Bayesian multivariate probability of success with applications to rare diseases

Ethan M. Alt*

September 25, 2020

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

*joint work with Matthew Psioda and Joseph Ibrahim

Introduction

- Increased interest among practitioners in computing the probability of having a successful clinical trial
- Framework: Chuang-Stein (*Pharmaceutical Statistics* 2006)
- Standard methods to compute sample size rely on *statistical power*
  - Power is a *conditional value*
  - Power is *not* the probability of a successful clinical trial
- Probability of success (POS) is the expected value of power with respect to a specified distribution for the effect size:

$$\text{POS} = \int P(\text{Trial meets success critera}|\Delta)p(\Delta|D)d\Delta$$

Case study

- Phase 2 trial of Ivacaftor in subjects with cystic fibrosis (Vertex Pharmaceuticals, 2007-12)
- Primary objectives were to assess safety and dose tolerance, but some efficacy endpoints measured
- Phase 2 data suggested treatment efficacious, but sample size was very small
- Phase 3 trial conducted 2012-15:
  - **Primary endpoint**: Absolute change from baseline in percent predicted forced expiratory volume in 1 Second (FEV1)
  - **Key secondary endpoints:**
    - Change from baseline in sweat chloride
    - Change from baseline in Cystic Fibrosis Questionnaire-Revised (CFQ-R) respiratory domain score
- **Key question:** How can we exploit this phase 2 data to find a sample size that yields a high probability of success in a Phase 3 trial across the endpoints?

Seemingly Unrelated Regression (SUR)

SUR Model

$$
\begin{aligned}
\boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i \\
\boldsymbol{u}_i &\overset{\text{ind}}{\sim} N_J(\boldsymbol{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} \in \mathbb{R}^{J\times J} \\
\boldsymbol{y}_i &= (y_{i1}, \ldots, y_{iJ})' \in \mathbb{R}^J \\
\boldsymbol{X}_i &= \text{blkdiag}\left\{\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iJ}'\right\} \in \mathbb{R}^{J\times p}, \qquad p = \textstyle\sum_{j=1}^J p_j \\
\boldsymbol{\beta} &= (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')' \in \mathbb{R}^p
\end{aligned}
$$

- Most general multivariate normal linear model
- Allows each response to have its own set of covariates

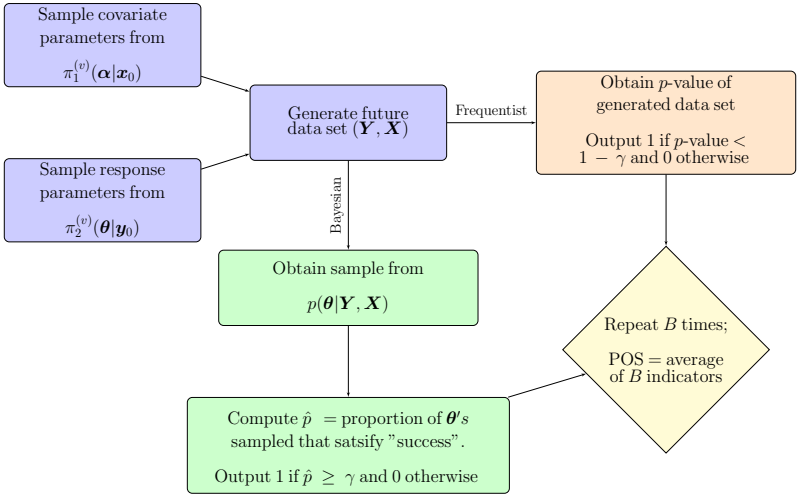
Copula regression

Gaussian copula regression model

$$\boldsymbol{z}_i \sim N_J(0, \boldsymbol{\Gamma})$$
$$y_{ij} = F_j^{-1}(\Phi(z_{ij})|\boldsymbol{\theta}, \boldsymbol{x}_{ij})$$

- $F_j(\cdot)$ is the CDF for the j^{th} margin
- If the j^{th} margin is a GLM, $\boldsymbol{\theta} = (\beta', \tau)$, where $\tau = 1$ in some cases
- $\boldsymbol{x}_{ij}$ is a vector of covariates for subject i and endpoint j, $i = 1, \ldots, n$, $j = 1, \ldots, J$
- Explicitly models the correlation between endpoints of possibly mixed types
- Equivalent to SUR model if all endpoints are normally distributed

