The optimal randomization ratio in time-to-event trials is not 1:1

Godwin Yung, Genentech/Roche

Acknowledgements

Yi Liu, Nektar Therapeutics (main collaborator)

Hans Ulrich Burger, Roche

Kaspar Rufibach, Roche

Marcel Wolbers, Roche

Outline

- The call for/against randomization
- To what ratio should patients be randomized?
  - Power: Schoenfeld, Freedman, and Luo
  - Efficacy, safety, trial duration, data maturity
- Example

The call for/against randomization

- Ethics: Human investigations should employ designs that minimize the total burden for study volunteers.[1-2]
- Science: RCTs provide the most valid means of establishing causal efficacy.
- Tension between the two are on full display during public health emergencies:
  - Ebola epidemic (2014-2016)[3]
  - COVID-19 pandemic (2019-)[4]

To what ratio should patients be randomized?

- 1:1 is the typical choice.
- There is an unspoken but common misconception that 1:1 randomization ratio (RR) optimizes statistical power in time-to-event trials.
- Continuous endpoint, equal variance[5]
  - "… However, 2:1 or 3:2 does not greatly alter statistical properties."
- Survival endpoint – Schoenfeld (1981)
  - Approximation for the log-rank statistic $Z \sim N(\mu_S, 1)$.
  - Alternative approximations to the log-rank statistic say otherwise.

In a trial comparing an active therapy to control, let HR denote the hazard ratio, k the sample size ratio of active to control, and d the minimum required number of events to achieve power $1-\beta$.

- Schoenfeld (1981)

$$Z \sim N(\mu_S, 1) \Longrightarrow d = \left(\frac{(z_{\alpha/2} + z_\beta)(1+k)}{\log HR}\right)^2 \frac{1}{k} \Longrightarrow \arg\max_k \mu_S = 1$$

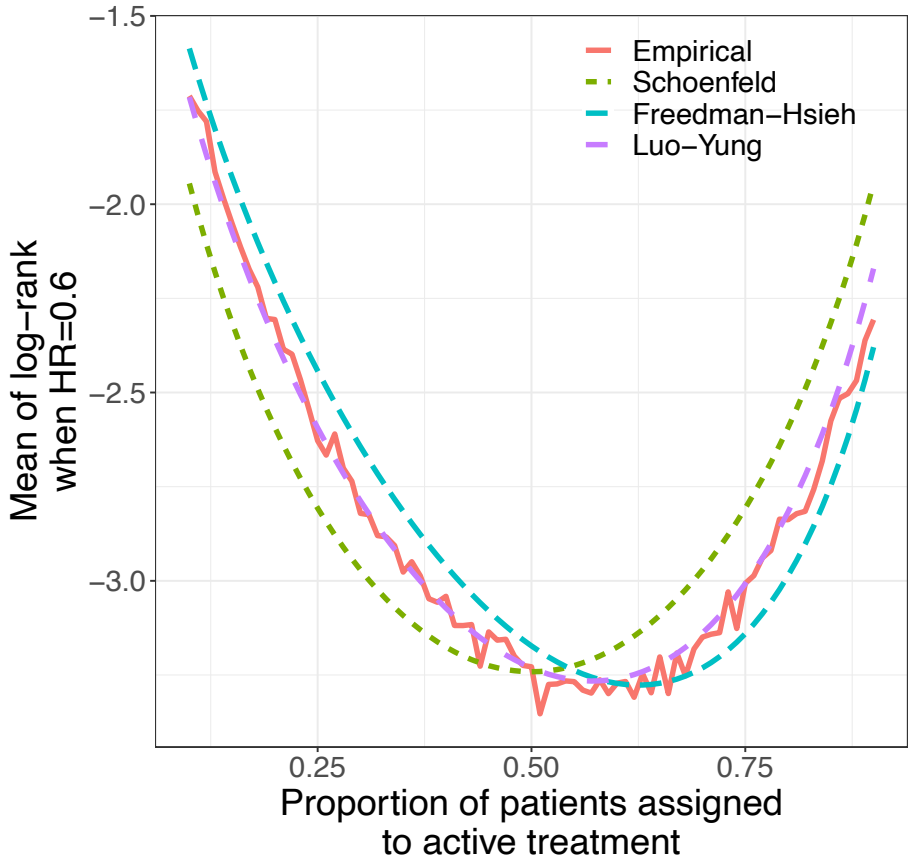
- Freedman (1982) and Hsieh (1992)

