A Novel Design for Decision Rules Based on Statistical Testing Strategies in a Definitive Go/No-Go Clinical Study

Ming Zhou

Bristol-Myers Squibb, New Jersey, USA

- *joint work with Dr. Larry Z. Shen*

PSI & DIA Virtual Journal Club
December 15, 2015

Outline

Motivation

- Investment in new drug development is both costly and risky.
- Go/No-go decisions are to be made between development phases
- Potential approaches to go/no-go decisions
  - Modeling and Simulation: Burman et al. (2005) and Kowalski et al. (2007) gave an overview of modelling by combining PK/PD data with clinical data and incorporating the modeling in trial simulations to improve decision-making in clinical development.
  - Meta-Analyses
  - Chuang-Stein et al. (2011) proposed a quantitative approach by combining the ideas of diagnostic tests and hypothesis tests for making go/no-go decisions in drug development.
  - Nothing beats collecting and using more information to inform decisions

A Definite Go/No-Go Study

- A definitive go/no-go clinical study is sometimes conducted before major investment to advance drug development into new phases.
- There can be many aspects for such a go/no-go study, e.g., whether certain type of adverse event could bring a potential safety risk for a Phase III program; whether a new device could be used, etc.
- Here we focus on binary endpoints, e.g., AE, device failure, etc.
- The go/no-go study is mainly for internal decision-making.

Study Design and Hypothesis Testing

- Consider a single arm study where subjects are asked to test a new device. Let p represent the failure rate of the device.
- Select $p_2 < p_1 < p_0$, such that
  - p_0: a failure rate at which it would not be prudent to move into the next phase.
  - p_1: an acceptable rate to move into next phase, where more information about the device would be collected.
  - p_2: ideal failure rate below which the sponsor would have total confidence to go to the next phase with the current product presentation.
- Very often it is unrealistic to power a go/no-go study to rule out p_2.

Objective

Focus on the one sample test of the binomial proportion

$$H_0 : p \geq p_0 \quad vs \quad H_1 : p \leq p_1.$$

Rewrite it as

$$H_0 : p \geq p_0 \quad vs \quad H_a : p < p_0 \tag{1}$$
$$K_0 : p \leq p_1 \quad vs \quad K_a : p > p_1 \tag{2}$$

- Examine certain situations where the difficult or ambiguous outcome might happen, e.g. we fail to reject either of H_0 and K_0.
- Try to avoid ambiguous outcomes by proposing a straightforward and intuitive procedure, equipped with easy-to-interpret graphical outputs.

A Simple Decision Rule Based on Hypothesis Testing

The main goal is to allow a sponsor to make clear-cut decisions:

H_0 is rejected	$\Rightarrow$	Move to the next phase
Fail to reject H_0	$\Rightarrow$	Do not move to the next phase

Sample Size Requirement

Let S_n be the number of failures based on n independent Bernoulli tests. H_0 will be rejected at one-sided level α if $S_n \leq c_0$, where (Fleming, 1982)

$$c_0 = np_0 + z_\alpha \sqrt{np_0(1-p_0)}. \tag{3}$$

The sample size required for the test in (1) to have one-sided significance level α and power $1-\beta$ is approximately

$$n_0 = \left(\frac{z_\alpha \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_1(1-p_1)}}{p_0 - p_1} \right)^2. \tag{4}$$

Wilson Confidence Interval (CI)

Let $[\hat{p}_L(\beta), \hat{p}_U(\alpha)]$ be $100 \times (1 - \alpha - \beta)\%$ Wilson CI (Wilson, 1927) for p such that

$$p \in [\hat{p}_L(\beta), \hat{p}_U(\alpha)] \Leftrightarrow z_\alpha \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq z_{1-\beta}.$$

- The Wilson CI can be used equivalently for testing (1).
- For simplicity of discussion, we assume that $\alpha = \beta$, and write the CI simply as $[\hat{p}_L, \hat{p}_U]$.

Hypothesis Testing and Wilson CI

For the designated sample size n_0 in (4), and H_0, K_0 in (1), (2),

- If we reject H_0, i.e., $S_{n_0} \leq c_0$, then
  - $\hat{p}_U \leq p_0$
- If we fail to reject H_0, i.e., $S_{n_0} > c_0$, then
  - $\hat{p}_U > p_0$
  - $\hat{p}_L > p_1$ (reject K_0).