POS Algorithm



A note on small historical data sets

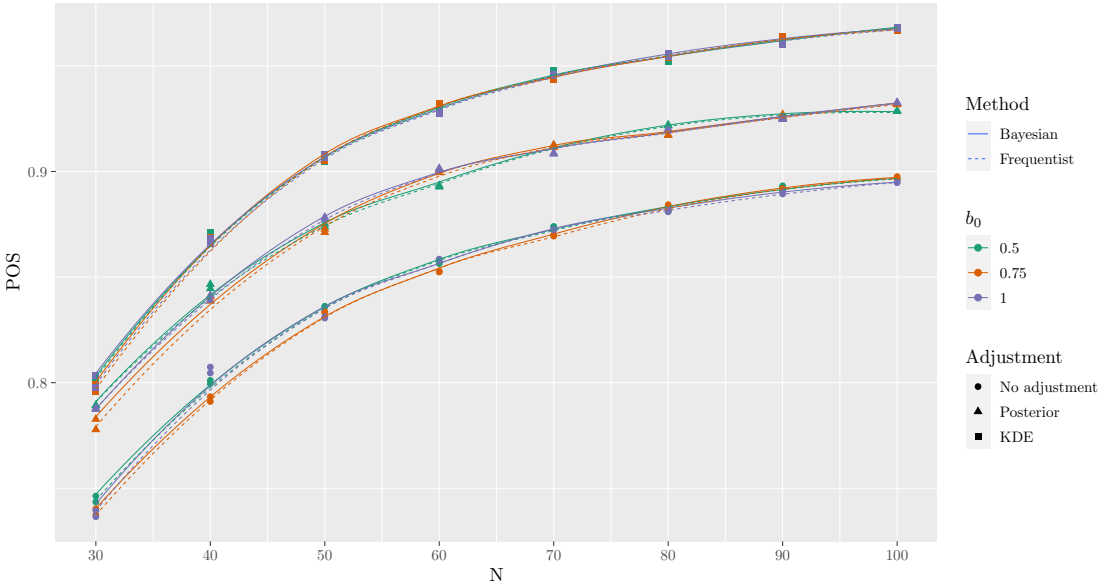
- For rare diseases, Phase II sample sizes are typically very small
- Implausible treatment effects may be sampled if the sample size is too low
- There are several possible adjustments one can make:
  1. Restrict samples for efficacy in the treatment effect: *Bayesian conditional expected power*
  2. Use informative priors for treatment effects
  3. Restrict samples of treatment effects to the q^{th} highest posterior density (HPD) region for some $0 < q < 1$.
- We focus on (3) and propose two different mechanisms:
  1. HPD region of all parameters
  2. HPD region of only treatment effects estimated using KDE

The simulated trial

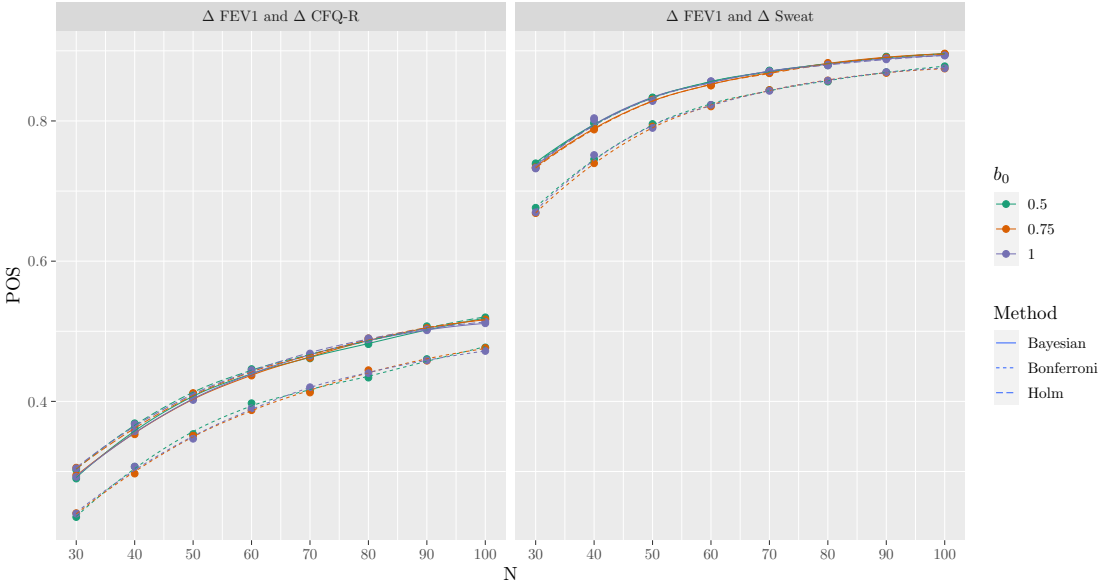
- $n_0 = n_1 = 8$
- $E(\Delta y_{ij}) = \beta_{0j} + \beta_{1j} z_i + \boldsymbol{x}_{ij}' \beta_{2j}$
- $\beta_1 = (6.4, 3.5, -49.1)'$
- $\text{diag}(\Sigma) = (5.12, 7.05, 12.27)'$
- $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, -0.25, -0.33)$
- For all outcomes, $\boldsymbol{x}_{ij}$ includes an intercept term, linear and square terms of age, weight, BMI, and sex
- Baseline levels for FEV1 and CFQ-R score are controlled for in their regressions
- Power computations yield $\tilde{n}_1 = 22$, $\tilde{n}_2 = 176$, and $\tilde{n}_3 = 14$ for 90% power

