on the proportion of Type I error among all hypothesis tests declared statistically significant and proposed the concept of false discovery rate (FDR), that concerns the proportion of false discoveries among the ones that are claimed statistically significant.

Table 1 summarizes the various outcomes that occur when testing  $m$  hypotheses.  $V$  is the number of type I errors (or false positive results).  $T$  is the number of type II errors (or false negative results).

In a multi-arm platform trial setting, the commonly used measure of family-wise type I error rate (FWER) can thus

Table 1. Outcomes when testing  $m$  hypotheses.

	Null is true	Alternative is true	Total
Rejected	$V$	$S$	$R$
Not rejected	$U$	$T$	$m - R$
Total	$m_0$	$m - m_0$	$m$

be defined to be  $Pr(V \geq 1)$ , while false discovery rate (FDR) is defined to be  $FDR = E\left(\frac{V}{R} \mid R > 0\right)$ .

As discussed in [15], our definition of false discovery rate (FDR) is referred as the definition of *positive false discovery rate* (pFDR). In [14], the positive false discovery rate describes the fact that we have conditioned on at least one positive finding having occurred. In practice, evidence-based decisions (e.g., regulatory approval, continuation of further clinical investment) are often made among those treatments that have shown positive results. It would be important to control the total rates of false positive decisions. If none of the positive findings has occurred, the decision on the treatment arms would likely be no plan for further development. Therefore, by conditioning on at least one positive finding is more relevant in practice in multi-arm platform trials.

Similarly, one can consider the type II error rate in a multi-arm platform trial setting as the probability of at least one alternative hypothesis tested negative. We call it family-wise type-II error rate in this manuscript. The family-wise-type II error can be defined as  $Pr(T \geq 1)$ . Similar to FDR, false non-discovery rate (FNR) in [14] is defined as  $FNR = E\left(\frac{T}{m-R} \mid m-R > 0\right)$ . Of note, our definition of false non-discovery rate is referred as the definition of *positive false non-discovery rate* (pFNR) in [14]. This definition conditions on at least one negative finding having occurred and it is more relevant in the multi-arm trial setting.

### 2.2 Algorithm

When  $m$  hypotheses are tested, Storey (2011) [15] gave the following algorithm. Denote the p-values of the  $m$  tests by  $\{p_1, p_2, \dots, p_m\}$ . For a given type-I error  $\alpha$ , let  $V(\alpha) = \#\{\text{false positive } p_i : p_i \leq \alpha\}$ , assuming  $m_0$  null hypotheses, and  $R(\alpha) = \#\{p_i : p_i \leq \alpha\}$  be the total number of hypotheses that are rejected.

When  $R(\alpha) > 0$ , Storey and Tibshirani (2001) [16] suggested estimating FDR by the approximation:

$FDR(\alpha) = E\left[\frac{V(\alpha)}{R(\alpha)} \mid R(\alpha) > 0\right] \approx \frac{E[V(\alpha)]}{E[R(\alpha)]}$ . They pointed out that a simple estimate of  $E[R(\alpha)]$  is the observed  $R(\alpha)$ .

Given the above, we use conventional notations  $\alpha$  and  $\beta$  to be the type I error and type II error for each of the hypotheses tested. If we assume  $m_0$  is known and here  $R > 0$ , we have  $FDR \approx \frac{m_0\alpha}{m_0\alpha + (m-m_0)(1-\beta)}$ . Similarly, assume  $m_0$  is known and assume  $T = m - R > 0$ ,  $FNR = E\left(\frac{T}{m-R} \mid m_0, m-R > 0\right) \approx \frac{(m-m_0)\beta}{m_0(1-\alpha) + (m-m_0)\beta}$ . Therefore, when  $m_0$  is given, FDR and FNR as well as their approximations are functions of  $\alpha$  and  $\beta$ . In [15], the FDR is controlled at pre-specified level, e.g., 0.05 (two-sided), as the

number of hypothesis tests grows large. In the master protocol setting, when multiple investigational drugs compare to a shared control, the control of simultaneous false discovery rate would be equivalent to control the proportion of positive findings out of the multiple pair-wise comparisons between each investigational arm to the shared control.

### 3. SIMULTANEOUS FALSE-DECISION ERROR RATES

In this section, we will define the simultaneous false-decision error using the framework of FDR and FNR and give the derivations.

#### 3.1 Definition and Clinical Interpretation

The simultaneous false-decision error would consist of two parts: simultaneous false-positive decision error and simultaneous false-negative decision error.

The simultaneous false-positive decision error would increase when the shared control by chance is weak, where the statistical tests would reject more than one treatment comparison when no treatment effect is actually true. Using the notations in section 2.1, it can be written as  $\Pr(V > 1)$ . In the FDR framework, we define the simultaneous false-positive decision error as simultaneous false-discovery rate (SFDR) =  $E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right)$ .

Similarly, the simultaneous false-negative decision error would increase when the shared control by chance is overoptimistic, where the statistical tests would be unable to reject the null for more than one treatment comparisons when treatment efficacy is true. Using the notations in section 2.1, it can be written as  $\Pr(T > 1)$ . In the FNR framework, we define the simultaneous false-negative decision error as simultaneous false non-discovery rate (SFNR) =  $E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0\right)$ .

The Simultaneous False-Decision Error (SFDE) consists of two parts: SFDR =  $E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right)$ , and SFNR =  $E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0\right)$ , corresponding to the commonly used measure  $\Pr(V > 1)$  and  $\Pr(T > 1)$ .

#### 3.2 Derivation

##### 3.2.1 Derivation in the Case of Independent Tests

When  $m_0$  is known, using conventional notations  $\alpha$  and  $\beta$  as the type I error and type II error for each of the hypothesis testing, assuming each hypothesis is independent (i.e.,  $m$  two-arm studies with no shared control), the margins of Table 1 would be known. To derive FDR and FNR, the calculation would be:

$$FDR = E\left(\frac{V}{R} \mid R > 0, m_0\right)$$

$$\begin{aligned} &= \sum_{j=1}^{m_0} \left[ \binom{m_0}{j, m_0 - j} \sum_{k=0}^{m-m_0} \binom{j}{k+j} \binom{m-m_0}{k, m-m_0-k} \right. \\ &\quad \times \Pr(V = j, U = m_0 - j, S = k, T = m - m_0 - k) \left. \right] \\ &= \sum_{j=1}^{m_0} \left[ \binom{m_0}{j, m_0 - j} \alpha^j (1 - \alpha)^{m_0 - j} \right. \\ &\quad \times \sum_{k=0}^{m-m_0} \binom{j}{k+j} \binom{m-m_0}{k, m-m_0-k} (1 - \beta)^k * \beta^{m-m_0-k} \left. \right]. \end{aligned}$$

FNR

$$\begin{aligned} &= E\left(\frac{T}{m-R} \mid m - R > 0, m_0\right) \\ &= \sum_{j=1}^{m-m_0} \left[ \binom{m-m_0}{j, m-m_0-j} \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} \right. \\ &\quad \times \Pr(T = j, U = k, V = m_0 - k, S = m - m_0 - j) \left. \right] \\ &= \sum_{j=1}^{m-m_0} \left[ \binom{m-m_0}{j, m-m_0-j} \beta^j (1 - \beta)^{m-m_0-j} \right. \\ &\quad \times \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} (1 - \alpha)^k * \alpha^{m_0-k} \left. \right]. \end{aligned}$$

As shown in section 2.2, when  $m_0$  is known, we can approximate FDR and FNR calculation to be:  $FDR \approx \frac{m_0 \alpha}{m_0 \alpha + (m - m_0)(1 - \beta)}$  and  $FNR \approx \frac{(m - m_0) \beta}{m_0(1 - \alpha) + (m - m_0) \beta}$ .

