OPEN ACCESS Freely available online

Journal of Clinical Toxicology

Commentary

Advancing Bioequivalence Assessment: A Bayesian Method in the Presence of Non-Lognormality and Extreme Values

Jing Wang[1], Meng Hu[1], Somesh Chattopadhyay[2], Lanyan Fang[1], Carl Peck[3,4*]

[1]Division of Quantitative Methods and Modeling, Office of Research and Standards, Office of Generic Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA; [2]Division of Biometrics VIII, Office of Biostatistics, Office of Translational Science, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA; [3]Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, CA, USA; [4]NDA Partners, a ProPharma Group Company, Washington DC, USA

DESCRIPTION

The Two One-Sided T-tests (TOST) procedure remains the regulatory standard for average Bioequivalence (BE), yet it may lead to inaccurate conclusions when assumptions of lognormality and/or absence of extreme data values are violated in practice. We compared TOST with a t-distribution-based, non-informative Bayesian approach (BayesT) on 2,341 pharmacokinetic datasets from 678 anonymized Abbreviated New Drug Applications (ANDAs). Lognormality was rejected in 26.8% of datasets and 17.5% contained extreme log(T/R) values. TOST and BayesT agreed (on BE or non-BE) in 98.9% of cases. When TOST assumptions were satisfied, performance was comparable. However, in datasets characterized by non-lognormality and extreme values, TOST rejected BE while BayesT concluded BE, reflecting greater robustness of the Bayesian approach to heavy-tailed distributions and extreme values. Simulation results of truly BE datasets showed improved accuracy and sample size efficiency for BayesT under assumption violations for TOST. These findings support consideration of pre-specifying BayesT analysis when TOST distributional assumptions are uncertain.

MAIN TEXT

The regulatory standard for evaluating average Bioequivalence (BE) has long relied on the frequentist, Two One-Sided T-tests (TOST) procedure [1]. While TOST has served as a practical and widely accepted BE analysis procedure, its validity depends on assumptions that are not always met in real-world data-most notably, lack of lognormality of the Test/Reference (T/R) ratios and influential extreme values are not uncommon.

Peck C, et al., showed that in simulations a t-distribution-based, non-informative Bayesian procedure ($Bayes_T$, derived from BEST [2]), compared favorably to the TOST in rendering accurate determinations of BE in truly BE datasets. $Bayes_T$ more accurately concluded BE or not BE, while requiring fewer subjects, when observed log (T/R) or T-R were normal and extreme values were present [3].

Preceded by extensive simulation analyses validating $Bayes_T$ against TOST [3], we evaluated both procedures in 2,341 pharmacokinetic datasets from 678 anonymized Abbreviated New Drug Applications (ANDAs) [4]. Our findings support consideration of pre-specifying $Bayes_T$ when distributional assumptions are uncertain.

ANDA BE data characteristics

Among the evaluated ANDA datasets:

- 26.8% rejected lognormality by Shapiro-Wilk testing
- 17.5% contained extreme log(T/R) values (identified using the interquartile range (IQR) method [5] with 3×IQR as the criterion [4])

These findings indicate that deviations from distributional assumptions for TOST are not uncommon in submitted BE data. Importantly, extreme values were not merely statistical curiosities; they materially influenced BE inference in certain cases.

Agreement and disagreement between tost and $Bayes_T$

Overall, TOST and $Bayes_T$ agreed in BE or non-BE conclusions for 98.9% of cases. When T/R data were lognormal and free of extreme values, both methods agreed on BE or not BE. This broad agreement reinforces that TOST remains appropriate under its assumed conditions.

However, the 1.1% of cases in which the methods disagreed were informative:

Correspondence to: Carl Peck, Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, CA, USA, E-mail: ccpeck@icloud.com

Received: 15-Jun-2026, Manuscript No. JCT-26-42543; **Editor assigned:** 17-Jun-2026, PreQC No. JCT-26-42543 (PQ); **Reviewed:** 24-Jun-2026, QC No. JCT-26-42543; **Revised:** 01-Jul-2026, Manuscript No. JCT-26-42543 (R); **Published:** 15-Jul-2026, DOI: 10.35248/2475-3181.26.16.621

Citation: Wang J, Hu M, Chattopadhyay S, Fang L, Peck C (2026). Advancing Bioequivalence Assessment: A Bayesian Method in the Presence of Non-Lognormality and Extreme Values. J Clin Toxicol. 16:621.

- 0.3% (7 cases): TOST concluded BE while $Bayes_T$ marginally rejected BE. These cases were all near the equivalence boundaries, with minimal interval differences. All these 7 cases didn't reject lognormality, and no extreme values were present.
- 0.8% (20 cases): TOST rejected BE while $Bayes_T$ concluded BE. All such cases failed lognormality testing, and 85% contained extreme values

Note that the 1.1% incidence of ANDAs submitted to FDA reported in [4] is probably an underestimate of that encountered in real-world BE studies not submitted to FDA, since sponsors limit such submissions to BE-likely datasets.

