MDG: Bayesian trial design

Duncan T. Wilson

Clinical Trials Research Unit
Leeds Institute of Clinical Trials Research
University of Leeds

d.t.wilson@leeds.ac.uk

18th July 2017

Overview

Previously...

- Last time, we talked about using the limited data provided by a pilot trial, in conjunction with prior knowledge, to **learn about the ICC** and use this knowledge to **better design the main trial**.
- How do we design trials **to test hypotheses**, using full distributions describing unknown parameters as opposed to point estimates or hypothesised values?
- Only after this has been described can we think about optimal pilot trial design.

Motivation

Power for a trial with clustering:

$$1 - \beta = \Phi\left(\sqrt{\frac{n\delta^2}{2\sigma^2[1+(m-1)\rho]}} + z_{\alpha/2}\right),$$

$$\begin{aligned}
n &= \text{number of patients per arm} \\
m &= \text{number of patients per cluster} \\
\delta &= \text{treatment effect} \\
\sigma_B^2 + \sigma_W^2 = \sigma^2 &= \text{total variance} \\
\frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \rho &= \text{intracluster correlation coefficient (ICC)}
\end{aligned}$$

Motivation

Power is defined with respect to some hypothesis H_1,

$$\text{Power} = Pr[\text{reject } H_0 \mid H_1],$$

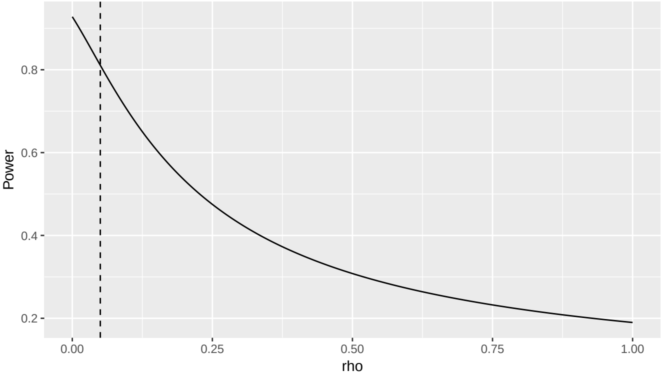
where H_1 specifies the values of all the parameters in our model:

$$H_1 : \delta = \delta^*, \sigma = \sigma^*, \rho = \rho^*.$$

We never know the true value of ρ. How do we choose ρ^*, and what are the implications of our choice?

Motivation

For a fixed sample size n, a higher than expected ICC will lead to a lower than expected power:



Motivation

For a fixed target power, a higher ICC will lead to a higher sample size requirement:

$$
\begin{aligned}
n &= n_i[1 + (m - 1)\rho^*] \\
n_i &= \text{sample size required when clustering is ignored}
\end{aligned}
$$

So, required sample size increases linearly with our chosen ρ^*.

Practical solution

Example (SHIFT)

"We anticipate that the level of clustering will be low for this particular trial - possibly around 0.01 but no higher than 0.05 . . . assuming an ICC of 0.05 effectively reduces the power from 90% to around 75%. If the ICC were as low as 0.01, then this reduces the power to around 85%."

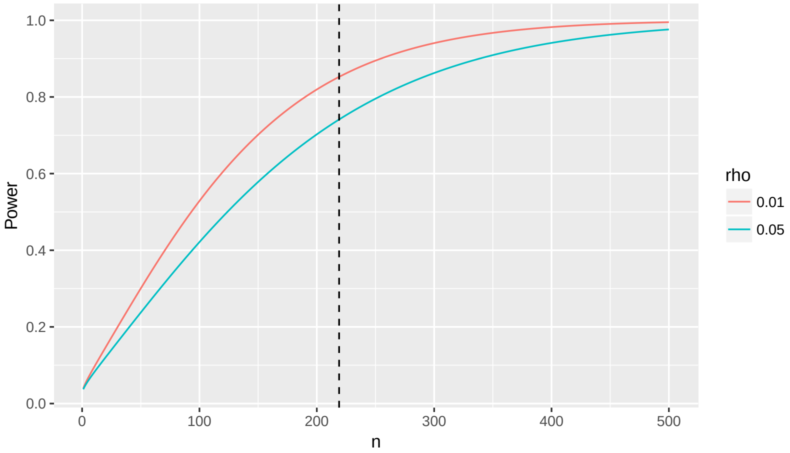
Practical solution

Trial design now considers another hypothesis:

$$
\begin{aligned}
&\text{Minimise sample size } n \\
&\text{Minimise type I error rate } \mid H_0 \\
&\text{Maximise power } \mid H_1 : \rho = 0.01 \\
&\text{Maximise power } \mid H_2 : \rho = 0.05
\end{aligned}
$$