Similarly, SFDR and SFNR can be expressed as the following:

SFDR

$$\begin{aligned} &= E\left(\frac{V}{R} * I(V > 1) \mid R > 0, m_0\right) \\ &= \sum_{j=2}^{m_0} \left[ \binom{m_0}{j, m_0 - j} \alpha^j (1 - \alpha)^{m_0 - j} \right. \\ &\quad \times \sum_{k=0}^{m-m_0} \binom{j}{k+j} \binom{m-m_0}{k, m-m_0-k} (1 - \beta)^k * \beta^{m-m_0-k} \left. \right] \end{aligned}$$

SFNR

$$\begin{aligned} &= E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0, m_0\right) \\ &= \sum_{j=2}^{m-m_0} \left[ \binom{m-m_0}{j, m-m_0-j} \beta^j (1 - \beta)^{m-m_0-j} \right. \\ &\quad \times \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} (1 - \alpha)^k * \alpha^{m_0-k} \left. \right] \end{aligned}$$

However, in reality, the number of null hypotheses  $m_0$  is unknown. In the independent trial setting,  $m_0$  follows Binomial distribution  $Bin(m, 1 - p)$ , where  $p$  is the probability that the arm is truly active. When  $p = 0$ , all arms would be inactive, and when  $p = 1$ , all arms would be active. When  $p$  is between 0 and 1, the expected number of active arms would be  $mp$ . Therefore, the  $SFDR = E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right)$  and  $SFNR = E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0\right)$  are calculated integrating all possible combinations when  $m_0 \geq 2, V > 1, T > 1$ :

$$\begin{aligned}
SFDR &= E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right) \\
&= E_{m_0} \left[ E\left(\frac{V}{R} * I(V > 1) \mid R > 0, m_0\right) \right] \\
&= \sum_{m_0=2}^m \left[ \binom{m}{m_0, m-m_0} p^{m-m_0} (1-p)^{m_0} \right. \\
&\quad \times \sum_{j=2}^{m_0} \binom{m_0}{j, m_0-j} \alpha^j (1-\alpha)^{m_0-j} \\
&\quad \left. \times \sum_{k=0}^{m-m_0} \binom{m-m_0}{k, m-m_0-k} \frac{j}{j+k} (1-\beta)^k \beta^{m-m_0-k} \right]
\end{aligned}$$

$$\begin{aligned}
SFNR &= E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0\right) \\
&= E_{m_0} \left[ E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0, m_0\right) \right] \\
&= \sum_{m_0=0}^{m-2} m_0 p^{m-m_0} (1-p)^{m_0} \left[ \sum_{j=2}^{m-m_0} \binom{m-m_0}{j, m-m_0-j} \right. \\
&\quad \times \beta^j * (1-\beta)^{m-m_0-j} \\
&\quad \left. \times \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} (1-\alpha)^k * \alpha^{m_0-k} \right].
\end{aligned}$$

Notice here,  $m_0$  is the number of true null hypotheses in the platform trials. In this SFDR and SFNR definition, we enumerate all possible values of  $m_0$ .

The statistical definition/interpretation of SFDR is the probability among  $m$  investigational arms that there can be **at least** two null hypotheses that are wrongly rejected. Similarly, the statistical definition/interpretation of SFNR is the probability among  $m$  investigational arms that there can be **at least** two alternative hypotheses that are wrongly not rejected. From clinical and regulatory perspectives, when we have multiple arms in a platform, SFDR further quantify the chance of wrongly approving at least two regimens among these arms, while SFNR quantifies the chance of wrongly

disapproving at least two regimens. It further calibrates the specificity and sensitivity of the multi-arm platform trial. In particular, when setting up the overall platform under a single master protocol, one of the key clinical and regulatory interests is how these error rates (SFDR and SFNR) change with the number of arms.

By enumerating all possibilities of  $m_0$ , SFDR and SFNR are calculated as expected values when simultaneous (at least two) errors can happen. The corresponding FDR and FNR definitions can easily be derived to sum from  $m_0 = 1$  instead of from 2.

As described in [5], the Simultaneous False-Decision Error (SFDE) has two parts and can be quantified separately by SFDR and SFNR. Note that the calculation of the joint probability is under the assumption that each arm is independently evaluated. This is the same as multiple independent two-arm trials. Under the platform trial setting, Bai et al (2020) [1] and Howard et al (2018) [8] discussed that with common control, the test statistics to evaluate treatment arms are positively correlated. We discuss the corresponding derivation incorporating correlations in section 3.2.2.

### 3.2.2 Derivation in the Case of Correlated Tests

As discussed above, the calculation of the joint probability in section 3.2.1 is under the assumption that each arm is evaluated independently in separate two-arm studies to support its claim of efficacy. Here we extend it to multi-arm studies with  $m$  treatment arms and one shared control arm. Let  $Z_i$  denote the standardized test statistic comparing each treatment arm  $i$  against the shared control arm in a multi-arm trial,  $i = 1, \dots, m$ . Assume a multivariate normal distribution, if treatment arm  $i$  does not have treatment effect under the null hypothesis,  $Z_i \sim N(0, 1)$ . On the other hand, denote  $\delta$  as the unified treatment benefit under the alternative hypothesis where  $\delta > 0$ . If treatment arm  $i$  has treatment effect,  $Z_i \sim N(\delta, 1)$ . Here  $\delta$  can be derived based on the desired  $\alpha$  and  $\beta$ . For simplicity, we assume the sample size in each treatment arm is same, as  $n_1$ , and the sample size in the shared control arm is  $n_0$ . The covariance between the two test statistics of different active arms can be calculated as  $cov_{ij} = \frac{n_1}{n_1+n_0}$ ,  $i, j = 1, 2, \dots, m$ . With equal allocation between treatment and shared control arms ( $n_1 = n_0$ ), the covariance is 0.5. In the illustrations below, we assume  $\alpha$ ,  $\beta$ , and  $\delta$  are given. Depending on the randomization ratio, sample size  $n_0$  and  $n_1$  would be different.

To derive  $E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right)$ , since there is no ordering of the treatment arms, from Table 2 a), without loss of generality, the joint distribution can be written as first  $j$  test statistics follow  $N(0, 1)$  with the next  $m_0 - j$  test statistics again follows  $N(0, 1)$ , the next  $k$  test statistics following  $N(\delta, 1)$ , the next  $m - m_0 - k$  test statistics follows  $N(\delta, 1)$ ,  $j = 2, \dots, k$ ,  $k = 2, \dots, m$ .

Therefore,  $Z = (Z_1, \dots, Z_j, Z_{j+1}, \dots, Z_{m_0}, Z_{m_0+1}, \dots, Z_{m_0+k}, Z_{m_0+k+1}, \dots, Z_m)$  follows the  $m$ -dimensional multivariate normal distribution with mean  $\mu_m = \underbrace{(0, \dots, 0)}_j$ ,

Table 2. Break-down of the rejected and not rejected hypotheses in multi-arm platform trials.

a) with  $R = k + j$  rejected treatment arms out of  $m$

	Null is true	Alternative is true	Total
Rejected	$V = j$	$S = k$	$R = k + j$
Not rejected	$U = m_0 - j$	$T = m - m_0 - k$	$m - R = m - k - j$
Total	$m_0$	$m - m_0$	$m$

b) with  $R = m - k - j$  rejected treatment arms out of  $m$

	Null is true	Alternative is true	Total
Rejected	$V = m_0 - k$	$S = m - m_0 - j$	$R = m - k - j$
Not rejected	$U = k$	$T = j$	$m - R = k + j$
Total	$m_0$	$m - m_0$	$m$

$\underbrace{0, \dots, 0}_{m_0-j}, \underbrace{\delta, \dots, \delta}_k, \underbrace{\delta, \dots, \delta}_{m-m_0-k}$  and covariance matrix

$$\sigma_m = \begin{pmatrix} 1 & \cdots & \frac{n_1}{n_1+n_0} \\ \vdots & \ddots & \vdots \\ \frac{n_1}{n_1+n_0} & \cdots & 1 \end{pmatrix}_{m \times m}.$$