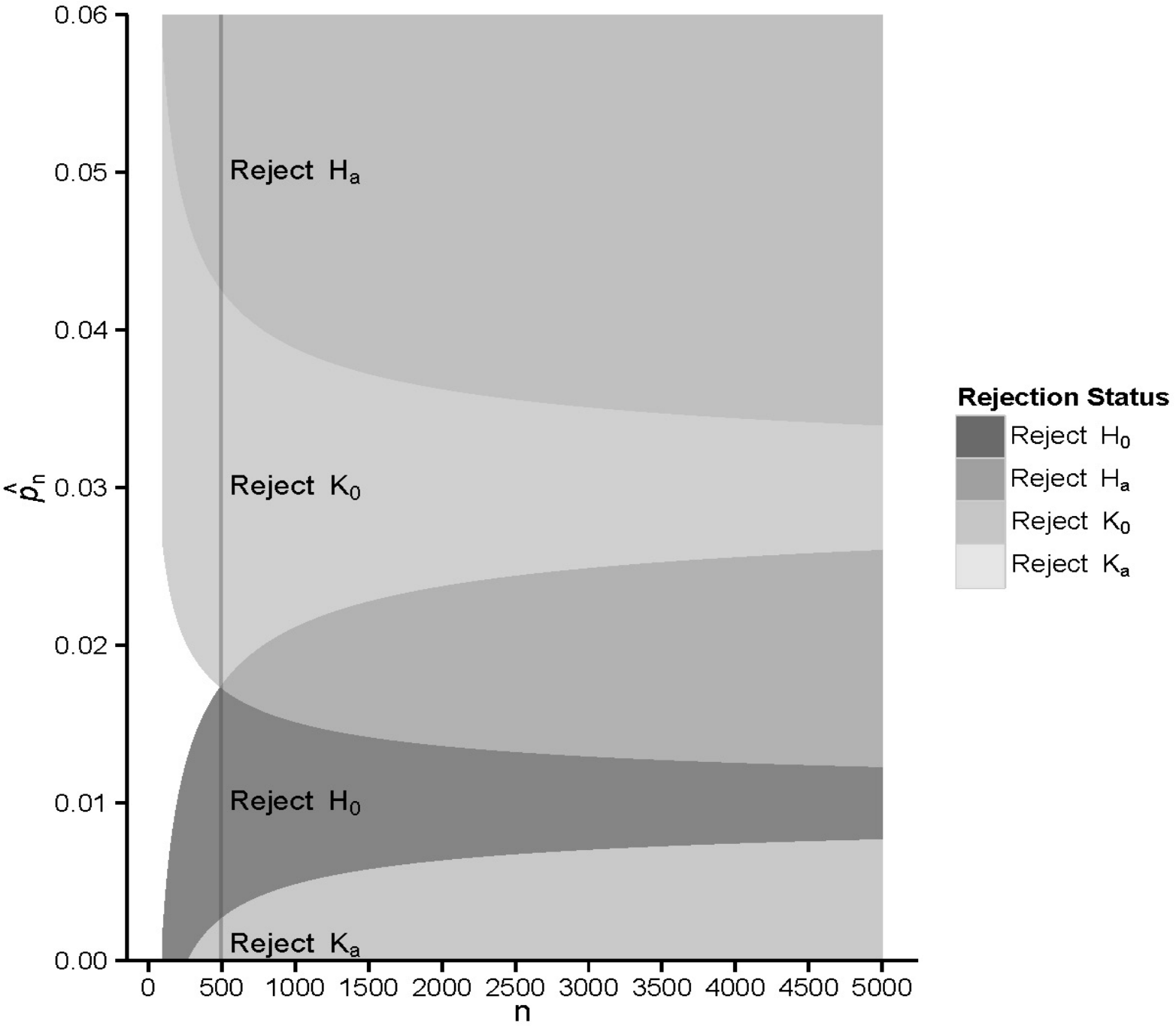
With designated sample size, we would always either reject H_0 or reject K_0.

Clear-Cut Decision Rules

The following two decision rules are equivalent:

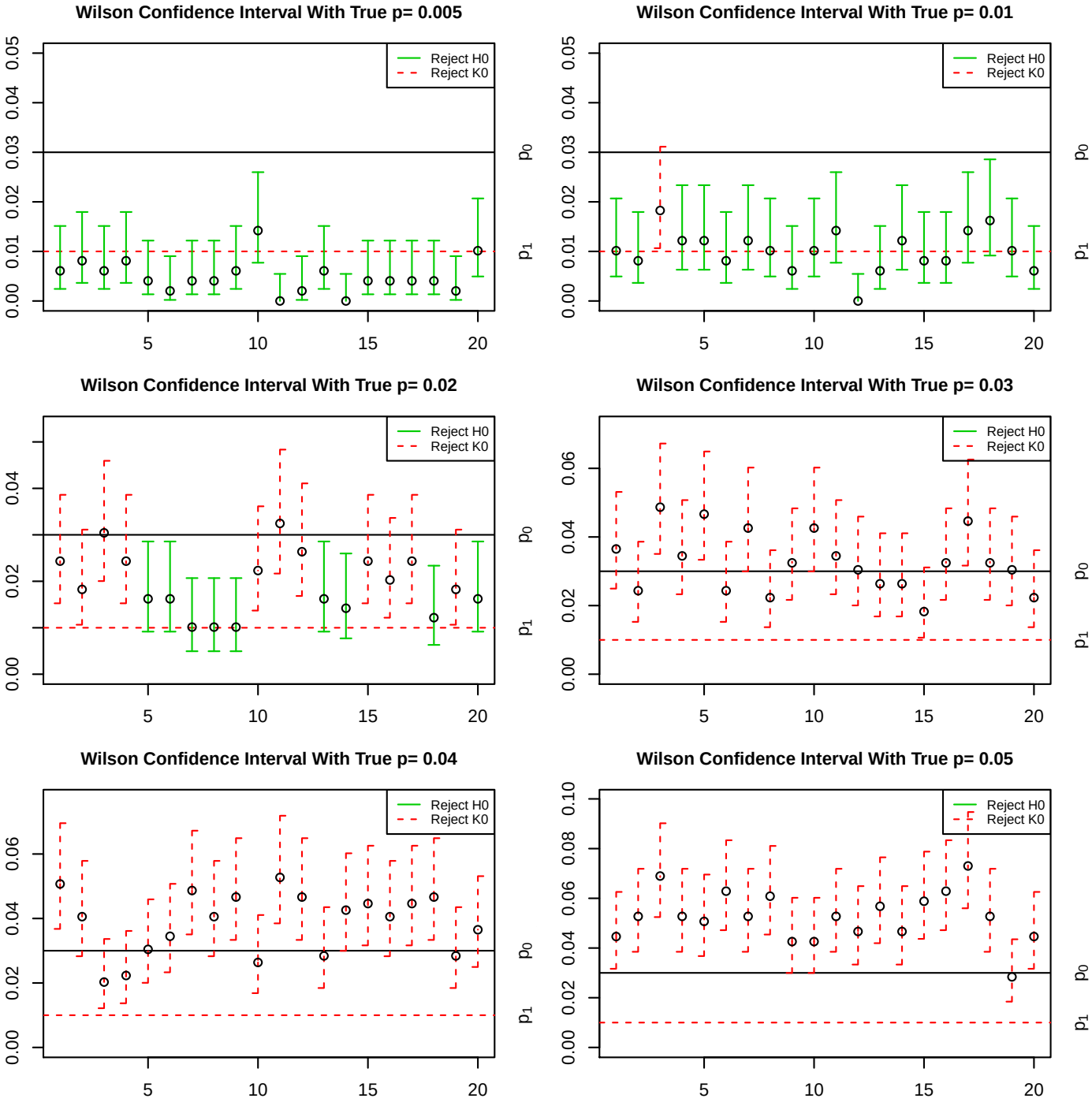
$$\begin{aligned} H_0 \text{ is rejected} &\Rightarrow \text{Move to the next phase} \\ \text{Fail to reject } H_0 &\Rightarrow \text{Do not move to the next phase} \end{aligned}$$

$$\begin{aligned} H_0 \text{ is rejected} &\Rightarrow \text{Move to the next phase} \\ K_0 \text{ is rejected} &\Rightarrow \text{Do not move to the next phase} \end{aligned}$$


0.06
0.05
0.04
0.03
0.02
0.01
0.00
$\hat{p}_n$
Reject H_a
Reject K_0
Reject H_0
Reject K_a
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
n
Rejection Status
Reject H_0
Reject H_a
Reject K_0
Reject K_a

Example

- Consider a situation where $p_0 = 0.03, p_1 = 0.01$ and $\alpha = \beta = 0.05$, which leads to
  - $(n_0, c_0) = (493, 8)$ and
  - a two-sided Wilson 90% confidence will be used.
- For each $p \in \{0.005, 0.01, 0.02, 0.03, 0.04, 0.05\}$, 20 binomial samples from $\mathrm{Binomial}(493, p)$ are generated.
- Wilson CIs (two-sided 90%) are plotted and the rejection status is indicated.


