Monitoring Rare Serious Adverse Events from a New Treatment and Testing for Difference from a Baseline Rate

Michael P. Fay[1], Chiung-Yu Huang[1], and Nana A. Y. Twum-Danso[2]
1. Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA.
2. Task Force for Child Survival and Development, Decatur, Georgia, USA.

Motivating Example

This poster describes the design of a study of a mass treatment program for the elimination of lymphatic filariasis (LF) using albendazole and ivermectin in areas that are co-endemic for *Wuchereria bancrofti*, the parasite that causes LF in Africa, *Onchocerca volulus* and *Loa loa* which have had ivermectin mass treatment programs for onchocerciasis control for many years. The focus of the study will be on whether the addition of albendazole to an existing ivermectin mass treatment program for onchocerciasis control increases the rates of neurologic *Loa loa* related serious adverse events (SAE) compared to ivermectin alone in areas that are endemic for *W. bancrofti*, *O. volulus* and *L. loa*. The sample size calculations for this trial are non-standard because of the following reasons:

- Our primary interest is not in showing that one of the two treatments works better in fighting disease, but in showing that the new treatment (combining albendazole and ivermectin) does not have higher neurologic *L. loa* related SAE rates than the standard treatment (ivermectin alone). This is a non-inferiority type of study on neurologic *L. loa* related SAE.
- Both treatments will not be applied concurrently. Instead the new treatment will be compared to historical data on the standard treatment in the same population, which has already been collected.
- Since the primary endpoint is neurologic *L. loa* related SAE, we will want to plan for early stopping in the study if the new treatment has a higher rate of neurologic *Loa loa* related SAE than the standard.

For this example, the historical data are from Gardon, et al (1997) who observed a rate of $2/17877$ (i.e., 1.12 per 10,000) of neurologic *L. loa* related SAE after receiving ivermectin alone.

Statistical Framing of the Problem

Let X be the number of SAE out of m, from the historical control group (e.g., $X = 2$ out of $m = 17877$ on ivermectin alone) and Y_n be the number of SAE out of the first n who get the new treatment (ivermectin and albendazole). Assume $X \sim Binomial(m, \mu_x)$ and $Y_n \sim Binomial(n, \mu_y)$

Non-inferiority hypotheses: show that rate of SAE for new treatment is not worse (within some indifference zone). We consider 2 choices, the ratio test and the difference test.

$$\begin{array}{rr} Ratio & Difference \\ Null: \mu_y \geq \Delta\mu_x & Null: \mu_y \geq \mu_x + \delta \\ Alternative: \mu_y < \Delta\mu_x & Alternative: \mu_y < \mu_x + \delta \end{array}$$

Inferiority hypotheses: show that SAE for new treatment is worse.

$$\begin{array}{r} Null: \mu_y \leq \mu_x \\ Alternative: \mu_y > \mu_x \end{array}$$

Desired Properties of Design:

1. The type I error of the inferiority hypotheses is less than or equal to $\alpha_I = 0.05$.
2. The type I error of the non-inferiority hypotheses is less than or equal to $\alpha_N = 0.05$.
3. The type II error of the non-inferiority hypotheses at $\mu_x = \mu_y$ is less than or equal to $\beta = .2$.

Non-inferiority Hypothesis: Rate Ratio Test

We consider first an exact rate ratio test for the non-inferiority hypothesis, treating the two groups as a randomized trial and ignoring the issue of early stopping. Using this test, we calculate the conditional power of the test when $\mu_x = \mu_y = X/m$.

In Figure 1 we see that for some Δ the conditional power goes to zero as the sample size increases.

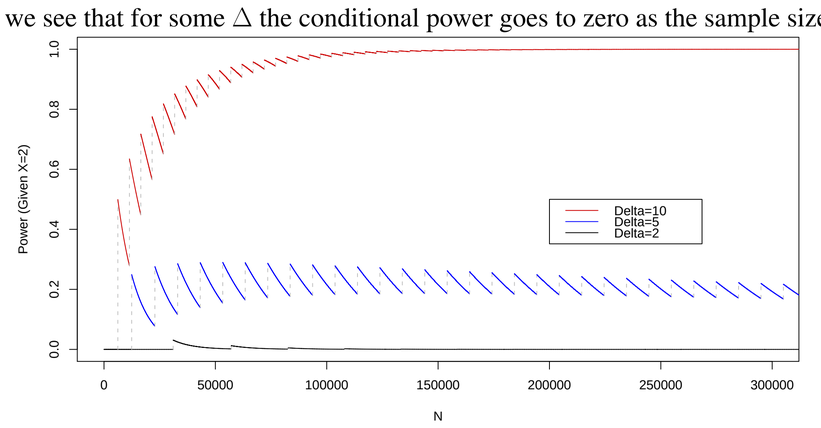

Figure 1: Conditional power of rate ratio non-inferiority tests, given $X = 2$, $m = 17877$ and $\mu_x = \mu_y = X/m$.