Given the  $m$ -dimensional multivariate normal distribution, the SFDR can be calculated as:

$$\begin{aligned} SFDR &= E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right) \\ &= \sum_{m_0=2}^m Pr(m_0) E\left(\frac{V}{R} * I(V > 1) \mid R > 0, m_0\right) \\ &= \sum_{m_0=2}^m \left[ \binom{m}{m_0, m-m_0} p^{m-m_0} (1-p)^{m_0} \right. \\ &\quad \times \sum_{j=2}^{m_0} \binom{m_0}{j, m_0-j} \sum_{k=0}^{m-m_0} \binom{j}{k+j} \binom{m-m_0}{k, m-m_0-k} \\ &\quad \left. \times \int_{a_k}^{u_k} f(Z_k) dz \right] \end{aligned}$$

Where the lower bound  $a_k = \underbrace{(1.96, \dots, 1.96)}_j, \underbrace{-1.96, \dots, -1.96}_k, \underbrace{1.96, \dots, 1.96}_{m-m_0-k}, \underbrace{-1.96, \dots, -1.96}_j$  and upper bound  $u_k = \underbrace{(\infty, \dots, \infty)}_j, \underbrace{1.96, \dots, 1.96}_{m_0-j}, \underbrace{\infty, \dots, \infty}_k, \underbrace{1.96, \dots, 1.96}_{m-m_0-k}$  and  $f(Z_k)$  is the pdf of the  $m$ -dimensional multivariable normal with mean  $\mu_m$  and variance-covariance matrix  $\sigma_m$ .

Similarly, for Table 2 b),  $Z = (Z_1, \dots, Z_{m_0-k}, Z_{m_0-k+1}, \dots, Z_{m-j-k}, Z_{m-j-k+1}, \dots, Z_{m-j}, Z_{m-j+1}, \dots, Z_m)$  follows the  $m$ -dimensional multivariate normal distribution

with mean  $\mu_m = \underbrace{(0, \dots, 0)}_{m_0-k}, \underbrace{0, \dots, 0}_k, \underbrace{\delta, \dots, \delta}_{m-m_0-j}, \underbrace{\delta, \dots, \delta}_j$  and covariance matrix

$$\sigma_m = \begin{pmatrix} 1 & \cdots & \frac{n_1}{n_1+n_0} \\ \vdots & \ddots & \vdots \\ \frac{n_1}{n_1+n_0} & \cdots & 1 \end{pmatrix}_{m \times m}.$$

The SFNR can be written as:

$$\begin{aligned} SFNR &= E\left(\frac{T}{m-R} * I(T > 1) \mid m-R > 0\right) \\ &= \sum_{m_0=0}^{m-2} \left[ \binom{m}{m_0, m-m_0} p^{m-m_0} (1-p)^{m_0} \right. \\ &\quad \times \sum_{j=2}^{m-m_0} \binom{m-m_0}{j, m-m_0-j} \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m-m_0}{j, m-m_0-j} \\ &\quad \left. \times \int_{a_k}^{u_k} f(Z_k) dz \right]. \end{aligned}$$

Where the lower bound  $a_k = \underbrace{(1.96, \dots, 1.96)}_{m_0-k}, \underbrace{-1.96, \dots, -1.96}_k, \underbrace{1.96, \dots, 1.96}_{m-m_0-j}, \underbrace{-1.96, \dots, -1.96}_j$  and upper bound  $u_k = \underbrace{(\infty, \dots, \infty)}_{m_0-k}, \underbrace{1.96, \dots, 1.96}_k, \underbrace{\infty, \dots, \infty}_{m-m_0-j}, \underbrace{1.96, \dots, 1.96}_j$  and  $f(Z_k)$  is the pdf of the  $m$ -dimensional

multivariable normal with mean  $\mu_m$  and variance-covariance matrix  $\sigma_m$ . Note here the SFNR calculation is where the  $k + j$  hypotheses are not rejected, while the SFDR calculation is where the  $k + j$  hypotheses are rejected. Similarly, the definition of FDR and FNR can be seen as sum from  $m_0 = 1$ .

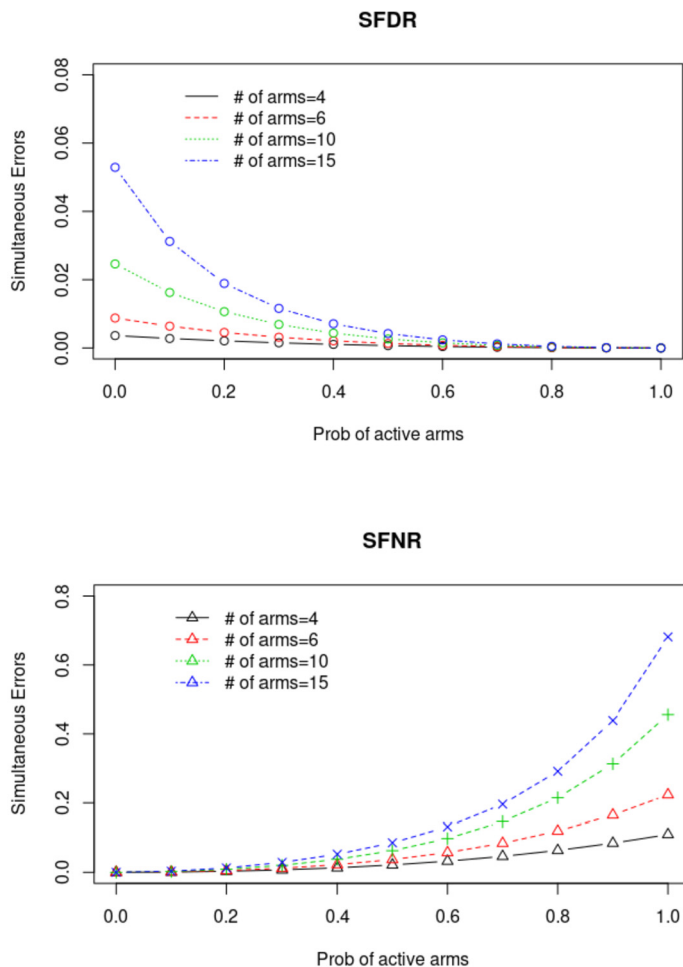


Figure 1: Plot of Simultaneous False-Discovery Rate (SFDR) (top) and Simultaneous False-Non-Discovery Rate (SFNR) (bottom) given probability of active arms in independent setting.

### 4. ANALYTICAL AND SIMULATION RESULTS

Given the derivations in section 3, we can calculate simultaneous false-decision errors (SFDE). To show the results, the number of investigational arms is generated for 4, 6, 10 and 15. We first evaluate the errors in the case of multiple independent two-arm trials as reference. Assuming each trial is tested at one-sided type-I error  $\alpha = 0.025$  and type-II error  $\beta = 0.15$ , and patients are equally randomized to treatment and control arm in its two-arm study, the SFDR and SFNR across multiple trials are shown in Figure 1 against the probability of truly active arms.

As shown on Figure 1 top plot, the simultaneous false-discovery rate (SFDR) decreases with the probability of active arms. Different colored lines show when there are 4, 6, 10, 15 arms in the platform respectively. When there are more arms in the platform, the SFDR increases. The bottom plot is the simultaneous false non-discovery rate (SFNR).

