The Incorporation of Prior Information Using Bayesian Statistics in Medical Device Clinical Trials

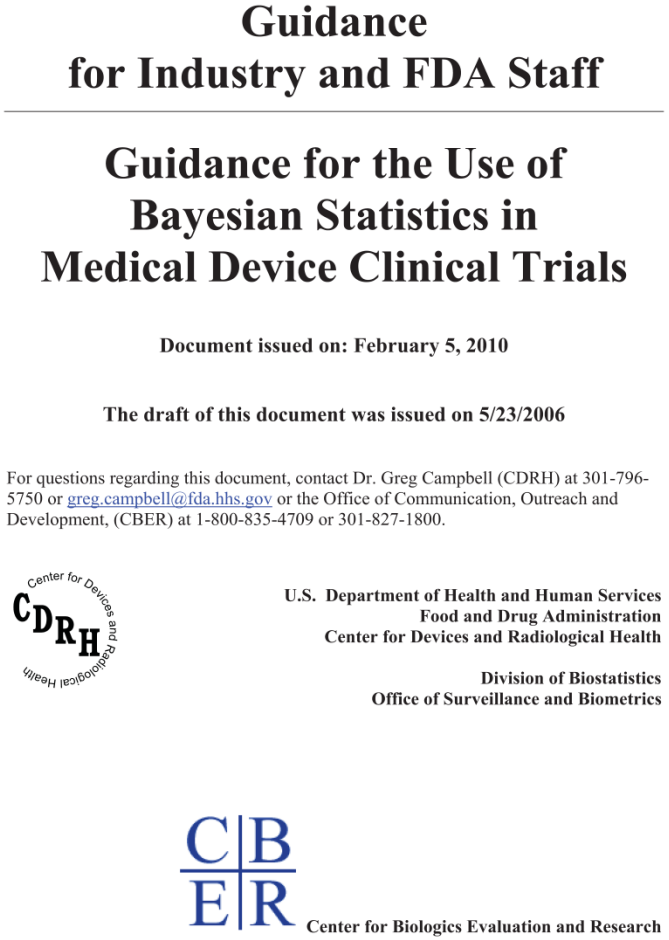
Gregory Campbell, Ph.D.
President, GCStat Consulting
GCStat@verizon.net

Regulatory-Industry Statistics Workshop September 23, 2020

Bayesian Statistics at FDA

- In 1998 CDRH signaled to industry that it was willing to consider Bayesian submissions for medical device applications.
- Devices often have prior information that can be leveraged and the mechanism of action is local not systemic and physical not pharmacokinetic.
- Conferences were held in 1998 and 2004 and a draft guidance issued in 2005.
- Goal is not to "lower the bar" but to arrive at the same decision in a more timely manner

FDA Bayesian Guidance



Guidance
for Industry and FDA Staff

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

Document issued on: February 5, 2010

The draft of this document was issued on 5/23/2006

For questions regarding this document, contact Dr. Greg Campbell (CDRH) at 301-796-5750 or greg.campbell@fda.hhs.gov or the Office of Communication, Outreach and Development, (CBER) at 1-800-835-4709 or 301-827-1800.

Center for Devices and Radiological Health
CDRH

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

CBER
Center for Biologics Evaluation and Research

- Finalized February 5, 2010.
- http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm

Statistical Philosophy

- Frequentist versus Bayesian
- Decision rule based on posterior distribution rather than p-values
- Confidence interval versus credible interval
- American Statistical Association statement on Statistical Significance and p-values followed by 43 articles in *American Statistician* in 2019
- p-value of 0.05 for Null Hypothesis Significance Testing (NHST) and reproducibility in science
- Bayesian focus on estimation (using the entire posterior distribution) and not just hypothesis testing

Bayesian Statistics: Submissions to CDRH

- At least 30 Original PMAs and PMA Supplements have been approved with a Bayesian analysis as primary.
  - The Supplements include stent systems, a heart valve, and spinal cage systems.
- Many IDEs (planned designs) have also been approved.
- Several applications for "substantial equivalence" (510(k)s)
- A list of publicly available submissions as of 2010 is in:

  Campbell G. (2011). Bayesian statistics in medical devices: innovation sparked by the FDA. *J. Biopharm Stat* **21**:871-887.