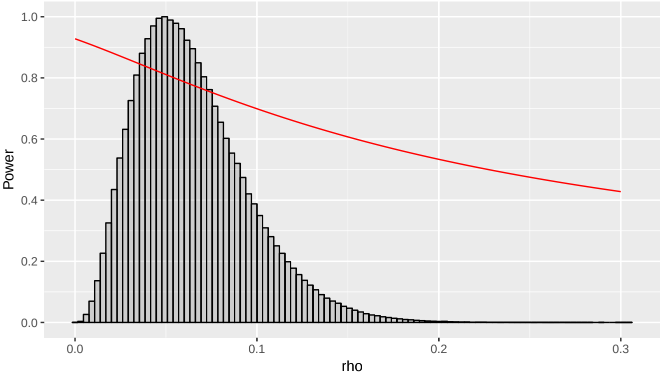
We down-weight the importance of H_2 compared to H_1 because we *believe it to be less likely*, but we do this *informally*.

Practical solution



Bayesian alternative

An alternative: express our belief about the true value of ρ using a probability distribution, e.g. $\rho \sim Beta(4, 58)$:



Bayesian alternative

Should we use probability to express our uncertainty? Axioms[1]:

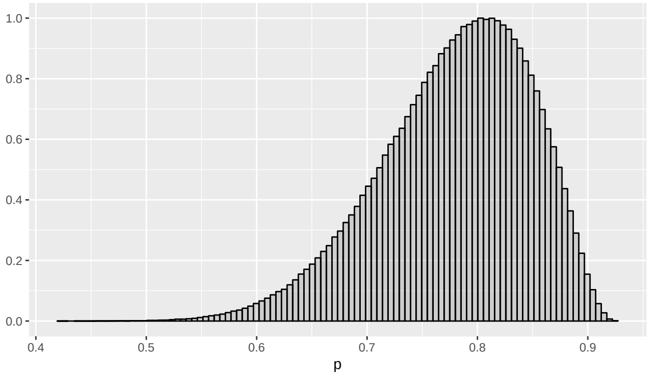
1. The relation $\succeq$, where $A \succeq B$ means A is at least as likely as B, is a weak ordering (i.e. complete and transitive) on S.
2. If A, B and D are such that $AD = BD = \emptyset$, then $A \succeq B \iff A \cup D \succeq B \cup D$.
3. $A \succeq \emptyset$ for any A. Furthermore, $S \succ \emptyset$.
4. If $A_1 \supset A_2 \supset \dots$ is a decreasing sequence of events and B is such that $A_i \succeq B$ for all i, then $\cap_{i=1}^{\infty} A_i \succeq B$.
5. There exists a random variable $X \sim Unif(0, 1)$.

Together, these imply that there is a unique probability distribution P that agrees with the relation $\succeq$.

[1] DeGroot 1970.

Bayesian alternative

Uncertainty about $\rho \sim Beta(4, 58)$ propagates to uncertainty about power:



Assurance

Key paper: O'Hagan, Stevens and Campbell (2005), "Assurance in clinical trial design", *Pharmaceutical statistics*.

- If event A denotes a 'successful' trial, and this is influenced by the true value of parameter(s) θ, what is $Pr[A]$?
- Unconditional, so we need to integrate out any parameters.

$$
\begin{aligned}
\text{assurance} = Pr[A] &= \int Pr[A \mid \theta] p(\theta) d\theta \\
p(\theta) &= \text{probability density function of } \theta
\end{aligned}
$$

Assurance

Example (Unknown treatment effect)

For a two-sample comparison of normal means, suppose we are uncertain what the true treatment difference δ is. We think $\delta \sim N(0.2, 0.25^2)$, and know that the common variance is $\sigma^2 = 1$. Our trial has $n = 500$ in each arm, and will test at a level of $\alpha = 0.025$ (one-sided). What is $Pr[\text{reject } H_0]$?