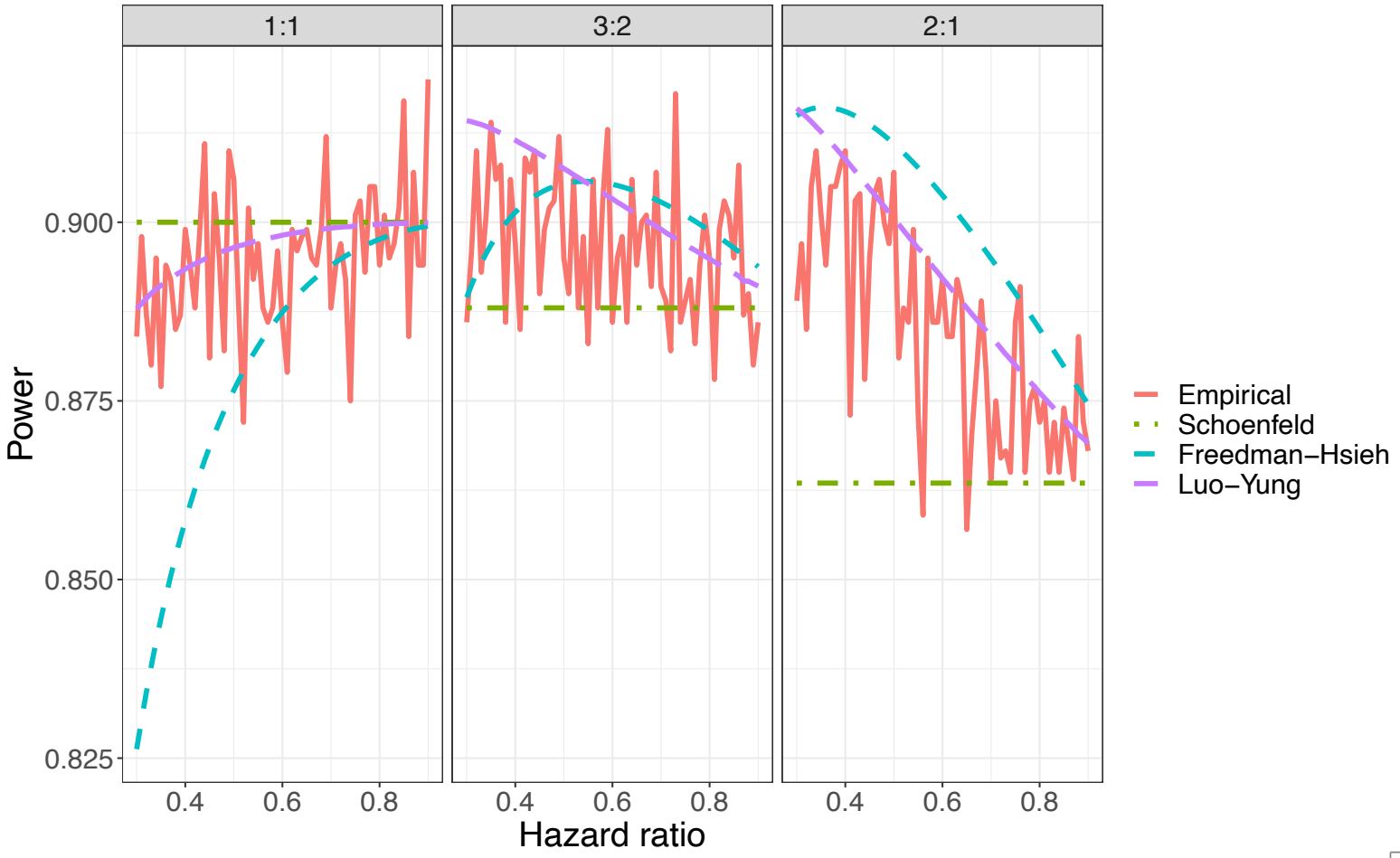
$$Z \sim N(\mu_F, 1) \Longrightarrow d = \left(\frac{(z_{\alpha/2} + z_\beta)(1+HR \times k)}{1-HR}\right)^2 \frac{1}{k} \Longrightarrow \arg\max_k \mu_F = \frac{1}{HR}$$

