POINT-COUNTERPOINT

Bayesian Approach to Establish Bioequivalence: Why and How?

Carl C. Peck[1,2] and Gregory Campbell[3,4]

A generic drug that is legally substitutable for its brand name predecessor drug (reference-listed drug (RLD)) is expected to be therapeutically equivalent (TE). Bioequivalence (BE) is a core requirement for TE and is established by statistical analysis of data in a small pharmacokinetic trial. Below, we discuss some concerns with the current statistical procedure used in this analysis and propose evaluation of a robust Bayesian approach to mitigate them.

DEFINITIONS AND CHALLENGES

BE is defined by an "absence of a significant difference in the rate and extent to which the active ingredient ... becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study." For systemically acting oral drugs, this requirement is usually satisfied in a small, healthy volunteer, crossover pharmacokinetic study of the generic and RLD, comparing their respective bioavailability (BA) properties. As the basis for approval of a new generic drug, the BE results are successful "surrogate" clinical trial end points, area under the curve (AUC), and maximum drug concentration (C_{max}) for the safety and efficacy end points of the RLD.

Given reliance on BE of a generic product, it is clear that conceptually, "prior" knowledge of the link of clinical safety and efficacy of the RLD to the BA observed at the time of its US Food and Drug Administration (FDA) approval (BA-0) plays a crucial role in concluding TE. Thus, generic drug approval can be viewed qualitatively as "Bayesian," in the sense that the newly observed BA of the RLD is assumed to be substitutable, in this case with BA-0.

Interestingly, the contemporary regulatory and statistical framework for evaluation of BE ignores the BA-0 data and relies solely on contemporaneous evaluation of the BE of an off-the-market-shelf formulation of the RLD.

EVOLUTION OF THE CURRENT STATISTICAL FRAMEWORK FOR BE

According to Skelly,[1] during the 1970s several statistical hypothesis testing procedures were proposed for BE studies, including the "Canadian±20% rule," "Power Approach," and "75/75" rule. In the 1980s, the FDA rejected Bayesian procedures in favor of the frequentist "average BE" "two one-sided test" (TOST) procedure of Schuirmann[2,3] as a test of the equivalence of the ratios of generic and RLD average AUC and C_{max}. The FDA's criteria for inferring BE require rejection of the two hypotheses that these ratios lie outside of the predetermined BE limits, usually from 0.8 to 1.25.[3] Alternatively, average BE is concluded when the 90% confidence intervals (CIs) of the ratios of the average AUC and C_{max} for the generic drug compared with the RLD each fall within the BE limits.[3] In TOST, the statistical method with prespecified type 1 error rate (one-sided $P < 0.05$ for each of the two one-sided hypothesis tests), usually with power of 80–90%, is based on the untestable assumption of long-run repetition of the BE trial.

INVESTIGATIONS AND SUGGESTIONS FOR BAYESIAN APPROACHES FOR BE

During the same period and continuing until today, statisticians have advanced procedures for using a bayesian framework for BE. For explanation of Bayesian approaches to clinical trials in contrast to non-Bayesian frequentist approaches, see the FDA's Guidance for the Use of Bayesian Statistics in Medical Device

[1]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California, USA;[2] NDA Partners LLC, San Luis Obispo, California, USA; [3]GCStat Consulting, LLC, Silver Spring, Maryland, USA; [4]NDA Partners LLC, Silver Spring, Maryland, USA. Correspondence: Carl C. Peck (ccpeck@icloud.com)

Received August 13, 2018; accepted October 11, 2018; advance online publication January 22, 2019. doi:10.1002/cpt.1288

Clinical Trials.[4] A good source of references for the early work is in Selwyn and Hall[5] using noninformative priors to generate the posterior distribution for the probability of BE rather than the frequentist symmetric CIs of BE. More important, Selwyn and Hall[5] emphasized the value of sensitivity analysis of competing prior distribution assumptions. Racine-Poon *et al.*[6] described a Bayesian adaptive procedure for efficient, sample size – sparing identification of non-BE drug formulations being tested during drug development. Ghosh and Gönen[7] described a semiparametric Bayesian analysis procedure for simultaneous multivariate (AUC and C_{max}) average BE assessment, using a Dirichlet process mixture prior, in which case it can be advantageous to put an informative prior on the correlation of AUC and C_{max}. Recently, Longford[8] presented the thesis that application of formal decision theory, informed by Bayesian or frequentist analytics, is more appropriate than hypothesis testing for evaluating drug effects in clinical trials, including evaluation of average BE.

CHALLENGES AND CAVEATS OF THE TOST

The TOST of Schuirmann[2] has several challenges and caveats. The first is that it is solely a hypothesis test that provides no other information than whether the hypothesis is rejected or not. Furthermore, it does not provide a direct estimate of the probability of BE, much less an estimate of its entire distribution. Although TOST is equivalent to a procedure that constructs a CI and then checks to see it is enclosed in the equivalence bounds, these statistical inferences assume that the distributions of the individual AUC and C_{max} are Normally distributed either before or after a logarithmic transformation. However, the number of subjects in BE trials is often small (12 or 24 is not uncommon), so the Central Limit Theorem does little to mitigate this strong assumption. A not uncommon problem is that of outliers; outliers threaten the Normality assumption and can dramatically inflate variance estimates. If the design is a parallel one, there are additional concerns about whether the variances in the two groups are equal or the sample sizes are similar. Last, TOST's resulting *P*-values rely on the assumed long-term repetition of the trial for its Normal theory–based type 1 error control rather than simulation.