Two Bayesian Approaches in the FDA Guidance

1. Bayesian Hierarchical Modeling using data from one or more prior studies, including power priors
2. Bayesian Adaptive designs, usually with non-informative priors
   - Usually these designs are Bayesian group sequential
   - The BIG advantage here is to model the primary outcome in terms of intermediate endpoints.

Hierarchical Bayesian Modeling

- Use a hierarchical model to place a usually non-informative prior at the highest level of the hierarchy
  - For example, consider a number of past studies and the current one, each with different numbers of patients and assume that the patients within a study are exchangeable and the studies are exchangeable among each other.
  - Place a (non-informative) prior to reflect the distribution of the studies.
  - This model borrows strength adaptively from past studies to model the current study.

Exchangeability

- If all patients and all the studies were exchangeable, you could pool all the data together (complete exchangeability).
- But there is variability from study to study so assume studies are exchangeable and that only patients within each study are exchangeable (and not patients between studies).
- Exchangeable studies mean the next study is no more likely to be better (or worse) than its predecessors. (It does not require that the data are all i.i.d. (independent and identically distributed) across studies)
- Exchangeability is not reasonable if the new device is expected to perform better.
- Evaluation of prior information and exchangeability requires non-statistical input.

Hierarchical Bayesian Modeling

Level 2: Study 1 Study 2 Study 3 New Study

Level 1: $y_{11},\ldots,y_{1n_1}$ $y_{21},\ldots,y_{2n_2}$ $y_{31},\ldots,y_{3n_3}$ y_{new}

Level 1: Patients (y) exchangeable within studies

$$y_j | \theta_j, \phi \sim P(y_j | \theta_j, \phi)$$

Level 2: Studies exchangeable within patient populations

$$\theta_j | \phi \sim P(\theta_j | \phi)$$

Level 3: Prior $\pi(\phi)$ (usually non-informative)

Effective Sample Size

- For a current one-arm study of size N and a prior distribution, the effective sample size is

$$\text{EffSS} = N \;\; V(\text{Posterior}) / \text{Var}(\text{no borrowing})$$

and the amount borrowed is EffSS – N.

Malec D. (2001). *Statistics in Medicine* **20**(12):1811-1824.

Hierarchical Bayesian Example of "Borrowing Strength"

A Hypothetical Coronary Stent Trial

- Outcome: Prop. of patients that have target vessel failure
- Prior distribution: size 250 of which 50 are TVF failures
  - $p_p = 50/250 = 0.20$
- Randomized study: treatment sample size 500, observe the proportion of failures:
  - $p_t = x/500$
- Aim: Estimate the failure rate in the treatment arm, using the prior distribution to improve the estimate. Call this estimate (the mean of the posterior dist.) p_B

Malec (2001)

Example (cont.)

- Inference
  - Suppose $p_t = 0.170$. Then $p_B = 0.177$. If we accepted the prior as equivalent evidence, the weighted average is

    $$[(0.17)500+(0.20)250]/750 = 0.18$$

  - The 95% posterior probability interval is (0.147, 0.207).
  - In contrast, the 95% *de novo* confidence interval is (0.140, 0.206).

Malec (2001)

Example (cont.)

- Sample Size Adjustment
  - The Bayesian interval is 10% shorter. It is equivalent to an "effective sample size" of over 600 with no prior information; if the target size was 600, the savings of using prior information is 100 patients.
  - The effective sample size is defined in terms of the ratio of the variances times the target sample size.
- If p_t is far away from prior, then very little if any borrowing. The model adaptively (dynamically) borrows.

Negative Borrowing

- "Figure 1 in Malec (2001) provides some insight into how the hierarchical model works. If the proportion is near 0.20 there is a lot of borrowing but very rarely as much as complete poolability, almost up to the entire additional 250. If the proportion is less than 0.08 or greater than 0.34, then there is no borrowing but if the proportion is say 0.12 (or 0.28), just far enough away, there is negative borrowing, meaning the model needs to overcome the incorrect prior and can result in an effective sample size that is less than 400 whereas the sample size with no borrowing is 500."

Campbell G (2017). *Communications for Statistical Applications and Methods* **24**(6):561-581.

Hierarchical Model within a Study

- In advance, plan to build a hierarchical model that identifies the partition into subgroups within a study that can be modeled. These subgroups are assumed to be exchangeable. These can then "borrow" from each other or "gain strength".
- This is quite similar to a random effects model (for example, for multiple centers) where the estimates shrink toward the mean depending on how much borrowing there is between groups.