Wilson Confidence Interval With True p= 0.005
Wilson Confidence Interval With True p= 0.01
Wilson Confidence Interval With True p= 0.02
Wilson Confidence Interval With True p= 0.03
Wilson Confidence Interval With True p= 0.04
Wilson Confidence Interval With True p= 0.05
Reject H0
Reject K0
p0
p1

(n, S_n, α, β)-Separable

With designated sample size and test statistic S_n, we can "separate" p_0 (and anything above) from p_1 (and anything below) by (n, S_n, α, β).

- By saying p_0 and p_1 are (n, S_n, α, β)-Separable, we mean

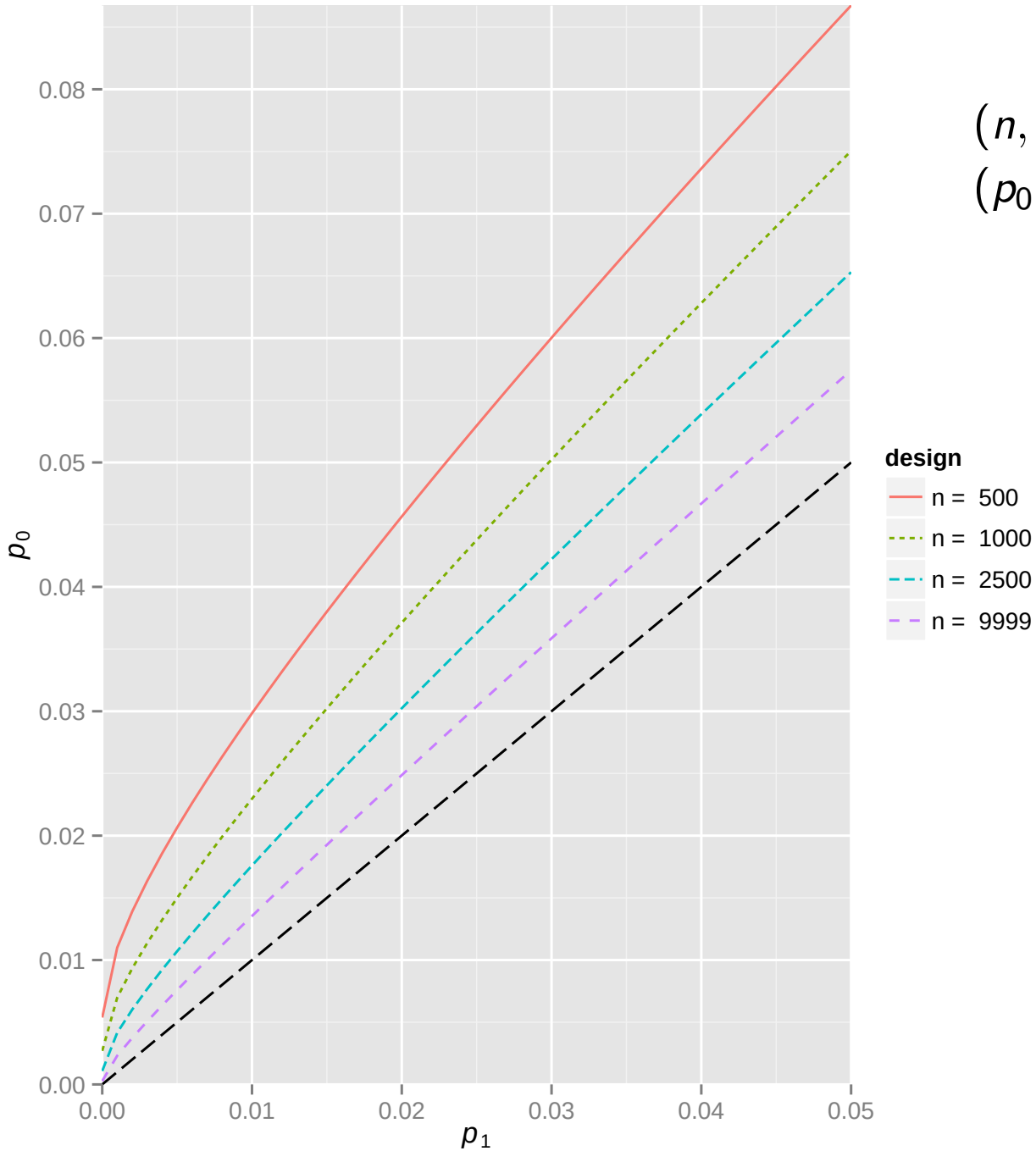
  True $p \leq p_1 \Rightarrow$ able to distinguish it from $[p_0, 1]$

  $i.e.$, significant evidence showing $p < p_0$ (type I error $\leq \alpha$)