PROPOSAL FOR A ROBUST BAYESIAN APPROACH FOR BE—CHALLENGES AND CAVEATS

Notwithstanding the challenge of proposing a Bayesian procedure in place of the traditional frequentist TOST analysis of BA data as a regulatory standard, an alternative BE statistical analysis that may mitigate the challenges of the TOST is a Bayesian procedure proposed by Kruschke.[9] We call this procedure "BE-BEST" (BE-Bayesian Estimation, adapted from Kruschke's[9] more general acronym, "BEST"). Using noninformative priors, the information output of BE-BEST exceeds that of TOST. In contrast to the TOST acceptance/rejection hypothesis tests of BE, the BE-BEST procedure enables checking whether the "credible interval" (Bayesian equivalent of CI) falls within the BE limits (e.g., region of practical equivalence, Kruschke and Liddell[10]) and can provide a direct assessment of the probability of the average BE, which can better ensure reproducibility. BE-BEST is robust to the Normality assumption of TOST and hence less sensitive to outliers by using a mixture model. Thus, BE-BEST can enable analysis of untransformed BE data, without the need to stabilize variance and Normalize the BA data via the FDA-recommended logarithmic transformation of the BE data. Computationally, the R Package BEST (https://cran.r-project.org/web/packages/BEST/index.html) is readily accessible and provides rapid computation of "Highest Density" (or symmetric) credibility intervals, model diagnostics, and plots. A drawback is that this software is not validated. The BE-BEST procedure provides rich information about the credibility intervals, distributions of BE ratios and SDs, and deviation from Normality of the data, which may offset perceptions of greater difficulty of the Bayesian approach.

If skeptically warranted, the BE-BEST procedure can also use informative prior distributions, which may broaden its applicability to sample size-sparing, formulation-bridging studies during development of a new or generic drug. As in best-practices employment of the Bayesian approach for any clinical trial, intense pretrial planning, and precise prespecification of the analysis plan, including transformations and the treatment of outliers, sensitivity analytics, and trial simulations under various design assumptions should be undertaken to ensure type 1 error control and adequacy of statistical power.

DISCUSSION

Acknowledging successful generic drug approval in many cases, use of the TOST procedure has limitations among which are the failure to provide a direct estimate of the probability of BE and reliance on vulnerable frequentist assumptions enumerated above. Shifting from the frequentist statistical framework for BE to a Bayesian method using noninformative priors, such as BE-BEST, offers robust solutions to all of these limitations. More importantly, the Bayesian framework presented herein opens the door to related novel applications, such as the following: (i) efficient employment in formulation-bridging studies, (ii) evaluation of BA "drift" of RLD off-the-market-shelf BA from BA-0, (iii) use in clinical efficacy trials other than for BE, (iv) evaluations of non-BE products in comparative BA studies for 505(b)(2) development programs, and (v) consideration of Bayesian decision analysis for regulatory approval. In the Bayesian BE approach, usually no influential prior information is used, and instead, noninformative priors are chosen for the parameters of interest. Simulations are required to ensure that the operating characteristics of the procedure are well controlled and understood, including the type I error probability and power for important alternatives. We propose that the BE-BEST procedure be further investigated for BE and BA bridging applications as an alternative to TOST.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

Dr Peck is Chairman of and consults for NDA Partners LLC. Dr Campbell consults for GCStat Consulting LLC, and NDA Partners LLC.

1. Skelly, J.P. A history of biopharmaceutics in the Food and Drug Administration 1968-1993. *AAPS J* **12**, 44–50 (2010).
2. Schuirmann, D.J. A comparison of the two-one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* **15**, 657–680 (1987).
3. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Guidance for industry: statistical approaches to establishing bioequivalence (2001).https://www.fda.gov/downloads/drugs/guidances/ucm070244.pdf.
4. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health. Guidance for industry and FDA staff: guidance for the use of Bayesian statistics in medical device clinical trials (2010). https://www.fda.gov/MedicalDevices/ucm071072.htm.
5. Selwyn, M.R. & Hall, N.R. On Bayesian methods for bioequivalence. *Biometrics* **40**, 1103–1108 (1984).
6. Racine-Poon, A., Grieve, A.P., Flühler, H. & Smith, A.F. A two-stage procedure for bioequivalence studies. *Biometrics* **43**, 847–856 (1987).
7. Ghosh, P. & Gönen, M. Bayesian modeling of multivariate average bioequivalence. *Statist. Med.* **15**, 2402–2419 (2008).
8. Longford, N.T. Comparing two treatments by decision theory. *Pharm. Stat.* **15**, 387–395 (2016).
9. Kruschke, J.K. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).
10. Kruschke, J.K. & Liddell, T.M. Bayesian data analysis for newcomers. *Psychon. Bull. Rev.* **25**, 155–177 (2018).