- Luo et al. (2018) and **Yung and Liu (2019)**

$$Z \sim N(\mu_L, 1) \Longrightarrow k \text{ such that events are balanced across arms}$$

Comparing the three approximations for the log-rank statistic




1:1
3:2
2:1
Power
0.900
0.875
0.850
0.825
0.4
0.6
0.8
Hazard ratio
Empirical
Schoenfeld
Freedman–Hsieh
Luo–Yung

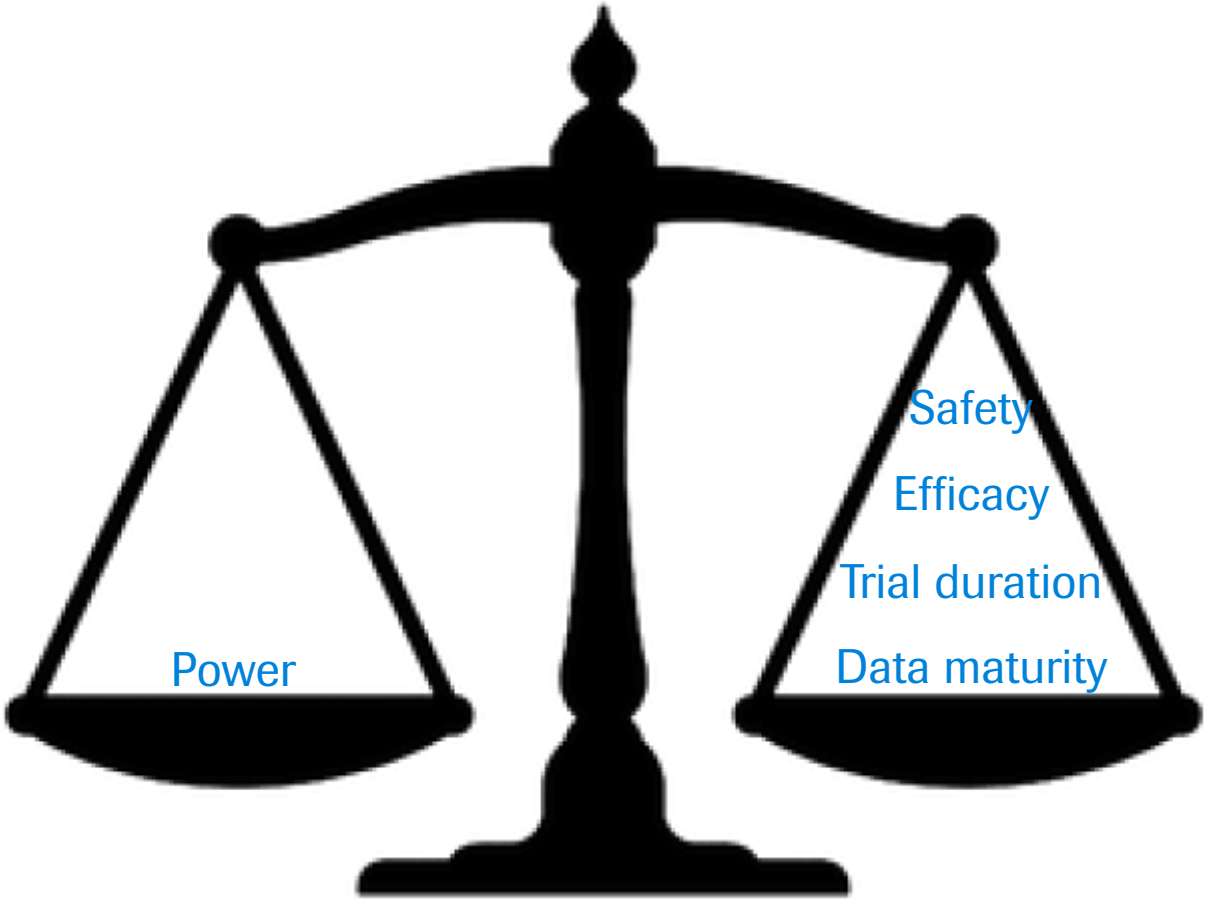
Recommended RR to maximize power given fixed sample and event size

