



University of Wisconsin – Madison

---

# Null Hypothesis Scales in Non-Inferiority Trials

Rick Chappell

Professor, Department of Statistics and  
Department of Biostatistics and Medical Informatics

University of Wisconsin

[chappell@stat.wisc.edu](mailto:chappell@stat.wisc.edu)

Xiaotian Zhu

Senior Research Statistician, AbbVie



# Outline

---

- I. CPORT – a Motivating Example**
- II. Specifying the Hypotheses**
- III. Consequences of the Choice**



# I. A Motivating Example - CPORT

---

**Sponsor:** Cardiovascular Patient Outcomes Research Team

**Comparison:** Angioplasty at hospitals with vs. without on-side cardiac surgery

**Follow-up & Outcome:** 6 week mortality

**Type:** Non-inferiority, .8% vs. 1.2% (-.4% additive margin? 2/3 multiplicative margin?), multicenter 1:3 randomized

**Sample size:** 18,500 patients



## II. Specifying the Hypotheses

---

Hypotheses in Superiority Trials for Power Calculations

$H_0: C - T = 0$  (Treatment may have no effect)

VS.

$H_A: C - T = \Delta$  (Treatment has Good Effect)

Hypotheses in Non-Inferiority Trials

$H_0: C - T \leq \Delta$  (Treatment might be much worse)

VS.

$H_A: C - T = 0$  (Treatment isn't much worse)

These may be on the raw, log, log-odds, or other scale.



---

"In Superiority Trials, all Null Hypotheses are the Same but all Alternatives are Different."

"In Non-inferiority Trials, all Alternative Hypotheses are the Same but all Nulls are Different."

- Tolstoy?



For example:

$$\begin{aligned}S_C(t) - S_T(t) &= \Delta, & S_C(t) / S_T(t) &= \Delta', \\ \lambda_C(t) - \lambda_T(t) &= \Delta'', & \lambda_C(t) / \lambda_T(t) &= \Delta''', \\ \text{median}_C - \text{median}_T &= \Delta''''\end{aligned}$$

are in general all incompatible unless

$$\begin{aligned}\Delta = \Delta'' = \Delta'''' = 0 & \text{ and} \\ \Delta' = \Delta''' = 1.\end{aligned}$$

And hypotheses for binary outcomes can be on various scales also: Chen (2000) gives methods for testing equivalence of **differences** of proportions, **ratios** of proportions, and **odds ratios**.



## III. Consequences of the Choice

---

In the above example with  $H_0$  of 1.2% vs. .8%, we have a choice:

1. Additive non-inferiority margin,  $\Delta = -.4\%$
2. Multiplicative non-inferiority margin,  $\Delta^* = 2/3$

There is much literature on choice of the margin's magnitude but almost none on its scale.



## A. Scale choice and balanced randomization

---

Consider the simple normal two-sample constant variance case:

$$X_{iT} \sim \text{iidN}(\mu_T, \sigma^2)$$

$$X_{iC} \sim \text{iidN}(\mu_C, \sigma^2)$$

Then the usual (unstandardized) test statistic for superiority is the difference in sample means.

Obviously, the allocation which minimizes its variance is 1:1.





---

Suppose we are designing an equivalence trial to test the hypothesis

$$H_0: \mu_C - \mu_T \leq \Delta .$$

Then the optimal allocation is still 1:1. However, if we want to test

$$H_0: \mu_C / \mu_T \leq \Delta^* ,$$

which is equivalent to

$$H_0: \mu_C - \Delta^* \times \mu_T \leq 0$$

instead, then the optimal allocation is  $\Delta^*$  : 1 in favor of the control group - a big difference.



---

“Suppose we are designing an equivalence trial to test the hypothesis

$$H_0: \mu_C - \mu_T \leq \Delta.$$

Then the optimal allocation is still 1:1.”

This statement is false if  $\mu_C$  and  $\mu_T$  are probabilities, even though we are dealing with additive non-inferiority. Sample proportions’ variances differ under  $H_0$ .

This also applies to other parameters whose estimates’ variances are not constant.