Asymptotic result (proof not shown):

$$\lim_{n\to\infty} CP(Ratio\ \ test|X_m, n, \mu_y, \Delta) = \begin{cases} 0 \text{ if } \Delta < \Delta^* \\ 1 \text{ if } \Delta \geq \Delta^* \end{cases}$$

Where $\Delta^* = 5.6$ for our example. Specifically,

$$\Delta^* = \max\left\{\Delta : X_m = F_P^{-1}\left(1-\alpha, \frac{m\mu_y}{\Delta}\right)\right\}$$

where F_P^{-1} is inverse Poisson distribution, i.e., $F_P^{-1}(q, \psi) = w$ is the smallest integer w such that $Pr[W \leq w] \geq q$ when $W \sim Poisson(\psi)$.

References

Berger, R.L., and Boos, D.D. (1994). "p-values maximized over a confidence set of nuisance parameter" *Journal of the American Statistical Association* **89:** 1012-1016.

Gardon, J., Gradon-Wendel, N., Demanga-Ngangue, Kamgno, J., Chippaux, J-P, Boussinesq, M. (1997). "Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for *Loa loa* infection" *Lancet* **350:** 18-22.

Noninferiority Hypothesis: Rate Difference Test

Second, we consider the exact unconditional test on the difference in proportions using the Berger-Boos (1994) method as used in StatXact (StatXact 7 Procs). The difference test has similar non-monotonic patterns in the conditional power as the ratio test. Also similarly, asymptotically there are some values of δ which will give zero power even as $n \to \infty$. Intuitively, as n goes to infinity then $\hat{\mu}_y \to \mu_y$, but we are still left with the uncertainty in $\hat{\mu}_x$. Since we want to show the alternative that $\mu_y < \mu_x + \delta$, if μ_y is known then we could create a test where we reject if $\mu_y < L_x(X) + \delta$, where $L_x(X)$ is the lower one-sided $100(1-\alpha_N)\%$ confidence limit. Thus, if $\delta < \mu_y - L_x(X)$ then the power would not exceed α_N as $n \to \infty$. For the Gardon, et al (1997) data $L_x(2) = 1.99 \times 10^{-5}$ and using $\mu_y = 2/17877$ we need $\delta > 9.20 \times 10^{-5}$ to obtain a test where the power goes to one as $n \to \infty$.

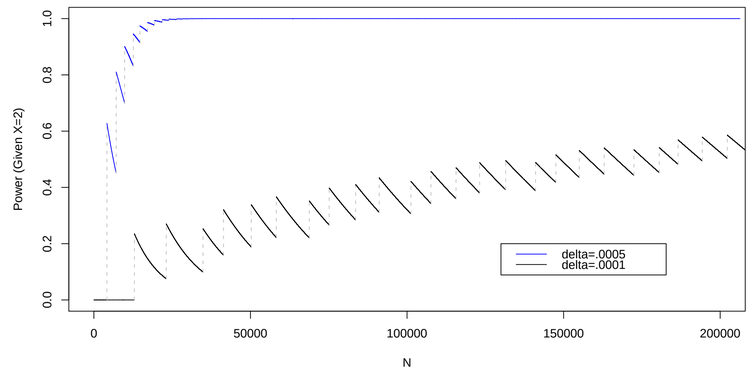

Figure 2: Conditional power of rate difference non-inferiority tests, given X, m and $\mu_x = \mu_y = X/m$.

Notice that although $\hat{\mu}_x \approx 0.0001$, there is much greater power to test the difference test with $\delta = 0.0001$ than the ratio test with $\delta = 2$ even though the alternative hypothesis for both tests when $\mu_x = .0001$ is approximately $\mu_y > 0.0002$. The reason for the difference is that μ_x is not known and the probability of very small values for μ_x must be accounted for in both tests, and very small values of μ_x more substantially affect the ratio than the difference. Further, the difference in risk of a SAE may be more important from a public health point of view rather than the ratio of risks, since cases with μ_x very small are the least important from the public health perspective. Thus, although the difference test is much more difficult to calculate, the test based on the difference will often be preferred.