1. Sample $\delta^i \sim N(0.2, 0.25^2)$ for $i = 1, \dots, N$
2. Calculate powers $\beta(\delta^i)$
3. Compute the Monte Carlo estimate:

$$Pr[\text{reject } H_0] \approx \frac{1}{N} \sum_{i=1}^{N} \beta(\delta^i) = 0.62. \quad (1)$$

Assurance

Example (Unknown treatment effect)

For a two-sample comparison of normal means, suppose we are uncertain what the true treatment difference δ is. We think $\delta \sim N(0.2, 0.25^2)$, and know that the common variance is $\sigma^2 = 1$. Our trial has $n = 500$ in each arm, and will test at a level of $\alpha = 0.025$ (one-sided). What is $Pr[\text{reject } H_0]$?

- Compare the unconditional power of 0.62 with the conditional power at $\delta^* = 0.2$ of 0.88.
- Is this good enough? Should we adjust the sample size?

Optimality

Given some measure(s) of the quality of a proposed trial design, how do we choose the best?

- Try to set some kind of standard threshold analogous to a power of 80%.
- Avoid thresholds and consider trade-offs[2];
- Maximise performance per-patient[3];
- Maximise expected utility[4], possibly through an economic model[5];

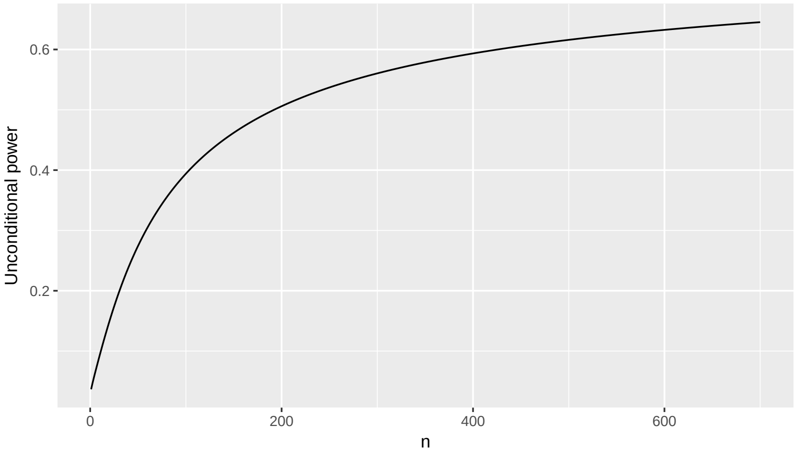
[2] Bacchetti, McCulloch, and Segal 2008.

[3] Stallard 2012.

[4] Lindley 1997.

[5] Patel et al. 2013.

Optimality



Assurances

The unconditional probability of rejecting the null hypothesis was the first hybrid Bayesian metric to be considered, but others have been proposed:

- Prob. rejecting the null *and* the true effect being meaningful[6];
- Prob. rejecting the null in *two* separate phase III trials[7];
- Prob. of a positive result in an updated meta-analysis[8].

'Fully Bayesian' criteria have also been proposed:

- Prob. posterior probability the treatment effect is meaningful[9];
- Conditional distribution of the treatment effect given a significant result[10].

[6] O'Hagan, Stevens, and Campbell 2005.

[7] Zhang and Zhang 2013.

[8] Sutton et al. 2007.

[9] Ibrahim et al. 2014.

[10] Walley et al. 2015.

Uncertainty

Various papers have addressed uncertainty in ...

- Treatment effects only[11];
- Nuisance parameters only, including ICCs[12] and parameters in survival models[13];
- Samples size[14];
- Adherance rates[15].

For trials of complex interventions following a pilot we might also expect uncertainty in recruitment and data collection rates[16].

[11] O'Hagan, Stevens, and Campbell 2005.

[12] Turner, Toby Prevost, and Thompson 2004.

[13] Ren and Oakley 2013.

[14] Ambrosius and Mahnken 2010.

[15] Fay, Halloran, and Follmann 2006.

[16] Avery et al. 2017.

Priors

How do we define the priors to be used in these calculations?

- Using early phase trial data (empirical Bayes)[17];
- Using other historical data, perhaps from a meta-analysis[18];
- Using expert judgement[19];
- Using defaults, e.g. pessimistic and optimistic priors[20];
- If using a fully Bayesian approach, consider differentiating between 'design' and 'analysis' priors[21];

---

[17] Jiang 2011.

[18] Sutton et al. 2007.

[19] Ren and Oakley 2013.

[20] Spiegelhalter, Abrams, and Myles 2004.

[21] Wang and Gelfand 2002.