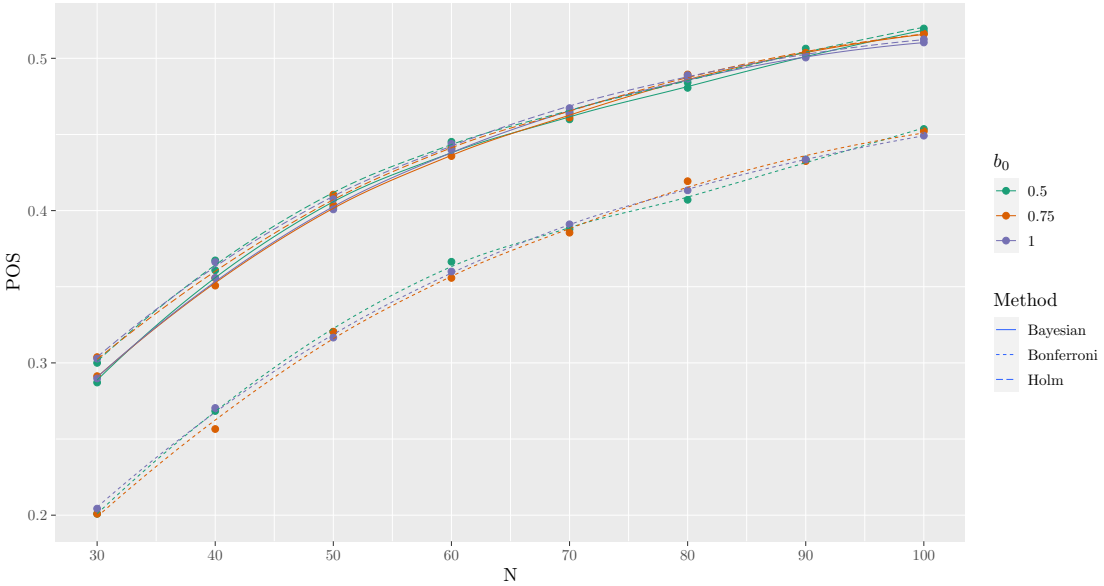
POS simulation

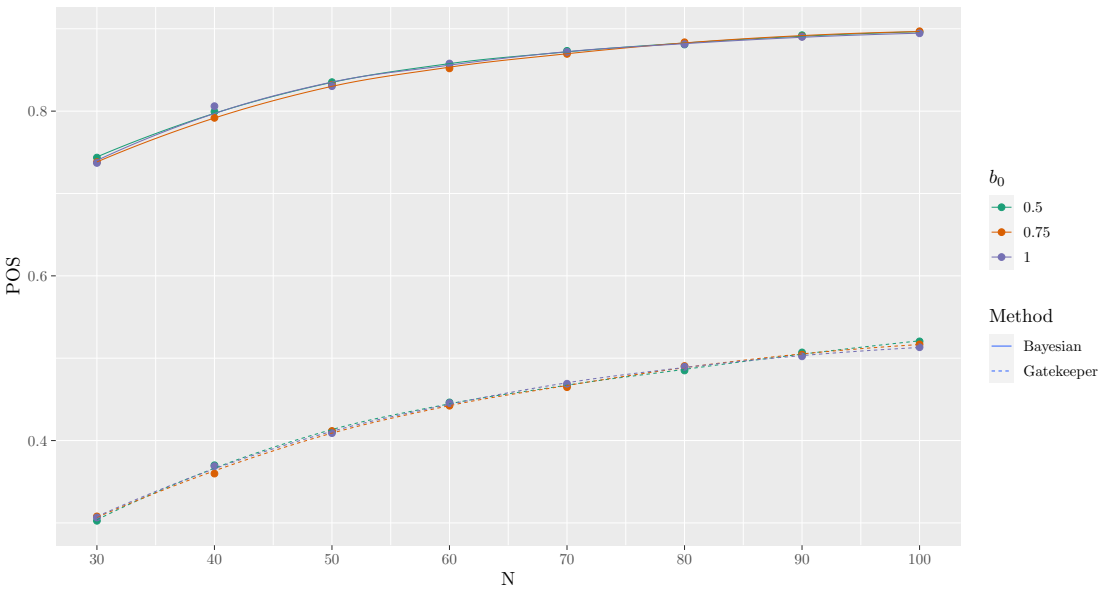
- The notion of "success" can be generalized to a set
- $$\Omega_1 := \{\boldsymbol{\theta} \mid \beta_{11} > 0\}$$
  $$\Omega_2 := \{\boldsymbol{\theta} \mid \beta_{12} < 0\}$$
  $$\Omega_3 := \{\boldsymbol{\theta} \mid \beta_{13} > 0\}$$
- $\Omega_{1j} := \Omega_1 \cap \Omega_j$, $j = 1, 2$
- $\Omega_{123} = \Omega_1 \cap \Omega_2 \cap \Omega_3$ (complete success)
- $\Omega_{1|23} = \Omega_1 \cap (\Omega_2 \cup \Omega_3)$