  True $p \geq p_0 \Rightarrow$ able to distinguish it from $[0, p_1]$

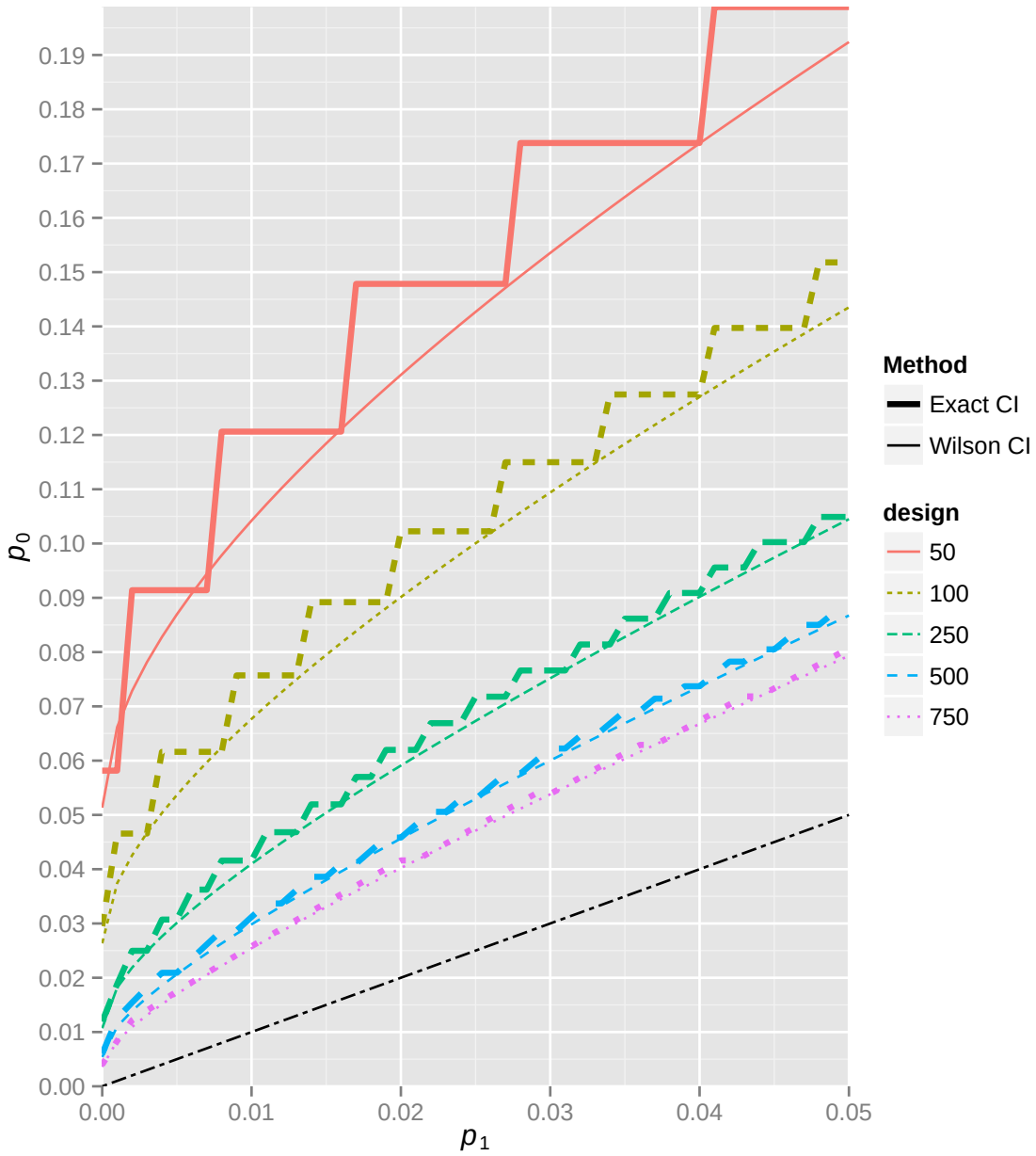
  $i.e.$, significant evidence showing $p > p_1$ (type I error $\leq \beta$)

- With designated sample size n_0 in (4), (p_0, p_1) is $(n_0, S_{n_0}, \alpha, \beta)$-separable. In fact, we can draw a "separation curve" for any generic (n, S_n).



$(n, S_n, 0.05, 0.05)$-separable (p_0, p_1) pairs by sample size n

Exact Binomial vs Wilson CI



$(n, S_n, 0.05, 0.05)$-separable (p_0, p_1) pairs by sample size n

Correlated Binary Data

- Outcomes from the same subjects, where multiple tests are taken.
- Shrinkage percentages of multiple tumors on the same subject.
- Defect rate of televisions from the same factory.
- Correlated binary data introduce intra-cluster correlation (ICC), which has been studied extensively in the literature, e.g., Fisher (1970).
- Designs ignoring the intra-cluster correlation can lead to inflated type I and type II error rates (Cox and Snell, 1989).