Computation

- Some papers focus on conjugacy[22];
- The more flexible approach is through simulation[23], sampling test statistics if possible or failing that, individual patient data;
- Bayesian (unconditional) simulation is no trickier than frequentist (conditional) simulation;
- But simulation is time-consuming - if the design problem is complex, we need efficient optimisation methods.

[22] Ibrahim et al. 2014.

[23] Wang et al. 2013.

Interpretation

- Can we have distributions for nuisance parameters but condition on treatment effects?[24]
- Can we have multiple priors?[25]
- Can we interpret a distribution of treatment effects from a frequentist perspective?[26]

[24] Turner, Toby Prevost, and Thompson 2004.

[25] Chen et al. 2011; Kirby et al. 2012.

[26] Gtte et al. 2015; Julious and Owen 2006.

Challenges

- Complex models, e.g. recruitment, missing data, adherance, multilevel outcomes, multivariate outcomes[27]
- Complex design problem, e.g. multilevel sample sizes, multivariate acceptance regions, choice of primary endpoint
- Qualitative gap between pilot and main trial settings, including changes to the intervention itself[28]
- Computational burden of searching for optimal pilot designs[29]

[27] Landau and Stahl 2013.

[28] Hampson et al. 2017.

[29] Strong et al. 2015.

Challenges

- Complex models, e.g. recruitment, **missing data**, adherance, multilevel outcomes, multivariate outcomes[30]
- Complex design problem, e.g. multilevel sample sizes, multivariate acceptance regions, **choice of primary endpoint**
- Qualitative gap between pilot and main trial settings, including changes to the intervention itself[31]
- Computational burden of searching for optimal pilot designs[32]

[30] Landau and Stahl 2013.

[31] Hampson et al. 2017.

[32] Strong et al. 2015.

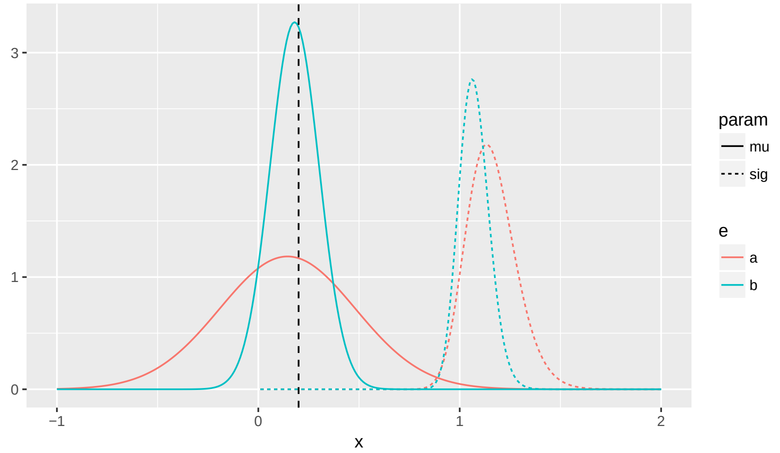
Example

Example (REACH(ish))

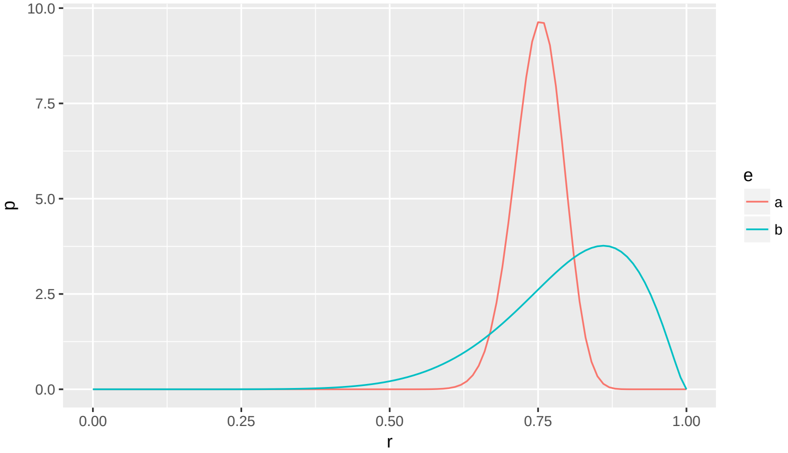
One of the goals of out pilot study was to assess two potential primary outcomes for the main trial. We weren't sure about what treatment effects we might see, the variability in the outcomes, or how complete the data are likely to be. Following the pilot, a Bayesian analysis gives us posterior distributions on the relevant parameters.