## B. Scale choice and power

---

Suppose in CPORT, which has a binary “failure” outcome, we are deciding between the null hypothesis of additive inferiority

$$\mathbf{H}_0: \pi_C - \pi_T \geq -.004$$

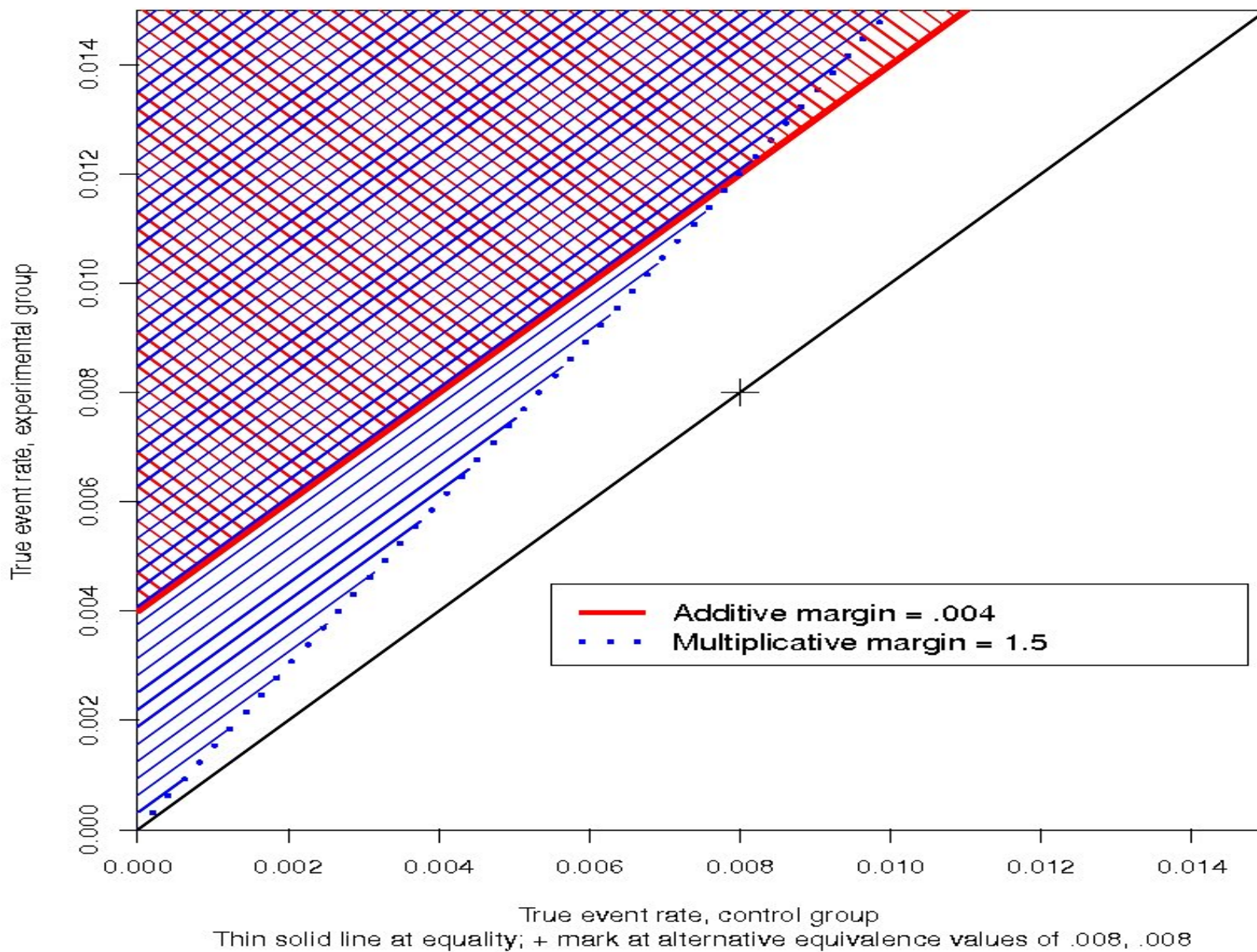
and that of multiplicative inferiority

$$\mathbf{H}_0: \pi_C / \pi_T \geq 2/3.$$

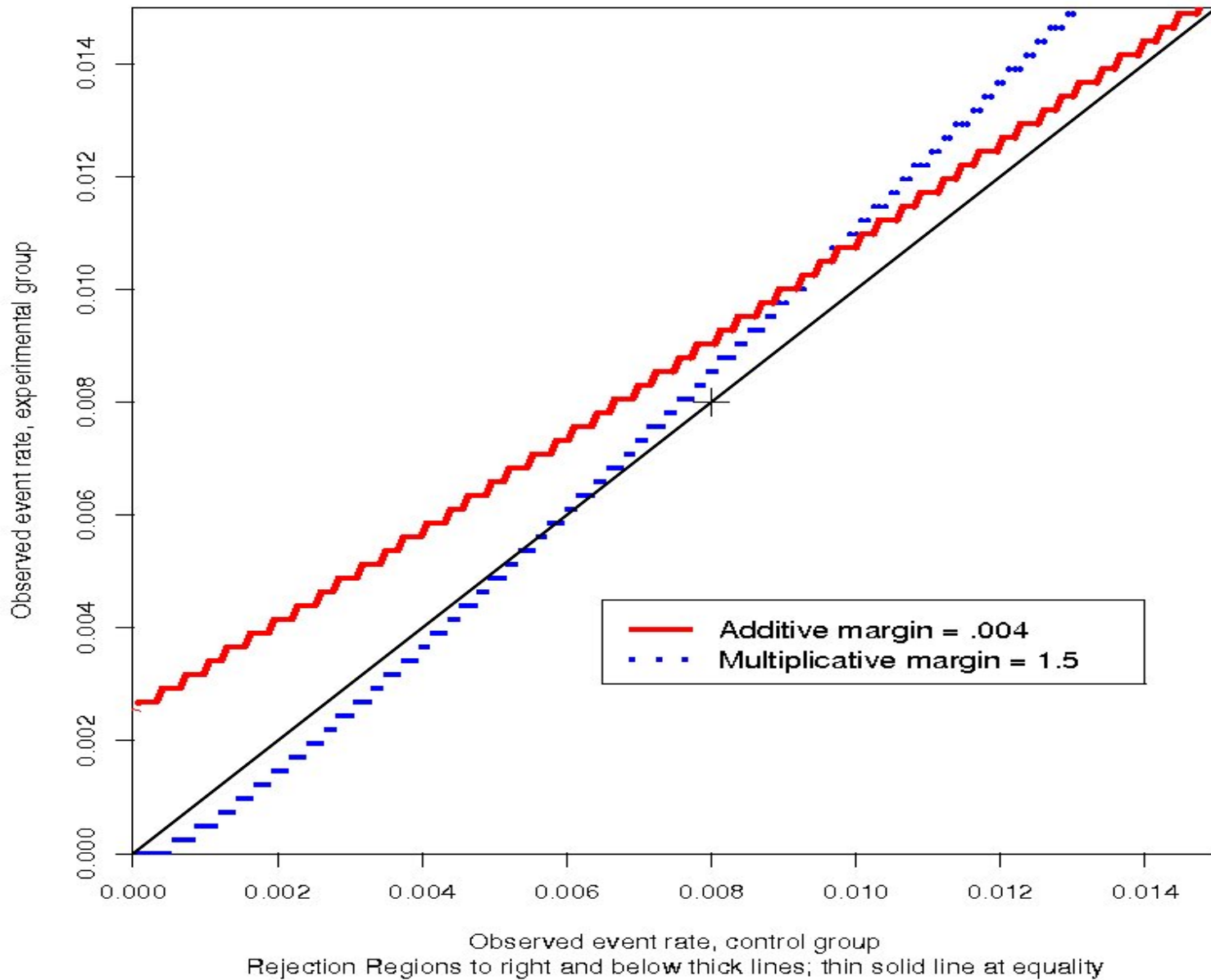
These are identical at  $\pi_C = .008$ . Should power be the same for the common alternative

$$\mathbf{H}_A: \pi_T = \pi_C = .008?$$

**Fig. 1: Null Hypotheses for CPORT Trial**



**Fig. 2: Rejection Boundaries for CPORT Trial (n=12300+4100)**





---

Answer: no (surprisingly, to me and the trial's principal investigator).

For proportions less than .01, the range in which we are interested, the hypothesis of multiplicative non-inferiority is much more demanding and requires a larger sample size:

About **21,000** instead of **16,400!**



## A similar current situation:

---

There is interest in toxicities due to an unnamed drug.

A non-inferiority trial comparing toxicities in patients treated with active vs. control compounds is planned under the alternative of

.02/year hazard in each group.

Absolute comparisons (hazard differences or 3-year survival differences) may be more relevant than relative ones (hazard ratios).

We investigated the number of events necessary in a non-inferiority clinical trial with a survival endpoint.