Pediatric Trials

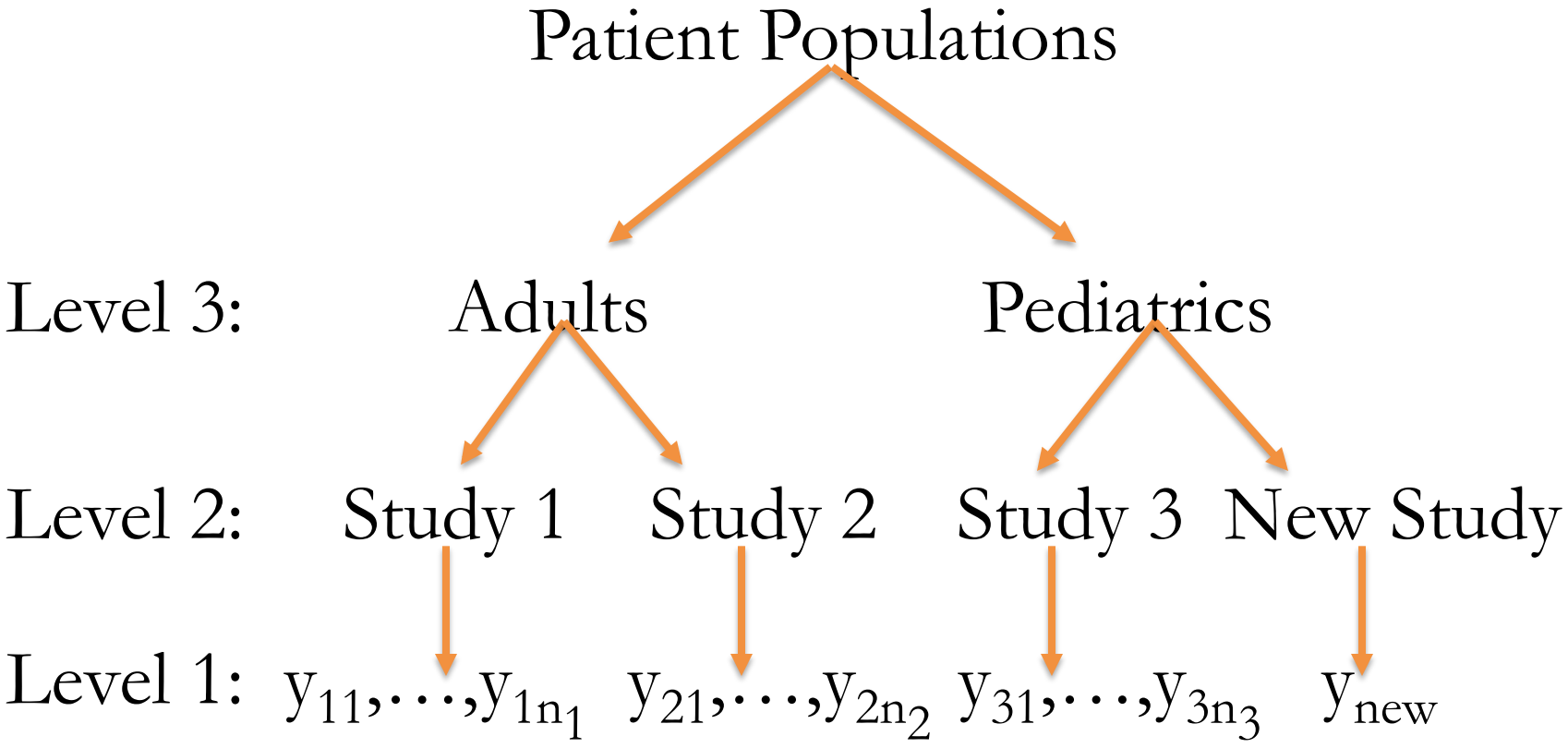
- Clinical trials for pediatric populations are difficult.
- (A response adaptive design is one possible approach, to minimize the inferior treatment, like ECMO.)
- Use Bayesian hierarchical modeling to "borrow strength" from
  - Other pediatric studies in US and overseas
  - Other adult studies

  in either
  - Control arm
  - Experimental arm (or both)

FDA Guidance on Pediatric Extrapolation

- FDA Guidance: Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices (issued June, 2016)
  - https://www.fda.gov/regulatory-information/search-fda-guidance-documents/leveraging-existing-clinical-data-extrapolation-pediatric-uses-medical-devices
- Response to Pediatric Medical Device Safety and Improvement Act of 2007
- Appendix on Bayesian Hierarchical Modeling