Endpoint	$\hat{\delta}$	$\hat{\sigma}^2$	$\hat{p}_{miss}$
A	0.145	1.16	0.75
B	0.18	1.08	0.8

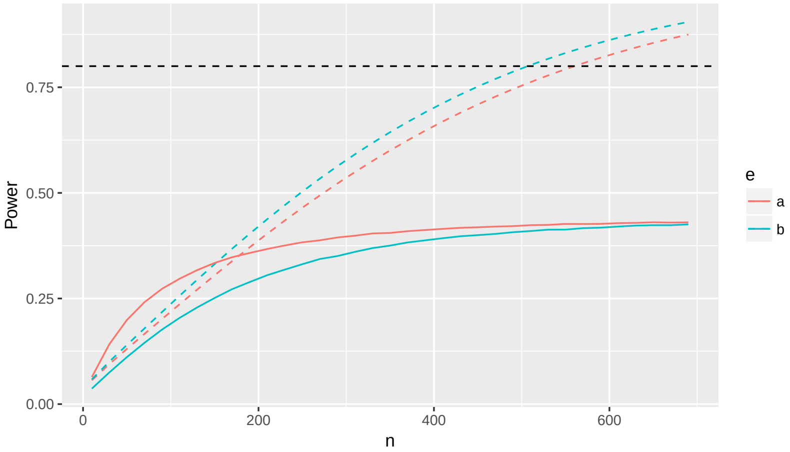
Example



Example



Example



510 → 370 patients per arm.

Ambrosius, Walter T. and Jonathan D. Mahnken (2010). "Power for studies with random group sizes". In: *Statistics in Medicine*, n/a–n/a. DOI: 10.1002/sim.3873. URL: https://doi.org/10.1002%2Fsim.3873.

Avery, Kerry N L et al. (2017). "Informing efficient randomised controlled trials: exploration of challenges in developing progression criteria for internal pilot studies". In: *BMJ Open* 7.2, e013537. DOI: 10.1136/bmjopen-2016-013537. URL: https://doi.org/10.1136%2Fbmjopen-2016-013537.

Bacchetti, Peter, Charles E. McCulloch, and Mark R. Segal (2008). "Simple, Defensible Sample Sizes Based on Cost Efficiency". In: *Biometrics* 64.2, pp. 577–585. DOI: 10.1111/j.1541-0420.2008.01004_1.x. URL: http://dx.doi.org/10.1111/j.1541-0420.2008.01004_1.x.

Chen, Ming-Hui et al. (2011). "Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data". In: *Biometrics* 67.3, pp. 1163–1170. DOI: 10.1111/j.1541-0420.2011.01561.x. URL: https://doi.org/10.1111%2Fj.1541-0420.2011.01561.x.

DeGroot, Morris H. (1970). *Optimal Statistical Decisions*.

Fay, Michael P., M. Elizabeth Halloran, and Dean A. Follmann (2006). "Accounting for Variability in Sample Size Estimation with Applications to Nonadherence and Estimation of Variance and Effect Size". In: *Biometrics* 63.2, pp. 465–474. DOI: 10.1111/j.1541-0420.2006.00703.x. URL: https://doi.org/10.1111%2Fj.1541-0420.2006.00703.x.

Gtte, Heiko et al. (2015). "Sample size planning for phase II trials based on success probabilities for phase III". In: *Pharmaceutical Statistics* 14.6, pp. 515–524. DOI: 10.1002/pst.1717. URL: https://doi.org/10.1002%2Fpst.1717.

Hampson, Lisa V et al. (2017). "A framework for prospectively defining progression rules for internal pilot studies monitoring recruitment". In: *Statistical Methods in Medical Research* 0.0. PMID: 28589752, p. 0962280217708906. DOI: 10.1177/0962280217708906. eprint: http://dx.doi.org/10.1177/0962280217708906. URL: http://dx.doi.org/10.1177/0962280217708906.

Ibrahim, Joseph G. et al. (2014). "Bayesian probability of success for clinical trials using historical data". In: *Statistics in Medicine* 34.2, pp. 249–264. DOI: 10.1002/sim.6339. URL: https://doi.org/10.1002%2Fsim.6339.

Jiang, Kaihong (2011). "Optimal Sample Sizes and Go/No-Go Decisions for Phase II/III Development Programs Based on Probability of Success". In: *Statistics in Biopharmaceutical Research* 3.3, pp. 463–475. DOI: 10.1198/sbr.2011.10068. URL: https://doi.org/10.1198%2Fsbr.2011.10068.