Correlated Binary Data

- Suppose there are k tests done on each of the r subjects (total number of tests done is $n = rk$)
- Let $Y_{ij}, Y_{ij'}$ be two of the responses from subject i.

$$Cov(Y_{ij}, Y_{ij'}) = \rho p(1-p), \text{ for } j \neq j'.$$

Central Limit Theorem for Correlated Binary Data

$$\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{\gamma(\rho)p(1-p)}} \xrightarrow{d} N(0,1),$$

where $\gamma(\rho) = 1 + (k-1)\rho$ is the variance inflation factor.

Effect of ICC

Sample Size formula

Suppose correlation is ρ, the sample size n testing (1),

$$n^* = \gamma(\rho)\left(\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_0 - p_1}\right)^2.$$

Let

$$\tilde{n} = \frac{n^*}{\gamma(\rho)},$$

then $\tilde{n}$ is the "effective" sample size.

If $\rho > 0$, the maximum effective sample size is

$$\tilde{n} = \frac{r}{\frac{1-\rho}{k} + \rho} \to \frac{r}{\rho}, \quad k \to \infty.$$

An Example of ICC Effect on Sample Size

Assume $k = 4$, $\alpha = 0.05, \beta = 0.1$.

ρ	r	n
0	50	200
0.2	80	320
0.4	110	440
0.6	140	560
0.8	170	680
1	200	800

Table: Sample size for testing (1) with $p_0 = 0.5$ and $p_1 = 0.3$.

ICC-adjusted Wilson CI

The ICC-adjusted $100 \times (1 - \alpha - \beta)\%$ Wilson CI is given by

$$p \in [\hat{p}_L(\beta; \rho), \hat{p}_U(\alpha; \rho)] \Leftrightarrow z_\alpha \leq \frac{S_n - np}{\sqrt{\gamma(\rho)\, np(1-p)}} \leq z_{1-\beta}.$$

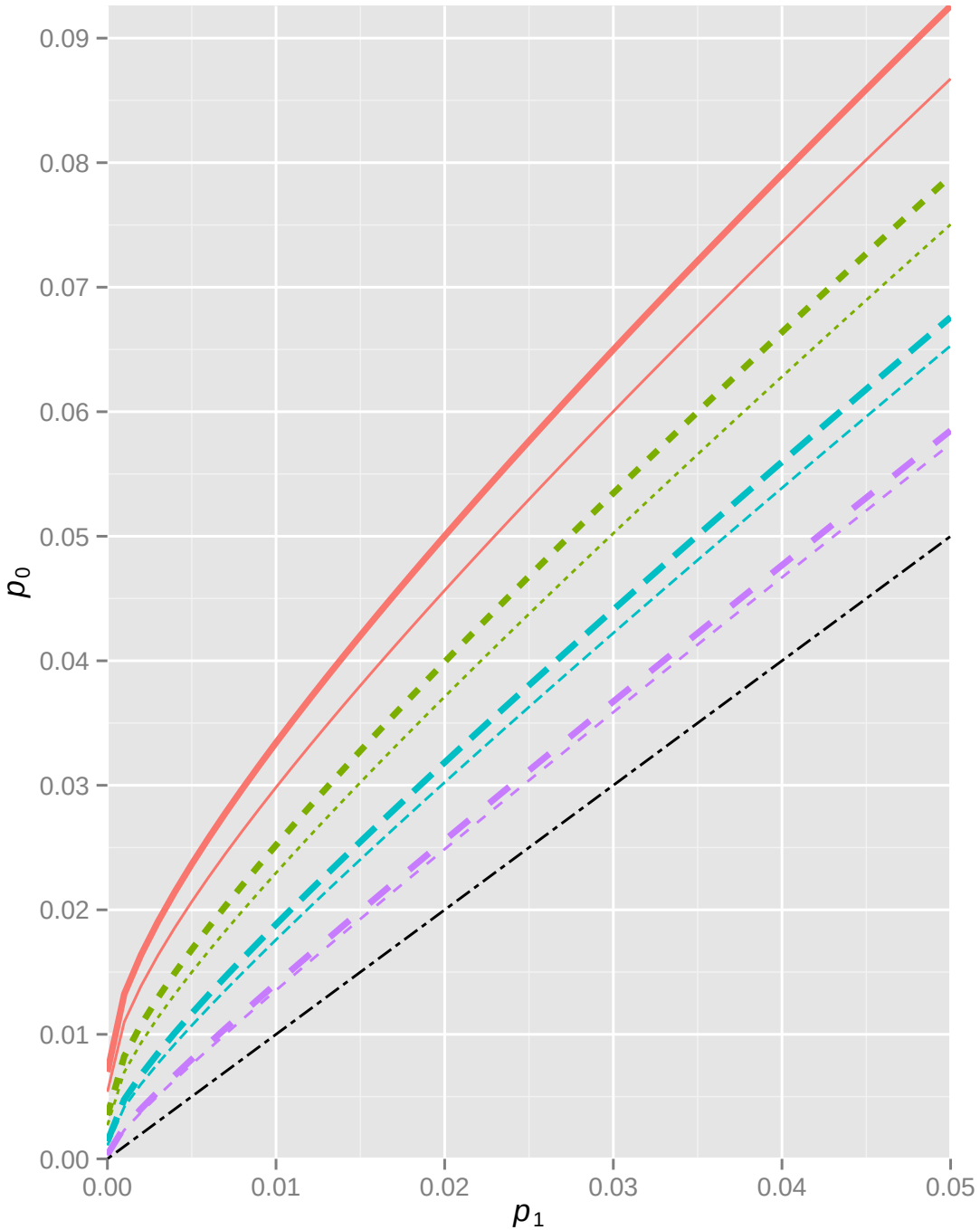
All the aforementioned properties and results for regular Wilson CI hold similarly for the ICC-adjusted Wilson CI.