- Luo approximates the empirical power most accurately.
- Its implication on RR is intuitive: balance the effective sample size.
- Contrary to common belief, 1:1 does not necessarily maximize power. A rule of thumb *to jumpstart considerations*:

Design HR	RR
0.7 to 1.0	1:1
0.5 to 0.7	3:2
Less than 0.5	2:1

Following this rule, unequal randomization will have similar power to 1:1.

Other factors for consideration



- Efficacy ↔ Safety: If the risk-benefit ratio is in favor of treatment, then patients stand to benefit from a higher chance of being assigned to treatment. However, the opposite (inverse) is also true.
- Trial duration ↔ Efficacy: If unequal randomization is attractive to patients and leads to faster accrual, then overall timeline may be shortened. But if accrual is not accelerated, then overall timeline may be delayed.
- Data maturity ↔ Trial duration: Randomizing more patients to active treatment may accelerate time at which “mature” data is achieved, e.g. more events contributed by active treatment may lead to higher chance of observing median survival in this arm.

Example

Base case:

- Survival: 18m vs 25.7m (HR=0.7)
- $\alpha = 0.05$ two-sided
- Target ~90% power
- N=530 patients, randomized 1:1
- Accrual=~27 months
- Final=331 events (~39 months)

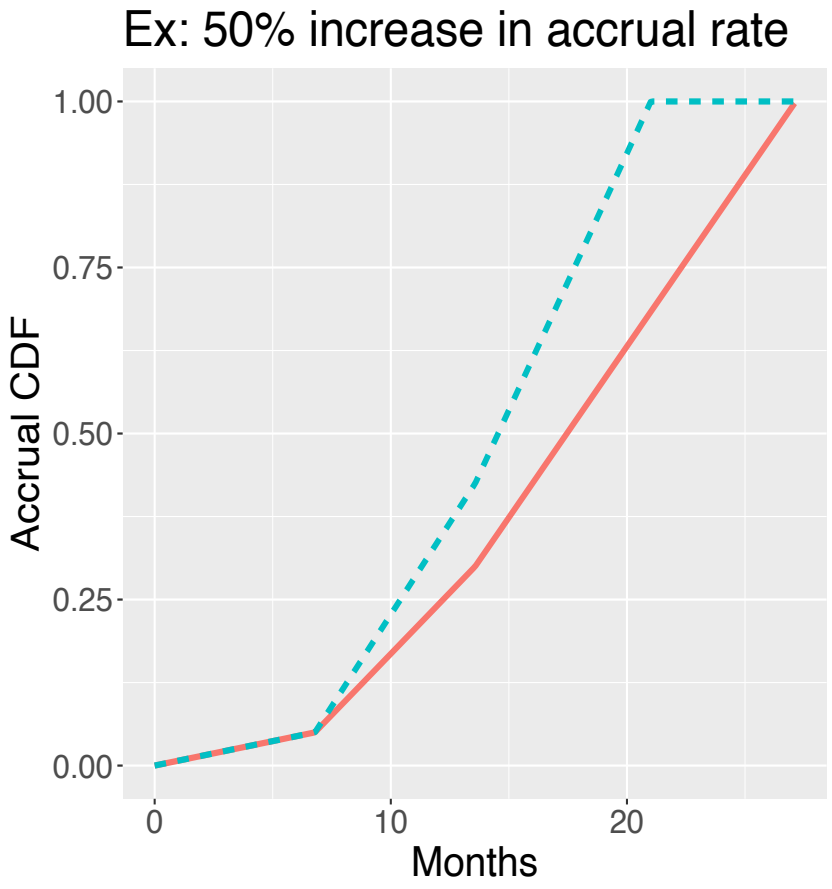
Questions: If we were to randomize more patients to active treatment while fixing sample size and event size, then