Pediatric Hierarchical Model



- Level 1: Patients (y) exchangeable within studies
- Level 2: Studies exchangeable within patient populations
- Level 3: Patient populations are exchangeable

Limitations of Hierarchical Modeling

- If the new study is thought to be better than the historical ones, then exchangeability is not reasonable.
- However if the prior studies are for the control group and the practice of medicine has not changed then it would seem reasonable.
- If there are only one or two prior studies, the model will try to estimate the study variance but it will be unstable and can be highly sensitive to the parameterization of the non-informative prior at the top of the hierarchy.

Power Priors for Historical Data

Conditional power prior:

$$\pi(\theta \mid D_0, \alpha_0) \propto L(\theta \mid D_0)^{\alpha_0} \pi_0(\theta)$$

•Parameter α_0 is the exponent for the likelihood based on the prior data and indicates a range of borrowing from none ($\alpha_0=0$) to all ($\alpha_0=1$), where the latter is complete patient-level exchangeability.

Posterior distribution q is:

$$q(\theta \mid D_0, D, \alpha_0) \propto \pi_0(\theta) L(\theta \mid D_0)^{\alpha_0} L(\theta \mid D)$$

Power Prior

- A simple idea is to agree to a fixed value for α_0 say 0.5. This would mean borrowing 50% of the prior.
- The Bayesian approach is to place a prior on α_0 and combine with the current data to calculate the posterior for α_0
  - Ibrahim and Chen (2000), Neelon and O'Malley (2010)

Commensurate Power Prior for Historical Data

- Commensurate Power Prior introduce a parameter τ that is a measure of how commensurate the current data are with the historical data.
- Conditional prior $\pi^{\text{CPP}}(\theta, \alpha_0, \tau \mid D_0)$
  - For a location parameter θ, the commensurate parameter τ can be a precision parameter with a prior distribution updated to the posterior

Hobbs, Carlin, Mandrekar and Sargent (2011), Hobbs, Sargent, Carlin (2012)

Adaptive Borrowing

- For both Bayesian hierarchical models and for Bayesian power priors, the amount of borrowing depends on how close the current study is to the prior.
- Bayesian models with prior information are adaptive.

Bayesian Adaptive Designs

- Campbell, G. (2014). Similarities and Differences of Bayesian Designs and Adaptive Designs for Medical Devices: A Regulatory View. *Statistics in Biopharmaceutical Research*. **5**(4): 356-368.
- Guidance for Industry and Food and Drug Administration Staff: Adaptive Designs for Medical Device Clinical Studies. Issued July, 2016

  https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-designs-medical-device-clinical-studies

Adaptive Clinical Study Designs for Medical Devices

- The Bayesian experience has been most helpful for adaptive designs generally.
- CDRH has seen about 250 adaptive submissions from 2007 to 2013, including most Bayesian ones.

  Yang, X., Thompson, L., et al. (2016). Adaptive design practice at CDRH, January 2007-May 2013. *Therapeutic Innovation and Regulatory Science* **50**(6):710-717.

- Campbell G. (2013). Similarities and differences of Bayesian designs and adaptive designs for medical devices: a regulatory view. *Statist. Biopharm. Research* **5**:356-364.
- Guidance for Industry and Food and Drug Administration Staff: Adaptive Designs for Medical Device Clinical Studies (Issued July 27, 2016) https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf

The Importance of Simulation

- It is important for FDA to understand the frequentist operating characteristics of the Bayesian submissions.
- Why? The Type 1 error probability (or some analog of it) protects the U.S. public from FDA approving products that are ineffective or unsafe.
- So simulate to show that Type 1 error (or some analog of it) is well-controlled.
- The use of prior information can inflate the Type I error
- Simulations can also be of help in estimating the approximate size of the trial and the strategy of interim looks. Usually Bayesian studies are not studies with a fixed size.

Trends

- More prospectively done studies—planned and executed as Bayesian studies (not a switch from frequentist to Bayesian)
- More adaptive trials without prior information, using the accumulating data in the trial to make preplanned decisions, and relatively fewer trials that rely on external prior information. (This may be a reflection of the realization that a Bayesian adaptive trial is a methodological leap forward and that exchangeability can be difficult.)
- All Bayesian trials require a lot of planning and a lot of work but can be quite beneficial.

Thank you for your attention!