## A similar current situation:

---

Simulations *under proportional hazards* show:

A test of difference in Kaplan-Meier curves at three years is *more powerful* than a test of the equivalent margin for hazard ratios from a Cox PH model.

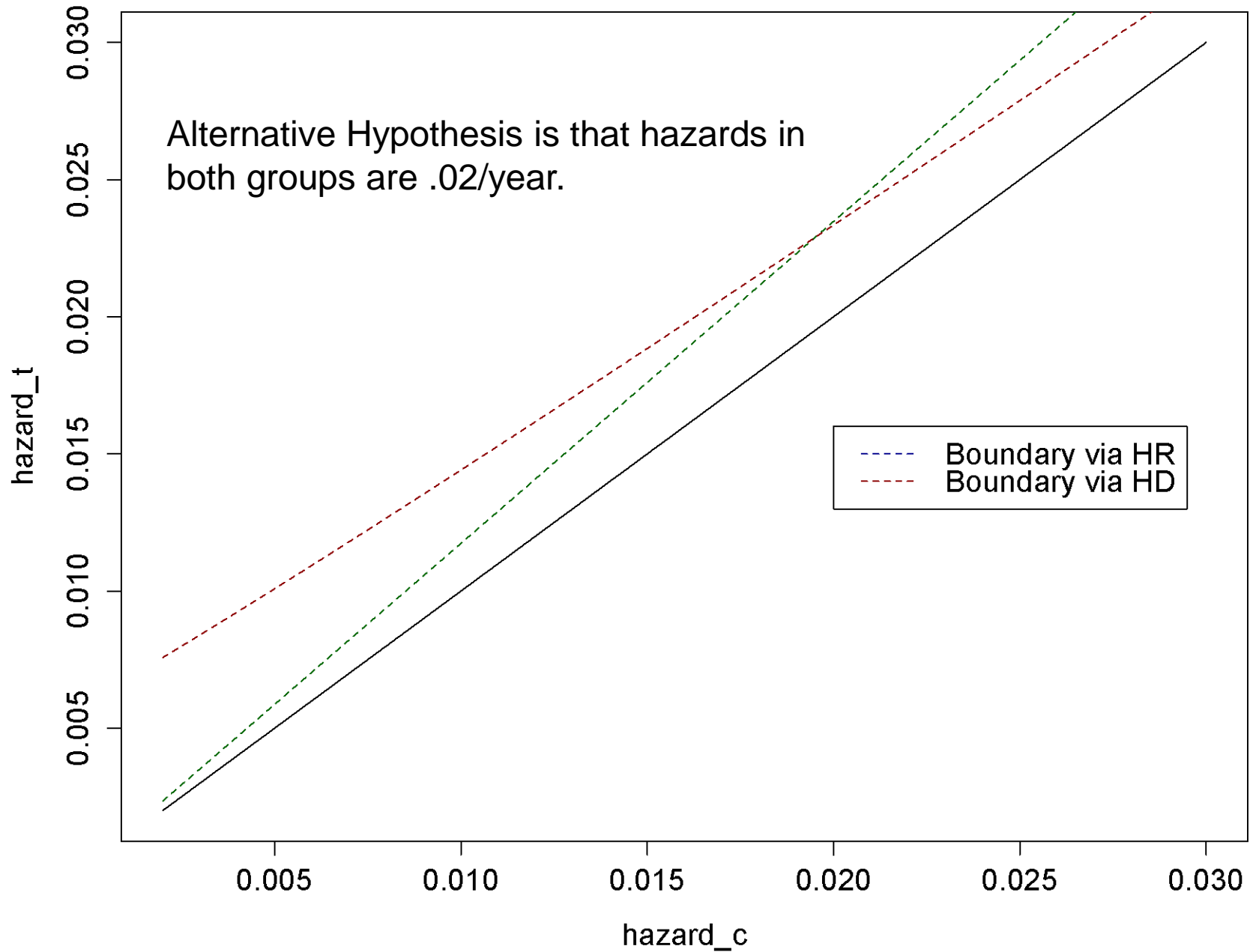
How could this be? It isn't true for superiority trials.

As with the above binary example, the non-inferiority null hypotheses had different interpretations.

Note that for small event rates, as here, the hazard and survival differences are nearly identical, as are the hazard and survival ratios are similar.



# Boundaries of Additive vs. Relative Hazard Hypotheses



# Critical Regions Associated with each Hypothesis