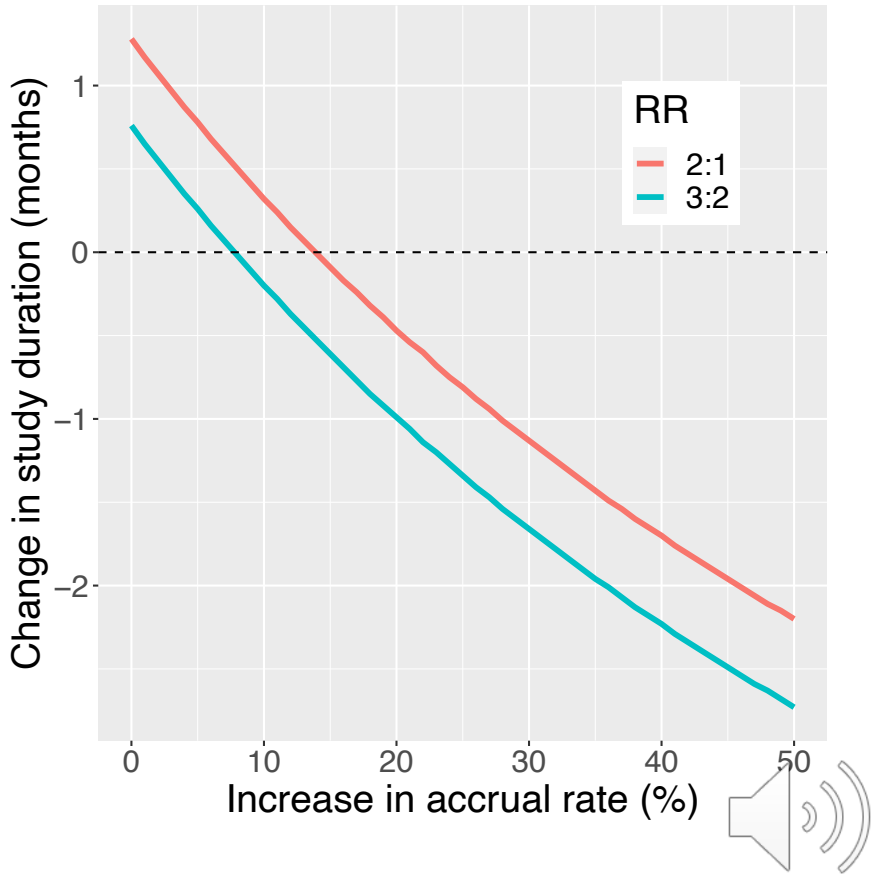
1. how would *power* change?

HR	Power
1:1	88.8%
3:2	89.1%
2:1	86.4%

2. how would *trial duration* change?
3. what would be the increase in probability that we observe the *median survivals*?

Q2: … how would trial duration change?