Julious, Steven A. and Roger J. Owen (2006). "Sample size calculations for clinical studies allowing for uncertainty about the variance". In: *Pharmaceutical Statistics* 5.1, pp. 29–37. DOI: 10.1002/pst.197. URL: http://dx.doi.org/10.1002/pst.197.

Kirby, S et al. (2012). "Discounting phase 2 results when planning phase 3 clinical trials". In: *Pharmaceutical statistics* 11.5, pp. 373–385. DOI: 10.1002/pst.1521.

Landau, Sabine and Daniel Stahl (2013). "Sample size and power calculations for medical studies by simulation when closed form expressions are not available". In: *Statistical Methods in Medical Research* 22.3, pp. 324–345. DOI: 10.1177/0962280212439578. eprint: http://smm.sagepub.com/content/22/3/324.full.pdf+html. URL: http://smm.sagepub.com/content/22/3/324.abstract.

Lindley, Dennis V. (1997). "The choice of sample size". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 46.2, pp. 129–138. ISSN: 1467-9884. DOI: 10.1111/1467-9884.00068. URL: http://dx.doi.org/10.1111/1467-9884.00068.

O'Hagan, Anthony, John W. Stevens, and Michael J. Campbell (2005). "Assurance in clinical trial design". In: *Pharmaceutical Statistics* 4.3, pp. 187–201. ISSN: 1539-1612. DOI: 10.1002/pst.175. URL: http://dx.doi.org/10.1002/pst.175.

Patel, Nitin R et al. (2013). "A mathematical model for maximizing the value of phase 3 drug development portfolios incorporating budget constraints and risk". In: *Statistics in medicine* 32.10, pp. 1763–1777. DOI: 10.1002/sim.5731.

Ren, Shijie and Jeremy E. Oakley (2013). "Assurance calculations for planning clinical trials with time-to-event outcomes". In: *Statistics in Medicine* 33.1, pp. 31–45. DOI: 10.1002/sim.5916. URL: https://doi.org/10.1002%2Fsim.5916.

Spiegelhalter, David J., Keith R. Abrams, and Jonathan P. Myles (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley.

Stallard, Nigel (2012). "Optimal sample sizes for phase II clinical trials and pilot studies". In: *Statistics in Medicine* 31.11-12, pp. 1031–1042. ISSN: 1097-0258. DOI: 10.1002/sim.4357. URL: http://dx.doi.org/10.1002/sim.4357.

Strong, Mark et al. (2015). "Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: a fast, nonparametric regression-based method". In: *Medical Decision Making* 35.5, pp. 570–583.

Sutton, Alexander J. et al. (2007). "Evidence-based sample size calculations based upon updated meta-analysis". In: *Statistics in Medicine* 26.12, pp. 2479–2500. DOI: 10.1002/sim.2704. URL: https://doi.org/10.1002%2Fsim.2704.

Turner, Rebecca M., A. Toby Prevost, and Simon G. Thompson (2004). "Allowing for imprecision of the intracluster correlation coefficient in the design of cluster randomized trials". In: *Statistics in Medicine* 23.8, pp. 1195–1214. ISSN: 1097-0258. DOI: 10.1002/sim.1721. URL: http://dx.doi.org/10.1002/sim.1721.

Walley, Rosalind J. et al. (2015). "Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study". In: *Pharmaceutical Statistics* 14.3, pp. 205–215. DOI: 10.1002/pst.1675. URL: http://dx.doi.org/10.1002/pst.1675.

Wang, Fei and Alan E. Gelfand (2002). "A Simulation-Based Approach to Bayesian Sample Size Determination for Performance under a Given Model and for Separating Models". English. In: *Statistical Science* 17.2, pp. 193–208. ISSN: 08834237. URL: http://www.jstor.org/stable/3182824.

Wang, Yanping et al. (2013). "Evaluating and utilizing probability of study success in clinical development". In: *Clinical Trials* 10.3, pp. 407–413. DOI: 10.1177/1740774513478229. URL: https://doi.org/10.1177%2F1740774513478229.

Zhang, Jianliang and Jenny J. Zhang (2013). "Joint probability of statistical success of multiple phase III trials". In: *Pharmaceutical Statistics* 12.6, pp. 358–365. DOI: 10.1002/pst.1597. URL: https://doi.org/10.1002%2Fpst.1597.