The authors also showed that removing extreme values substantially narrowed TOST confidence intervals and often changed TOST's BE determinations - especially in disagreement cases, whereas after extreme value removals, $Bayes_T$ credible interval widths and BE conclusions remained robust [3,4]. A representative example [4] showed that a single extreme reference value caused the TOST upper bound to exceed 1.25 (1.66), whereas the Bayesian credible interval remained within range (1.24), leading to divergent BE conclusions.

These findings do not suggest that TOST is inappropriate in general. Rather, they indicate that when its core assumptions are materially violated-particularly in the presence of extreme values-its inferences may become sensitive to data irregularities.

Relationship to the 2026 FDA draft Bayesian guidance

The FDA's January 2026 draft guidance on Bayesian methodology [6] provides an important regulatory context for these findings. The guidance discusses:

- Use of non-informative and minimally informative priors
- Posterior probability–based success criteria
- Evaluation of operating characteristics
- Sensitivity analyses and documentation standards

Although the draft guidance does not specifically address bioequivalence, it outlines a general framework in which Bayesian methods may be used to support primary inference when appropriately justified and pre-specified.

The $Bayes_T$ approach used in our analysis aligns with these principles:

- It employs non-informative priors, allowing the observed data to dominate inference in a robust way when the sample size is adequate.
- It relies on posterior credible intervals analogous to confidence interval criteria.
- It can be fully pre-specified with operating characteristics evaluated simulation

Thus, while our work was motivated by methodological considerations in BE analysis, it is also consistent with broader regulatory interest in Bayesian frameworks.

Broader implications beyond bioequivalence

Although our published study focused on bioequivalence datasets, the methodological implications extend beyond BE.

Many pharmacokinetic and clinical datasets in drug development may:

- Deviate from assumed parametric distributions
- Contain extreme or influential observations
- Exhibit small sample sizes that challenge asymptotic approximations
- Yield inaccurate estimations and inferences when analyzed by frequentist methods

In such settings, a t-distribution-based Bayesian framework provides greater robustness to heavy-tailed distributions, stability in the presence of outliers, and transparent and more direct probabilistic interpretation through credible intervals.

Accordingly, while BE provides a clear and structured test case, the underlying principles may be applicable more broadly across pharmaceutically relevant analyses when distributional assumptions are uncertain.

A balanced perspective

Our findings suggest the following balanced observations:

- When lognormality holds and extreme values are absent, TOST and $Bayes_T$ perform comparably.
- When lognormality is violated or extreme values are present, $Bayes_T$ offers greater accuracy, efficiency and robustness.
- Disagreements (TOST and $Bayes_T$) are rare but may have meaningful regulatory consequences in individual cases.
- Prespecifying $Bayes_T$ when a BE dataset is expected to be vulnerable to non-lognormality or extreme values can be advantageous.

CONCLUSION

Overall, both simulations and analysis of real-world BE datasets revealed that TOST and $Bayes_T$ perform comparably on lognormal T/R datasets without extreme values, while $Bayes_T$ is more robust, requiring fewer subjects when T/R datasets are normal or contain extreme values.

Importantly, we do not advocate replacing TOST universally. Rather, we suggest that pre-specified Bayesian procedures such as $Bayes_T$ may be considered as scientifically grounded alternatives in situations where TOST assumptions are not satisfied. The evolving regulatory framework described in the 2026 FDA draft guidance provides an opportunity for continued dialogue on how Bayesian methodologies can complement established frequentist standards.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare no competing interests.

DISCLAIMER

This article reflects the views of the authors and should not be construed to represent FDA's views or policies. Readers are encouraged to refer to the discussions of extreme values and outliers in the final FDA guidance for industry, "Statistical Approaches to Establishing Bioequivalence" (May 2026).

REFERENCES

1. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987;15(6):657-680.
2. Kruschke JK. Bayesian estimation supersedes the t test. J Exp Psychol Gen. 2013;142(2):573.
3. Peck C, Campbell G, Yoo I, Feng K, Hu M, Zhao L. Comparing a bayesian approach (best) with the two one-sided t-tests (tosts) for bioequivalence studies. AAPS J. 2022;24(5):97.
4. Wang J, Campbell G, Lee JH, Hu M, Feng K, Chattopadhyay S, et al. Bioequivalence of ANDA Data using a Non-Informative Bayesian Procedure (BEST) compared with the Two One-Sided t-Tests (TOST). AAPS J. 2025;27(2):47.
5. Kraaikamp FD, Meester HL. A modern introduction to probability and statistics. Springer: Berlin/Heidelberg, Germany. 2005.
6. US Food and Drug Administration. Use of bayesian methodology in clinical trials of drug and biological products. 2026.