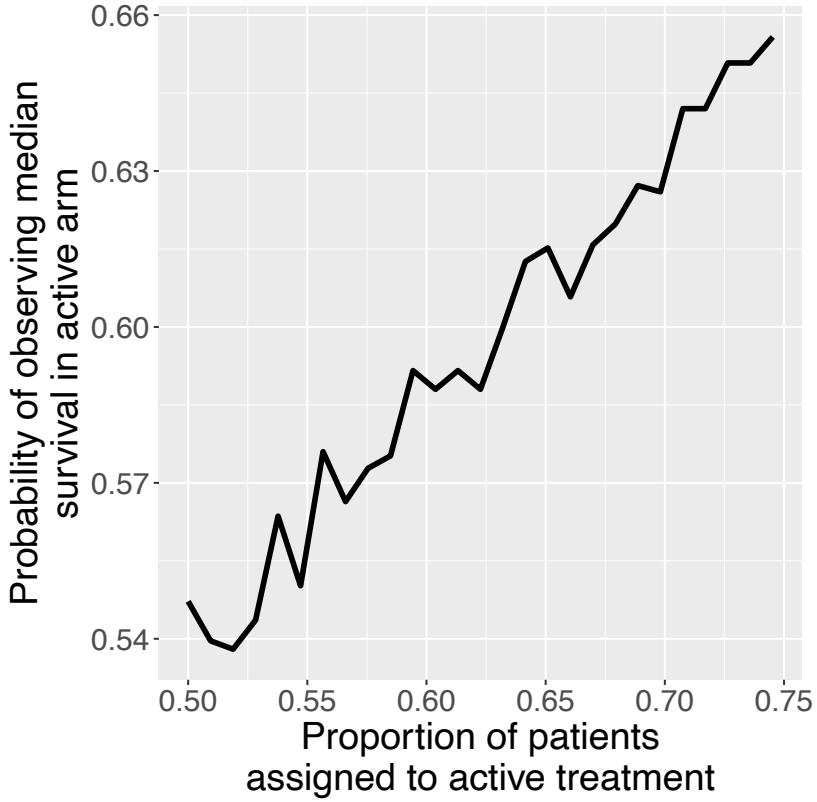




Q3: … what would be the increase in probability that we observe the median survivals?

Interim analysis at 166 events (50% of total 331)

5000 simulations



Discussion

- When comparing active treatment to control, randomizing more patients to the active arm may lead to similar power compared to 1:1.
- In practice, other factors need to be taken into consideration, e.g. safety, efficacy, trial duration, data maturity.
- Unequal randomization may be especially attractive in settings where
  i. Patients are reluctant to enroll in a study with 1:1 RR (e.g. debilitating disease, vulnerable population, ineffective control).
  ii. Randomization is required to prevent bias.
  iii. True equipoise is absent (e.g. prior information on efficacy and safety)
- Combine unequal randomization with use of historical controls

References

1. Avins, AL. (1998) “Can Unequal be More Fair? Ethics, Subject Allocation, and Randomised Clinical Trials.” *Journal of Medical Ethics* 24:401-408.

2. Hey, SP and Kimmelman, J. (2013) “The questionable use of unequal allocation in confirmatory trials.” *Neurology* 82:77-79.

3. National Academy of Sciences, Engineering, and Medicine 2017. *Integrating Clinical Research into Epidemic Response: The Ebola Experience*. Washington, DC: The National Academies Press.

4. Lane, HC and Fauci, AS. (2020) “Research in the Context of a Pandemic.” *New England Journal of Medicine.* doi: 10.1056/NEJMe2024638.

5. Pocock, S. (1979) Allocation of Patients to Treatment in Clinical Trials. *Biometrics* 35:183-197.

6. Schoenfeld, D. (1981) “The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions.” *Biometrika* 68:316-319.

7. Freedman, LS. (1982) “Tables of the Number of Patients Required in Clinical Trials Using the Logrank Test.” *Statistics in Medicine* 1:121-129.

8. Hsieh, FY. (1992) “Comparing Sample Size Formulae for Trials with Unbalanced Allocating Using the Logrank Test.” *Statistics in Medicine* 11:1091-1098.

References cont.

9. Luo, X, Mao, X, Chen, X, Qiu, J, Bai, S, and Quan, H. (2018) “Design and Monitoring of Survival Trials in Complex Scenarios.” *Statistics in Medicine* 38:192–209.

10. Yung, G and Liu, Y. (2019) “Sample Size and Power for the Weighted Log-rank Test and Kaplan-Meier Based Tests with Allowance for Nonproportional Hazards.” *Biometrics.* doi:10.1111/biom.13196.

Doing now what patients need next