Estimation of ICC

- Estimation of ICC has been studied extensively in the literature. For example, Ridout et al. (1999) reviewed over 20 different methods.
- Three methods were recommended by Ridout et al. (1999), they are ANOVA estimator, the Pearson pairwise estimator and kappa-type estimator (Zou and Donner, 2004).
- When the "success" rate p is low, the Pearson pairwise estimator is recommended for use (Ridout et al., 1999).

Separation Curve for Correlated Binary Data

- Due to presence of ICC, the variance inflation factor $\gamma(\rho) \geq 1$.
- Separability can be similarly defined by including the component of "variance inflation factor".
- The notation of $(n, S_n, \gamma(\rho), \alpha, \beta)$-separable will be used.



Example: $k = 4, \rho = 0.1$, then $\gamma(\rho) = 1.3$.

The figure shows $(n, S_n, 1.3, 0.05, 0.05)$-separable (p_0, p_1) pairs by sample size n

Entertain the Idea of Separability

- Sample size can be viewed as being determined to separate the null from the alternative.
- The closer two target effects (e.g., p_0 and p_1) are, the harder they can be distinguished.
- More information and better tests give better separability.
- If prior information for p_1 is available, the distinguishable $p_0(p_1; n, \alpha, \beta)$ as a function of p_1 also has a distribution.

Other Types of Data

- For normally distributed data, the separability is sufficiently captured by effect difference, θ^*, such that

$$\frac{\theta^*}{\sigma} = \frac{z_{1-\alpha} + z_{1-\beta}}{\sqrt{n}}.$$

- If we fail to reject $H_0 : \theta \leq 0$ at level α, then we reject $\theta > \theta^*$ at level β. In particular, set $\beta = \alpha$, if we fail to reject H_0, we claim $\theta < 2\sigma z_{1-\alpha}/\sqrt{n}$ at level α.

- For survival data, the separability (for log-rank test) is captured by log hazard ratio, $\log(\mathrm{HR}) = \log(\lambda_1/\lambda_0)$ (or equivalently $\log(m_0/m_1)$, where m_0, m_1 are median survival time) , such that

$$\log(\mathrm{HR}) = -\frac{2(z_{1-\alpha} + z_{1-\beta})}{\sqrt{L}},$$

where L is the number of events.

Summary

- We discussed a go/no-go type of decision-making in the framework of hypothesis testing.
- The equivalence between Wilson CI and hypothesising testing was established.
- A clear-cut decision can be made based on study outcome.
- An investment (e.g., sample size) does not only give us certain power to test a specific hypothesis, but grant us the ability to distinguish different effects.