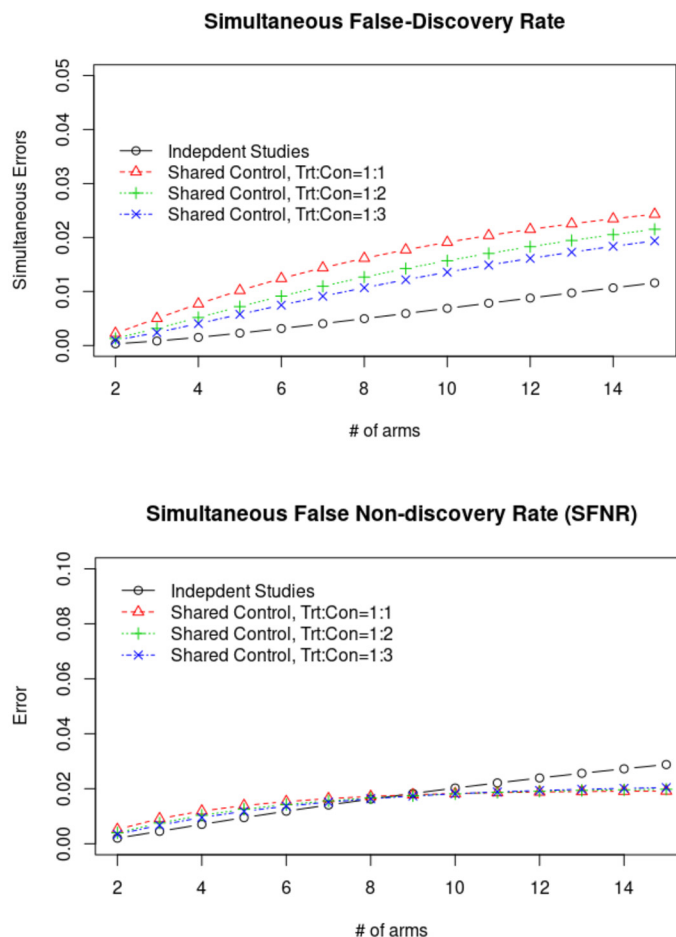


Figure 2: Equal randomization and unequal randomization for multi-arm platform trials: simultaneous false-discovery rate (SFDR) (top), simultaneous false non-discovery rate (SFNR) (bottom).

Recall that SFNR evaluates the probability among  $m$  investigational arms that there can be **at least** two alternative hypotheses that are wrongly not rejected. It increases with the probability of active arms. As more and more arms are truly active, the chance of simultaneous error will also increase. SFNR is also a lot higher when there are 15 arms than when there are only 4 investigational arms indicating more chances for simultaneous error.

Next, we evaluate SFDR and SFNR for cases in a platform trial with shared control, in which case the test statistics of the multiple hypotheses are positively correlated. We assume each hypothesis is still evaluated at the pairwise type-I error  $\alpha = 0.025$ , and its own type-II error  $\beta = 0.15$ . We use  $p = 0.3$  (30% of the arms are truly active) for the illustration. We evaluate the number of treatment arms from 2 to 15. SFDR and SFNR in these scenarios are plotted in Figure 2. The black line in Figure 2 (top) is the simultaneous false-discovery rate (SFDR) when each drug is evaluated independently (i.e., independent two-arm trial setting) to es-

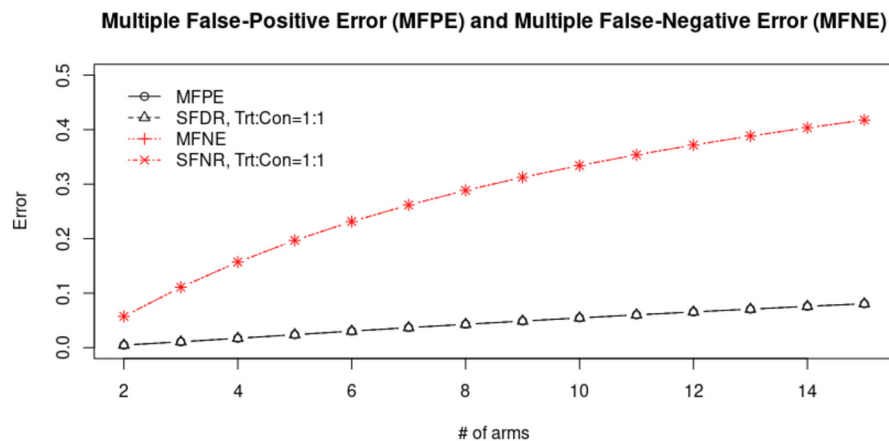


Figure 3: Multiple False-Positive Error (MFPE) and Multiple False-Negative Error (MFNE) with shared control in platform trial.

establish its efficacy and patients are equally randomized to treatment and control arms.

The red line in Figure 2 (top) plotted the case where the test statistics are correlated with equal randomization (i.e., platform trial with shared control setting). As expected, when the test statistics are correlated, the SFDR would be higher than that with independent testing statistics. Recall we define multiple false-positive error (MFPE) as the probability of making multiple (at least two) simultaneous false positive decisions when all null is true. When  $p = 0$ , the SFDR is same as MFPE. Note that unlike MFPE, SFDR is calculated for mixed scenarios, where the true number of alternative hypotheses are unknown, and the calculation is an exhaustive search when there can be at least two null hypotheses that are wrongly rejected. When all arms are inactive, the SFDR would be the same as MFPE.

The black line in Figure 2 (bottom) is the simultaneous false non-discovery rate (SFNR) when each drug is evaluated independently and patients are equally randomized to treatment and control arms. The red line is the corresponding SFNR where the test statistics are correlated with equal randomization. The SFNR is similar when the test statistics are correlated compared to that with independent test statistics and stays smaller than independent test statistics when the number of arms is larger than 8. As discussed in [1], for the case where the shared control arm is at random high, it is more likely that all tests are all negative at the same time, however, in independent setting, the probability of all control arms being random high is much smaller. similarly, the corresponding SFNR, when  $p = 1$ , would be same as MFNE (Figure 3). Unlike MFNE, SFNR is calculated as an exhaustive search when there can be at least two alternative hypotheses that are wrongly not rejected.

In addition to equal randomization, we also evaluate cases where unequal randomization ratio between each of the investigational drugs and the shared control is used. For il-

lustration here, we evaluate the situation where the ratio of each treatment arm to the control arm is 1:2, i.e., more patients would be randomized to the shared control in the multi-arm platform trials compared to each investigational drug. The test statistics of the arms are still positively correlated. Assuming the pairwise type-I error  $\alpha = 0.025$  and type-II error  $\beta = 0.15$ , the updated sample size would be 38 for each treatment arms and 75 for the shared control arm. The results are plotted in Figure 2 shown as the green line.

After increasing the ratio of shared control, even though the SFDR is still larger than when the treatment arms are evaluated independently, the difference became smaller than that with equal randomization since the correlation decreased. For SFNR, it stays similar as the case of equal randomization.

If we further increase the randomization ratio between each treatment arms and shared control to 1:3, this corresponds to 33 patients in each treatment arms and 101 in shared control arm given the same type-I, type-II errors and  $\delta$ . From Figure 2 shown as the dark blue dotted line, the SFDR is reduced to be similar as the case where the treatment arms are evaluated independently. For SFNR, it continues to be similar as the case of equal randomization and randomization ratio of 1:2, approaching the independent studies.

Additional simulation studies are conducted given various scenarios on the number of investigational arms, proportion of truly active arms and the randomization ratio between the experimental arms and the control. The number of arms in the platform is either 5 or 8 and the % of active arms are either 30% or 50% based on practical scenarios seen in current master protocol landscape. The randomization ratio between the investigational arm and the control is either 1:1 or 1:2. The goal of the simulation is not to do exhaustive simulations covering all scenarios, but to evaluate the practical scenarios that seen in current master protocol studies. The results are included in the Table 3.



Table 3. Simulation Results given different number of investigational arms in the platform, % of truly active arms and the randomization ratio between the investigational arm and the control.