POS for Δ FEV1
POS
0.9
0.8
30
40
50
60
70
80
90
100
N
Method
Bayesian
Frequentist
b_0
0.5
0.75
1
Adjustment
No adjustment
Posterior
KDE

POS for Primary and one Secondary Endpoint




POS for Complete Success
POS
0.2
0.3
0.4
0.5
30
40
50
60
70
80
90
100
N
b_0
0.5
0.75
1
Method
Bayesian
Bonferroni
Holm


POS for Primary and ≥ 1 Secondary
POS
0.8
0.6
0.4
30
40
50
60
70
80
90
100
N
b_0
0.5
0.75
1
Method
Bayesian
Gatekeeper

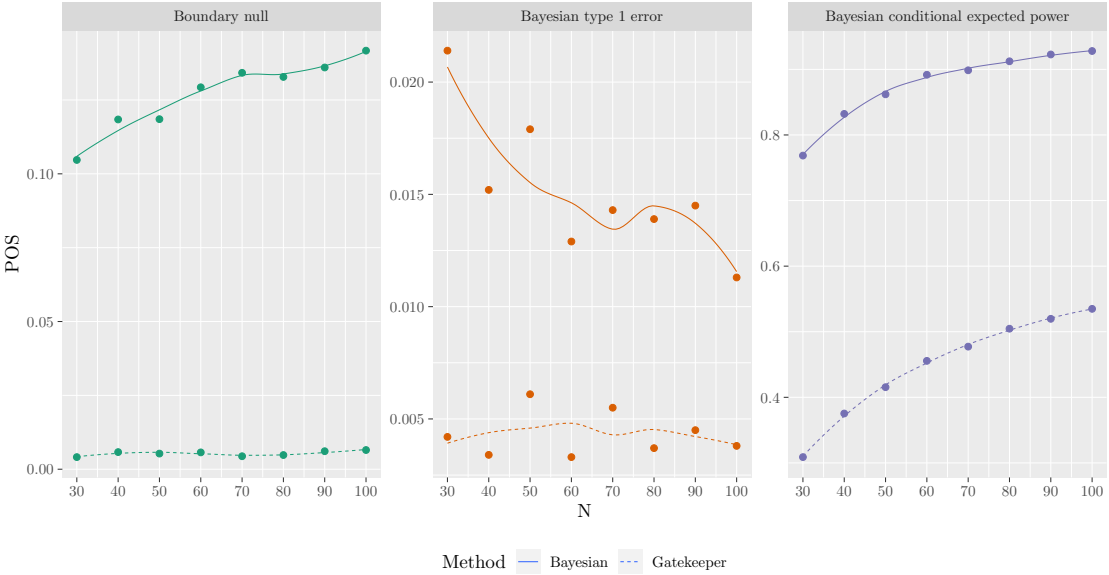
Power and Type I Error Rate

- In order to understand the increase in success, we must characterize the type I error rate of the method
- Amend the posterior samples of the historical data

$$\tilde{\pi}^{(v)}(\theta|D_0) \propto \pi^{(v)}(\theta|D_0) I(\theta \in \Theta) \quad (1)$$

- Three important Θ's:
  1. Boundary null:
     $\Theta_B = \{\theta|\beta_{12} = 0 \cap \beta_{13} = 0\}$ (for example)
  2. Bayesian conditional expected power:
     $\Theta_{H_1} = \{\theta|\beta_{11} > 0 \cap \{\beta_{12} < 0 \cup \beta_{13} > 0\}\}$
  3. Bayesian type I error:
     $\Theta_{H_0} = \overline{\Theta}_{H_1} = \{\theta|\{\beta_{11} \leq 0\} \cup \{\beta_{12} \geq 0 \cup \beta_{13} \leq 0\}\}$

Power and Type I Error for Primary and ≥ 1 Secondary



Conclusion

- Only POS method that considers multiple endpoints
- Fully Bayesian approach to jointly model multivariate LMs and GLMs
- Unifies hypothesis testing into one framework
- Interpretation far simpler than frequentist methods
- Useful whether a frequentist or Bayesian analysis will be used
- Methods of Ibrahim et al. and Chaung-Stein can be seen as special cases
- R packages available on CRAN: `surbayes` and `bayescopulareg`