Sample Size for Binomial Sample Using Exact Distribution

Let $F_n(c \mid p)$ be the cumulative distribution function of $C_n \sim Binomial(n, p)$, then the sample size for testing (1) with type I error α and power $1 - \beta$ is

$$n^* = \min\{n \in \mathbb{N} \mid F_1^{-1}(1 - \beta \mid p_1) \leq F_2^{-1}(\alpha \mid p_0)\},$$

where

$$F_1^{-1}(u \mid p) = \inf\{c \in \mathbb{N} \mid F_n(c \mid p) \geq u\},$$
$$F_2^{-1}(u \mid p) = \sup\{c \in \mathbb{N} \mid F_n(c \mid p) \leq u\}.$$

Separable (p_0, p_1) Using Exact Binomial

$$p_0 = \left\{ 1 + \frac{n - c^*}{(c^* + 1) F(\alpha; 2(c^* + 1), 2(n - c^*))} \right\}^{-1},$$

where

$$c^* = \inf\{c \mid F_n(c \mid p_1) \geq 1 - \beta\},$$

$F(\cdot; \lambda_1, \lambda_2)$ is the CDF of F-distribution with (λ_1, λ_2) degrees of freedom.

Generating Correlated Binary Data

- Suppose $Z, Y_1, \ldots, Y_k$ are independent and identically distributed bernoulli variables with probability of "success" being p.
- Let $U_1, \ldots, U_n$ are independent bernoulli variables with probability of "success" being $\sqrt{\rho}$ and they are also independent of $Z, Y_1, \ldots, Y_k$.
- Define $X_i = (1 - U_i)Y_i + U_i Z$, which is the mixture distribution of Y_i and Z. Then $X_1, \ldots, X_k$ are identically distribution as $Bernoulli(p)$ and $Corr(X_i, X_j) = \rho$.

Finite Sample Correlated Binomial Modeling

- Bahadur (1961) gave the general formula for modeling correlated binary data.
- Kupper and Haseman (1978)

$$\frac{Pr(X=c)}{\binom{n}{c}p^c(1-p)^{n-c}} = 1 + \frac{\rho}{2p(1-p)}\left\{(c-np)^2 + c(2p-1) - np^2\right\}.$$

- Madsen (1993)

$$Pr(X=c) = (1-\rho)\binom{n}{c}p^c(1-p)^{n-c} + \rho[pI\{c=n\} + (1-p)I\{c=0\}].$$

Explicit Form of the ICC-adjusted Wilson CI

$$\tilde{p}_L = \frac{\hat{p}_n + z^2_{1-\alpha/2}/(2\tilde{n}) - z_{1-\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/\tilde{n} + z^2_{1-\alpha/2}/(4\tilde{n}^2)}}{1 + z^2_{1-\alpha/2}/\tilde{n}}, \quad (5)$$

$$\tilde{p}_U = \frac{\hat{p}_n + z^2_{1-\alpha/2}/(2\tilde{n}) + z_{1-\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/\tilde{n} + z^2_{1-\alpha/2}/(4\tilde{n}^2)}}{1 + z^2_{1-\alpha/2}/\tilde{n}}, \quad (6)$$

where

$$\tilde{n} = \frac{n}{\gamma_g(\rho)}.$$

References I

Bahadur, R. R. (1961). A representation of the joint distribution of responses to, n dichotomous items. In *Studies in item analysis and prediction*, pages 158–168, Stanford, Calif. Stanford Univ. Press.

Burman, C.-F., Hamren, B., and Olsson, P. (2005). Modelling and simulation to improve decision-making in clinical development. *Pharmaceutical Statistics*, 4(1):47–58.

Chuang-Stein, C., Kirby, S., French, J., Kowalski, K., Marshall, S., Smith, M. K., Bycott, P., and Beltangady, M. (2011). A quantitative approach for making go/no-go decisions in drug development. *Drug Information Journal*, 45(2):187–202.

Cox, D. and Snell, E. J. (1989). *Analysis of Binary Data*. Chapman and Hall, London, 2 edition.

Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Macmillan Publishing Company.

Fleming, T. R. (1982). One-sample multiple testing procedure for phase ii clinical trials. *Biometrics*, 38(1):pp. 143–151.

References II

Kowalski, K., Ewy, W., Hutmacher, M., Miller, R., and Krishnaswami, S. (2007). Model-based drug developmentâĂŤa new paradigm for efficient drug development. *Biopharmaceutical Report*, 15(2):2–22.

Kupper, L. L. and Haseman, J. K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, 34(1):69–76.

Madsen, R. W. (1993). Generalized binomial distributions. *Communications in Statistics-Theory and Methods*, 22:3065–3086.

Ridout, M. S., DemÃľtrio, C. G. B., and Firth, D. (1999). Estimating intraclass correlation for binary data. *Biometrics*, 55(1):137–148.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.

Zou, G. and Donner, A. (2004). Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. *Biometrics*, 60:807–811.

Thank You!