Randomization Ratio (Trt: Con)		sFDR			
		Indep. Simulation	Analytical Derivation, Indep.	Shared Control Simulation	Analytical Derivation Shared control
1 : 1	$m = 5$ , % Active arm = 30%	0.0085	0.008	0.019	0.025
	$m = 8$ , % Active arm = 30%	0.016	0.019	0.031	0.037
	$m = 5$ , % Active arm = 50%	0.0036	0.0039	0.01	0.012
	$m = 8$ , % Active arm = 50%	0.0075	0.0075	0.015	0.018
1 : 2	$m = 5$ , % Active arm = 30%	0.01	0.01	0.014	0.019
	$m = 8$ , % Active arm = 30%	0.015	0.019	0.028	0.032
	$m = 5$ , % Active arm = 50%	0.003	0.004	0.008	0.009
	$m = 8$ , % Active arm = 50%	0.007	0.008	0.012	0.015

In Table 3, simulations are conducted to show the impact of sFDR when shared control vs independent control for each experimental arm is used. For each of the scenario, a total of 10000 simulations are run. The sample size for each experimental arm is 50 and the control arm is 50 when 1:1 equal randomization is used. The control arm is 100 when 1:2 randomization ratio is used. Endpoint of interest is a continuous endpoint assuming following a normal distribution. Data from  $H_0$  follow a normal distribution  $N(0, 1)$  and data from  $H_a$  follow a normal distribution  $N(3, 1)$  corresponding to 85% power and one-sided 2.5% type-I error rate. Two-sample t-test is conducted for each experimental arm vs. control arm at 5% level (two-sided).

Based on the simulations, the analytical derivations in section 3 generally match with the simulation scenarios. As shown in the table, the derivation may be slightly higher than the simulations, thus reflecting being more conservative. This may be due to the limited sample size in each investigational arm. Here we only report SFDR because it is more interest to the regulatory agencies. Similar results in SFNR are seen.

## 5. REAL-WORLD SETTING CONSIDERATIONS

In practice, a platform trial can be implemented solely by one organization or a collaboration with various organizations. When the platform trial is solely sponsored by one organization with a common control arm in a particular disease/indication, the objective is usually to study drugs' efficacy and safety either as monotherapy or in combination in one overarching framework to advance one or more arms into the next phase development.

For the case when the platform trial is in collaboration between organizations, the decisions are likely made for each investigational arm independently for each sponsor. In this situation, due to "random high" or "random low" in the common control arm, SFDR may get inflated comparing to the case with independent two-arm studies, Furthermore,

even if multiple treatment arms are sponsored by different organizations, the overall SFDR and SFNR across these treatment arms may have public health impacts and could be evaluated in that context. As introduced in the introduction section, the ACTIV-2 COVID-19 platform trial with shared control is one such example.

For SFNR, it stays close to independent cases and crosses with independent cases when the number of arms is around 8. Based on the results, design changes do not deem necessary. However, if the goal is to reduce the error close to independent cases as shown by SFDR and SFNR, two options can be implemented: 1) by randomizing more patients to the common control arm, for example, 1:3 ratio of investigational arm to common control can shrink the simultaneous error close to independent two-arm studies; 2) if equal randomization is used, making more stringent type-I error for individual arms will reduce the simultaneous false-positive error. Either option can offer a better chance to make the right conclusion out of the trial but would reduce the original efficiency of using the platform trial framework. Trade-offs between the two options can be explored given the number of planned arms in the platform trial in addition to other factors typically considered in drug development. Table 4 illustrates the comparison of SFDR and SFNR to the independent two-arm trials when either type-I error on individual arm or randomization ratio changes at design stage. From the table and the plots, unequal randomization offers the chance of approaching both simultaneous errors to independent cases. However, if the goal of the development is to randomize more patients into investigational arms to learn, keeping equal randomization while increasing individual type-I error rate to be 0.05 would reduce the SFNR smaller than the independent studies, therefore a great chance to advance the right arms to further development.

## 6. DISCUSSION

Master protocol trials with a shared control, referred in this paper as platform trials with a shared control, has re-

Table 4. Comparison of SFDR and SFNR given design changes in the multi-arm platform trials, assuming 30% arms are active.

# of Arms	1:1 Randomization Ratio and Type-I error = 0.0125		1:1 Randomization Ratio and Type-I error = 0.05		1:3 Randomization Ratio and Type-I error = 0.025		1:1 Randomization and Type-I error = 0.025		Independent Two-Arm Trial*	
	SFDR	SFNR	SFDR	SFNR	SFDR	SFNR	SFDR	SFNR	SFDR	SFNR
5	0.0042	0.016	0.025	0.0011	0.006	0.012	0.01	0.014	0.002	0.009
10	0.0084	0.023	0.043	0.0012	0.014	0.018	0.019	0.018	0.007	0.02
15	0.011	0.026	0.052	0.0012	0.019	0.02	0.024	0.019	0.012	0.029

\*For independent two-arm trial, each experimental arm and control is randomized 1:1 and type-I error rate for each arm is 0.025.

cently gained increasing interest in clinical drug development. Extensive discussions have taken place on type I error considerations for platform trial with a shared control. A consensus achieved from these discussions is that, with the ‘random high’ and ‘random low’ performances from the shared control and due to the fact the same control has been used as the comparator in comparisons for multiple treatments, the chances of SFDE (some also refer it as ‘clustered errors’ to be brief) could increase [5]. This paper conducted detailed evaluations of different errors by formulating the SFDE under the FDR framework and presenting numerical results to demonstrate the factors impacting SFDE, the magnitude of the error and some recommendations to potentially reduce the error.

The reason FDR-based framework is concerned in this research is that the concept of FDR in the platform trial context is focusing on the question that out of the positive arms identified by the platform trial, how many of them are not truly effective and on the flip side, how many of them are. The relationship between type I error, power and FDR is similar to sensitivity (analogous to 1 minus type I error), specificity (analogous to power) and positive predictive value (analogous to 1 minus FDR) where the latter are terminologies widely used in the field of diagnosis tests. Therefore, unlike type I and type II errors that only focus on  $\alpha$  or  $\beta$  alone, the concept of FDR is a function of  $\alpha$  and  $\beta$  with the consideration of proportion of truly active arms among total number of arms being investigated (analogous to the prevalence in the diagnostics field). This is also seen in the mathematical formula derived for SFDR and SFNR in section 3 and the analytical results. Therefore, this may provide a different angle to the multiple simultaneous errors introduced by the common shared control.

As indicated by the analytical results, in platform trial design, SFDR is larger than in the scenario when separate two-arm studies are conducted. This trend is consistent with multiple type I error (or Multiple False Positive Error in [8]) in the master protocol design comparing to the design with multiple independent studies. The proposed methods can readily be applied in the signal finding studies. These studies not only drive further development of investigational compounds but also have impacts on regulatory decisions. Although not strict adjustment is deemed as necessary, to un-

derstand what level of type-I error for each treatment-shared control needs to match to the level of conventional independent two-arm trials, evaluation is conducted and presented in Appendix B. Except when there are 3 arms in the platform, milder multiplicity adjustments than Bonferroni can be made to each arm-shared control comparison. If more patients are allocated to control than randomization ratio of 1:1, even milder adjustment is shown between treatment and shared control. Notice here, as shown in Table 4, the relative SFDR inflation shrinks as the number of arms increase in the platform, i.e., 5-fold inflation when there are 5-arm in the platform versus 2-fold inflation when there are 15-arm. Nonetheless, the absolute magnitude of SFDR inflation is still small. Our exercise in the Appendix B indicates the level of adjustment relates to the relative SFDR inflation rather than the absolute values of inflation. Again, we emphasize that the adjustments provided in the Appendix B are demonstrations of the magnitude of adjustment if performed but not necessarily recommending adjustments like such in real practices.

Without loss of generality, in the derivation of SFDR and SFNR, the probability of truly active arms is assumed to be  $p$ , the same for each arm. In reality, the probability  $p$  may be different given different MOAs (mechanism of actions) of the arms. As discussed in the Appendix B, the number of arms in the platform rather than the proportion of truly active arms impacts the SFDR. Based on biological knowledge, certain MOAs may be more promising than others. In the case of combination therapy, the arms with combination drugs may offer better chance of being effective than arms with monotherapy. In both cases, the probability  $p$  may be higher in certain MOA or combination than other arms. In understanding the SFDR and SFNR when different probabilities of success can exist, we can use the minimum and maximum probabilities to quantify the range of simultaneous errors. Alternatively, if we understand the rough distribution of probabilities of treatment success, the expectations on the probabilities can be implemented to estimate the SFDR and SFNR. For more detailed derivation based on mixed scenarios of success given the treatment arms in the platform trial, this can be explored as future research.