Combining Both Hypotheses

The proposal for these types of studies is the following. Stop early for inferiority of the new treatment if any uncorrected exact ratio test of the data up until that point rejects, but we require $Y_n > 1$. Otherwise stop the study at the value n such that the conditional power at $\mu_x = X/m$ is greater than 80% for all subsequent n and use the exact unconditional difference in proportions test with $\delta = 0.0005$. When $X = 2$ and $m = 17877$ the resulting design is shown in Figure 3. Figure 4 shows the properties given X and m.

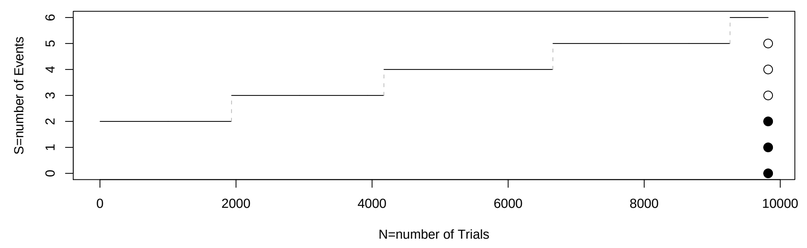

Figure 3: Proposed Design: Solid lines denote stop and find new treatment worse for SAE rate, solid circles denote stop and find new treatment non-inferior with $\delta = .0005$, open circles denote inconclusive.

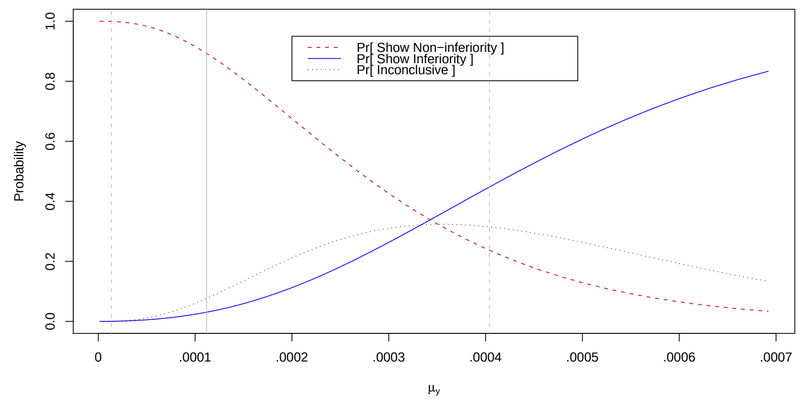

Figure 4: Probabilities for 3 decisions of proposed design (where $\delta = .0005$ for the non-inferiority hypotheses), given $X = 2$ and $m = 17877$. Gray lines are $\hat{\mu}_x = X/m$ (solid) and the 95% confidence intervals.

Unconditional Probabilities of Procedure

We create designs for $X = 1, \ldots, 17$ as above. Then since $Pr[X \leq 17|\mu_x = .0004] \geq .9995$, we can calculate the unconditional probabilities of the three decisions when applying designs following the proposal (given $X > 1$). The results are in Figure 5.

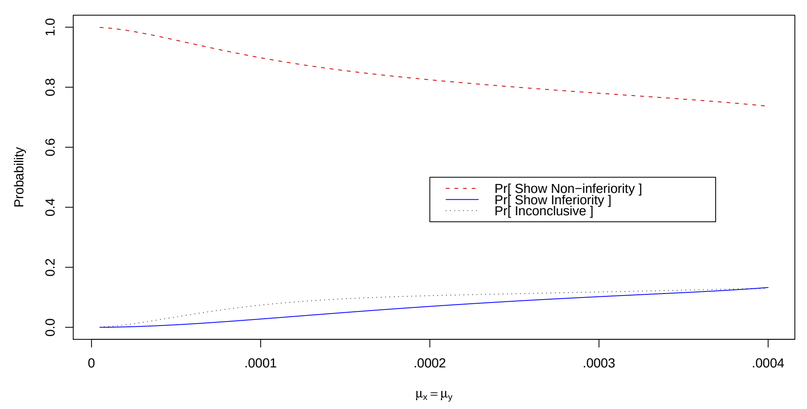

Figure 5: Unconditional probabilities for 3 decisions for all possible studies using same algorithm as proposed with $\delta = .0005$ for non-inferiority hypotheses, given $X > 1$ and $m = 17877$.