It is worth noting that [11] derived a step-up procedure on k-FWER and a less constrained k-FDR procedure. In

[11], the  $k$ -FWER is defined as  $k - FWER = \Pr(V \geq k)$ , while  $k - FDR = E(k - FDP)$ , where

$$k - FDP = \begin{cases} 0, & \text{otherwise} \\ \frac{V}{R}, & \text{if } V \geq k \end{cases}$$

Sarkar (2007) [11] further derived the procedure using the upper bound to control the  $k$ -FDP at certain level rather than deriving a closed form. The objective of [11] is to generalize the procedure of Benjamini-Hochberg (BH) and establish the critical value for rejection and control the  $k$ -FDP at  $\alpha$ . It should be noted that he derived the procedure under the assumption that individual hypotheses are independent or weakly dependent; while our attempt is to quantify the error understanding that the test statistics are positively correlated. Moreover, Sarkar (2007) [11] focused on the FDR, while our goal is to quantify the error in both directions (false positives and false negatives). Additionally, we examine the error in relationship to the randomization ratio in the platform trial design because the correlation is closely tied to the randomization ratio between the arms and the common control.

There is now an increasing trend in industry as a single sponsor to use platform trial as proof-of-concept studies. The simultaneous decision error concept is readily applicable to exploratory platform trials. The decision for each sponsor may include moving forward or stopping any investigational arm at the end of phase 2 in the platform trial. The exploratory platform trials may have a common control or without. Even when the exploratory platform trial is without a common control, the simultaneous decision errors can still occur when the arm is actually inactive compared to historical control or against a benchmark.

One of the hallmarks of the platform trials is the ability for investigational drugs to enter and exit the platform at different times. For example, the platform trial may start with one investigational drug A and the control. During mid-trial, additional investigational drug B is added to the platform sharing the same control. The common control thus includes concurrent common control for drug B and non-concurrent common control for those enrolled prior to drug B entering the platform. The simultaneous decision errors can still occur; however, this will depend on the common control overlap between drug A and drug B. As reported in [10], as the overlap increases, positive correlation increases. It is the largest when the overlap is 100%, i.e., drug A and drug B share the entire concurrent common control. The SFDR and SFNR derived in our paper thus represent the maximum values since it represents all investigational arms share concurrent common control. When the overlap decreases, so does the positive correlation. The SFDR and SFNR would approach independent studies.

Through our evaluation, the magnitude of potential decision errors has been quantified through SFDR and SFNR when regulatory agencies make the decision of approving or

declining an application based on a platform trial design. With such small magnitude, our recommendation is that at the study design stage, we need be aware of this type of error, and their impact on the conclusions of a platform trial. Overall, the multiple error rate controls are generally adequate, further adjustment to a pre-specified level on SFDR or SFNR or reduce the alpha allocated to each individual treatment comparison to the shared control is deemed unnecessary.

## APPENDIX A. R CODE

The R code to calculate the simultaneous multiple testing errors is provided in Source Code 1.

## APPENDIX B. THE LEVEL OF TYPE-I ERROR ADJUSTMENT RELATES TO THE RELATIVE SFDR INFLATION

Because of shared control in platform trial, the SFDR can be larger compared to that in the conventional independent two-arm trials. To further evaluate the relationship between the SFDR and  $p$ , different proportion of active arms in the platform, we plotted the SFDR against the  $p$  in Figure 4, the proportion ranges from 0% to 95%. The black line is the SFDR given independent two-arm trials, and the red line is the corresponding SFDR in platform trial with shared control. Four panels indicate when there are 3, 5, 8 and 10 arms in the platform respectively. As can be seen from the plot, the SFDR shrinks to independent two-arm trials as the proportion of active arms increases. The difference between the platform trial with share control and independent two-arm becomes very small when the over half of the arms are truly active. When there are more investigational arms in the platform, the SFDR is also higher. This can be seen on the y-axis when SFDR is around 0.01 when there are 3 arms in the platform with shared control and increases to 0.06 when there are 10 arms in the platform. However, the relative inflations from independent two-arm trials are not much since the SFDR for independent two-arm trials also increases. Note here even though the SFDR is inflated in platform trial with shared control, the absolute magnitude is still small for all scenarios.

Although the magnitude is small, potentially multiplicity can be considered when multiple treatment arms are included in the same platform if the intention is to minimize the inflation of SFDR. We evaluate the adjusted type-I error  $\alpha$  for each treatment-shared control comparison to match the independent two-arm trials. The proportion of active arms from 0% to 100% is evaluated and commonly used  $\beta$ , e.g., 0.1, 0.15 and 0.2, are also evaluated. The results showed that the adjusted  $\alpha$ -levels only vary slightly given different proportion of active arms. For example, given  $\beta = 0.1$ , with 3-arm platform trial with shared control, the adjusted  $\alpha$ -level ranges from 0.0062–0.0067 when the proportions of

Listing 1: R code to calculate the simultaneous multiple testing errors.

```

1  #Simultaneous false-discovery rate (SFDR)
2  #m is the number of treatment arms in multi-arm trials
3  #alpha is type-I error for each treatment-control comparison
4  #beta is type-II error for each treatment-control comparison
5  #p is the probability that the arm is active
6  #multiple false-positive error
7  SFDR <- function(m,alpha,betas,p){
8    tempm <- rep(NA,m-1)
9    for(m0 in 2:m){
10     pm <- choose(m,m0)*(1-p)^m0*p^(m-m0)
11     temp2 <- rep(NA,m0-1)
12     for(j in 2:m0){
13       temp <- choose(m0,j)*alpha^j*(1-alpha)^(m0-j)
14       psum <- rep(NA,m-m0+1)
15       for(k in 0:(m-m0)){
16         psum[k+1] <- ((j/(k+j))*choose(m-m0,k)*(1-betas)^k*betas^(m-m0-k))
17       }
18       temp2[j-1] <- temp*sum(psum)
19     }
20     tempm[m0-1] <- pm*sum(temp2)
21   }
22   pFDR <- sum(tempm)
23   pFDR
24 }
25 #Simultaneous false non-discovery rate (SFNR)
26 #m is the number of treatment arms in multi-arm trials
27 #alpha is type-I error for each treatment-control comparison
28 #beta is type-II error for each treatment-control comparison
29 #p is the probability that the arm is active
30
31 SFNR <- function(m,alpha,betas,p){
32   tempm <- rep(NA,m-1)
33   for(m0 in 0:(m-2)){
34     pm <- choose(m,m0)*(1-p)^m0*p^(m-m0)
35     temp2 <- rep(NA,m-m0-1)
36     for(j in 2:(m-m0)){
37       temp <- choose(m-m0,j)*betas^j*(1-betas)^(m-m0-j)
38       psum <- rep(NA,m0+1)
39       for(k in 0:m0){
40         psum[k+1] <- ((j/(k+j))*choose(m0,k)*(1-alpha)^k*alpha^(m0-k))
41       }
42       temp2[j-1] <- temp*sum(psum)
43     }
44     tempm[m0+1] <- pm*sum(temp2)
45   }
46   pFNR <- sum(tempm)
47   pFNR
48 }
49 #Examples
50 #When there are 4 arms of treatments comparing to common control, alpha is 0.025 and beta is 0.15
51 # Simultaneous multiple false-positive error
52 mpFDRr <- SFDR(4,0.025,0.15,0.3)
53 # Simultaneous multiple false-negative error
54 mpFNRr <- SFNR(4,0.025,0.15,0.3)
55

```

```

56 #Simultaneous false discovery-rate (SFDR), correlated test statistics
57 #variance_covariance function
58 #n0 is the number of patients in control arm
59 #n1 is the number of patients in each treatment arm in the multi-arm trials
60 covvar <- function(m,n0,n1){
61   corrM <- n1/(n1+n0)
62   covM <- matrix(corrM,m,m)
63   diag(covM) <- 1
64   covM
65 }
66 library(mvtnorm)
67 #SFDR, correlated test statistics
68 #n0 is the number of patients in control arm
69 #n1 is the number of patients in each treatment arm in the multi-arm trials
70 #delta is the mean of treatment arm that alternative hypotheses are true,
71 # determined by the alpha and beta
72 #p is the probability that the arm is active
73 SFDRcor <- function(alpha,m,n0,n1,beta,p){
74   delta<-qnorm(1-alpha)+qnorm(1-beta)
75   tempm <- rep(NA,m-1)
76   for(m0 in 2:m){
77     pm <- choose(m,m0)*(1-p)^m0*p^(m-m0)
78     temp2 <- rep(NA,m0-1)
79     for(j in 2:m0){
80       temp <- choose(m0,j)
81       psum <- rep(NA,m-m0+1)
82       for(k in 0:(m-m0)){
83         muk <- c(rep(0,j),rep(0,m0-j),rep(delta,m-m0)) #mean
84         covM <- covvar(m,n0,n1) #covariance matrix
85         lk <- c(rep(qnorm(1-alpha),j),rep(qnorm(alpha),m0-j),rep(qnorm(1-alpha/2),k),
86             rep(qnorm(alpha),m-m0-k)) #lower limits
87         uk <- c(rep(Inf,j),rep(qnorm(1-alpha),m0-j),rep(Inf,k),
88             rep(qnorm(1-alpha),m-m0-k)) #upper limits
89         psum[k+1] <- ((j/(k+j))*choose(m-m0,k)*pmvnorm(lower=lk, upper=uk, mean=muk, corr=covM)[1])
90       }
91       temp2[j-1] <- temp*sum(psum)
92     }
93
94     tempm[m0-1] <- pm*sum(temp2)
95   }
96   pFDR <- sum(tempm)
97   return(pFDR)
98 }
99 #Simultaneous false non-discovery rate (SFNR), correlated test statistics
100 #multiple false-negative error with correlation
101 #SFNR with correlation
102 SFNRcor <- function(alpha,m,n0,n1,beta,p){
103   delta<-qnorm(1-alpha)+qnorm(1-beta)
104   tempm <- rep(NA,m-1)
105   for(m0 in 0:(m-2)){
106     pm <- choose(m,m0)*(1-p)^m0*p^(m-m0)
107     temp2 <- rep(NA,m-m0-1)
108     for(j in 2:(m-m0)){
109       temp <- choose(m-m0,j)
110       psum <- rep(NA,m0+1)
111       for(k in 0:m0){
112         muk <- c(rep(0,m0-k),rep(0,k),rep(delta,m-m0)) #mean
113         covM <- covvar(m,n0,n1) #covariance matrix
114         lk <- c(rep(qnorm(1-alpha),m0-k),rep(qnorm(alpha),k),rep(qnorm(1-alpha),m-m0-j),
115             rep(qnorm(alpha),j)) #lower limits
116         uk <- c(rep(Inf,m0-k),rep(qnorm(1-alpha),k),rep(Inf,m-m0-j),

```

```

118   psum[k+1] <- (j/(k+j))*choose(m0,k)*pmvnorm(lower=lk, upper=uk, mean=muk, corr=covM)[1]
119   }
120   temp2[j-1] <- temp*sum(psum)
121   }
122   tempm[m0+1] <- pm*sum(temp2)
123   }
124   pFNR <- sum(tempm)
125   pFNR
126 }
127 #Example
128 #correlated test statistics, SFDR
129 mpFDRcorr <- SFDRcor (0.025,10, 60,60,0.15,0.3)
130 #correlated test statistics, SFNR
131 mpFNRcorr <- SFNRcor (0.025,10, 60,60,0.15,0.3)

```

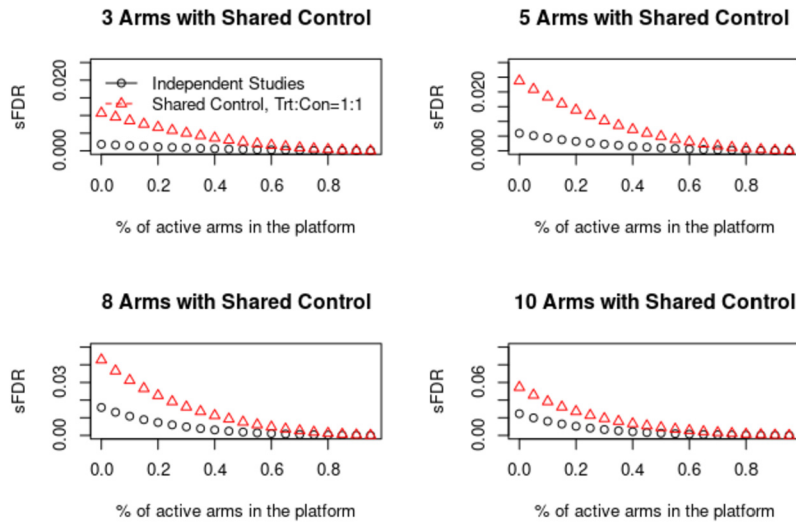


Figure 4: The relationship between SFDR and the proportion of active arms in the platform trial with shared control.

Table 5. Adjusted  $\alpha$ -level for each treatment-shared control comparison to match independent two-arm trials on the SFDR, when all arms are inactive.

# of Arms	1:1	1:3	Bonferroni
3	0.0067	0.013	0.0083
5	0.0083	0.014	0.005
8	0.011	0.016	0.0031
10	0.0123	0.016	0.0025

truly active arms are less than 95%. The difference between different  $\beta$  is also very small. For example, when there are 30% truly active arms in the 3-arm platform trial, the adjusted  $\alpha$ -level is around 0.0066 for  $\beta = 0.1, 0.15$  and  $0.2$ . The number of arms in the platform, and the randomization ratio rather than the proportion of active arms impacts the adjusted  $\alpha$ -level. Therefore, the adjusted  $\alpha$ -level given number of arms in the platform trial, and randomization ratio is shown in Table 5 as 0% of arms is active and with different randomization ratios.

In Table 5, all arms are inactive. When there are 3 arms in the platform, each comparison will be conducted at 0.0067 level instead of 0.025 and this will give the same SFDR as if the evaluation is done in independent two-arm trial. For Bonferroni-adjustment,  $0.025/3 = 0.0083$ . Compared to Bonferroni multiplicity, the adjusted  $\alpha$ -level for 3-arm platform trial with shared control is more conservative than Bonferroni. However, when the arms increase, the adjusted  $\alpha$ -level would be milder and milder. When there are 10 arms in the platform, the adjusted  $\alpha$ -level is 0.0123 for each treatment-shared control comparison while Bonferroni-adjustment would require  $0.025/10 = 0.0025$ . The randomization ratio of treatment to shared control as 1:3 would yield even milder adjustment compared to 1:1 randomization ratio.

When all arms are active, there is no  $\alpha$  penalty for each treatment-shared control comparison and each treatment-shared control comparison can use the full 0.025. However, Bonferroni multiplicity control would require the same level of adjustment.

## APPENDIX C. DETAILED DERIVATION OF SFDR AND SFNR IN INDEPENDENT CASE

When  $m_0$  is known, SFDR and SFNR can be derived as:

*SFDR*

$$\begin{aligned} &= E\left(\frac{V}{R} * I(V > 1) \mid R > 0, m_0\right) \\ &= \sum_{j=2}^{m_0} \left[ \binom{m_0}{j, m_0-j} \sum_{k=0}^{m-m_0} \binom{j}{k+j} \binom{m-m_0}{k, m-m_0-k} \right. \\ &\quad \left. \times \Pr(V = j, U = m_0 - j, S = k, T = m - m_0 - k) \right] \\ &= \sum_{j=2}^{m_0} \left[ \binom{m_0}{j, m_0-j} \alpha^j (1 - \alpha)^{m_0-j} \right. \\ &\quad \left. \times \sum_{k=0}^{m-m_0} \binom{j}{k+j} \binom{m-m_0}{k, m-m_0-k} (1 - \beta)^k * \beta^{m-m_0-k} \right]. \end{aligned}$$

*SFNR*

$$\begin{aligned} &= E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0, m_0\right) \\ &= \sum_{j=2}^{m-m_0} \left[ \binom{m-m_0}{j, m-m_0-j} \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} \right. \\ &\quad \left. \times \Pr(T = j, U = k, V = m_0 - k, S = m - m_0 - j) \right] \\ &= \sum_{j=2}^{m-m_0} \left[ \binom{m-m_0}{j, m-m_0-j} \beta^j (1 - \beta)^{m-m_0-j} \right. \\ &\quad \left. \times \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} (1 - \alpha)^k * \alpha^{m_0-k} \right]. \end{aligned}$$

When  $m_0$  is unknown, SFDR and SFNR can be derived as:

*SFDR*

$$\begin{aligned} &= E\left(\frac{V}{R} * I(V > 1) \mid R > 0\right) \\ &= E_{m_0} \left[ E\left(\frac{V}{R} * I(V > 1) \mid R > 0, m_0\right) \right] \\ &= \sum_{m_0=2}^m \left[ \Pr(m_0) E\left(\frac{V}{R} * I(V > 1) \mid R > 0, m_0\right) \right] \\ &= \sum_{m_0=2}^m \left[ \binom{m}{m_0, m-m_0} p^{m-m_0} (1-p)^{m_0} \right. \\ &\quad \left. \times \sum_{j=2}^{m_0} \binom{m_0}{j, m_0-j} \alpha^j (1 - \alpha)^{m_0-j} \right] \end{aligned}$$

$$\times \sum_{k=0}^{m-m_0} \binom{m-m_0}{k, m-m_0-k} \frac{j}{j+k} (1 - \beta)^k \beta^{m-m_0-k} \Big].$$

*SFNR*

$$\begin{aligned} &= E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0\right) \\ &= E_{m_0} \left[ E\left(\frac{T}{m-R} * I(T > 1) \mid m - R > 0, m_0\right) \right] \\ &= \sum_{m_0=0}^{m-2} \left[ \Pr(m_0) E\left(\frac{T}{m-R}, I(T > 1) \mid m - R > 0, m_0\right) \right] \\ &= \sum_{m_0=0}^{m-2} \left[ \binom{m}{m_0, m-m_0} p^{m-m_0} (1-p)^{m_0} \right. \\ &\quad \times \sum_{j=2}^{m-m_0} \binom{m-m_0}{j, m-m_0-j} \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} \\ &\quad \left. \times \Pr(T = j, U = k, V = m_0 - k, S = m - m_0 - j) \right] \\ &= \sum_{m_0=0}^{m-2} m_0 p^{m-m_0} (1-p)^{m_0} \left[ \sum_{j=2}^{m-m_0} \binom{m-m_0}{j, m-m_0-j} \right. \\ &\quad \times \beta^j * (1 - \beta)^{m-m_0-j} \\ &\quad \left. \times \sum_{k=0}^{m_0} \binom{j}{k+j} \binom{m_0}{k, m_0-k} (1 - \alpha)^k * \alpha^{m_0-k} \right]. \end{aligned}$$

## ACKNOWLEDGEMENTS

The authors thank Dr. Dong Xi (Gilead), Dr. Cong Chen (Merck), Dr. Haitao Pan (St. Jude), reviewers and the associate editor for helpful comments to improve the manuscript.

## FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Accepted 25 March 2023

## REFERENCES

- [1] BAI, X., DENG, Q. and LIU, D. (2020). Multiplicity issues for platform trials with a shared control arm. *Journal of Biopharmaceutical Statistics* **30**(6) 1077–1090.
- [2] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**(1) 289–300. [MR1325392](#)
- [3] BRETZ, F. and KOENIG, F. (2020). Commentary on parker and weir. *Clinical Trials* **17**(5) 567–569.
- [4] GOV, C. (2020). *ACTIV-2: A Study for Outpatients With COVID-19*. <https://clinicaltrials.gov/ct2/show/NCT04518410>.

- [5] COLLIGNON, O., GARTNER, C., HAIDICH, A. -B., JAMES HEMMINGS, R., HOFNER, B., PÉTAVY, F., POSCH, M., RANTELL, K., ROES, K. and SCHIEL, A. (2020). Current statistical considerations and regulatory perspectives on the planning of confirmatory basket, umbrella, and platform trials. *Clinical Pharmacology & Therapeutics* **107**(5) 1059–1067.
- [6] FDA (2022, March). *Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics Guidance for Industry*. <https://www.fda.gov/media/120721/download>.
- [7] FLEMING, T. R. (2010). Clinical trials: discerning hype from substance. *Annals of internal medicine* **153**(6) 400–406.
- [8] HOWARD, D. R., BROWN, J. M., TODD, S. and GREGORY, W. M. (2018). Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Statistical methods in medical research* **27**(5) 1513–1530. <https://doi.org/10.1177/0962280216664759>. MR3777772
- [9] LU, C., LI, X., BROGLIO, K., BYCOTT, P., JIANG, Q., LI, X., MCGLOTHLIN, A., TIAN, H. and YE, J. (2021). Practical considerations and recommendations for master protocol framework: basket, umbrella and platform trials. *Therapeutic Innovation & Regulatory Science* **55**(6) 1145–1154.
- [10] REN, Y., LI, X. and CHEN, C. (2021). Statistical considerations of phase 3 umbrella trials allowing adding one treatment arm mid-trial. *Contemporary Clinical Trials* **109** 106538.
- [11] SARKAR, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. <https://doi.org/10.1214/009053607000000398>. MR2382652
- [12] SORIĆ, B. (1989). Statistical “discoveries” and effect-size estimation. *Journal of the American Statistical Association* **84**(406) 608–610. <https://doi.org/10.1080/01621459.2020.1762614>. MR4107669
- [13] SRIDHARA, R., MARCHENKO, O., JIANG, Q., PAZDUR, R., POSCH, M., REDMAN, M., TYMOFYEV, Y., LI, X., THEORET, M., SHEN, Y. L. et al. (2022). Type I error considerations in master protocols with common control in oncology trials: report of an American statistical association biopharmaceutical section open forum discussion. *Statistics in Biopharmaceutical Research* **14**(3) 349–352.
- [14] STOREY, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *The annals of statistics* **31**(6) 2013–2035. <https://doi.org/10.1214/aos/1074290335>. MR2036398
- [15] STOREY, J. D. (2011). False Discovery Rate. *International encyclopedia of statistical science* **1** 504–508. MR2703303
- [16] STOREY, J. D. and TIBSHIRANI, R. (2001). Estimating false discovery rates under dependence, with applications to DNA microarrays. Technical Report, Technical Report 2001-28, Department of Statistics, Stanford University.
- [17] WASON, J. M. and ROBERTSON, D. S. (2021). Controlling type I error rates in multi-arm clinical trials: a case for the false discovery rate. *Pharmaceutical statistics* **20**(1) 109–116.
- [18] WOODCOCK, J. and LAVANGE, L. M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine* **377**(1) 62–70.

Jingjing Ye. Global Statistics and Data Sciences (GSDS), BeiGene, Fulton, MD, 20759, USA.

E-mail address: [jingjing.ye@beigene.com](mailto:jingjing.ye@beigene.com)

Xiaoyun (Nicole) Li. Global Statistics and Data Sciences (GSDS), BeiGene, Fulton, MD, 20759, USA.

Chengxing Lu. Analytics and Data Sciences, Biogen Inc., Cambridge MA, 02142, USA.

William Wang. Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ 07033, USA.