



Comment on: “Why Were More Than 200 Subjects Required to Demonstrate the Bioequivalence of a New Formulation of Levothyroxine with an Old One?”

Axel Krebs-Brown¹ · Alain Munafo² · Sumedh Gaikwad¹ · Bogumila Urgatz¹ · Claire Castello-Bridoux³

Published online: 5 December 2019
© The Author(s) 2019

Having read with great interest the Current Opinion article by Concordet et al. [1] in this journal, we would like to comment on a number of points where we feel further clarification is required, or where we disagree with the authors' statements about the Levothyrox[®] trial [2].

The authors cite their previous opinion paper [3] and point out the number of subjects outside the bioequivalence acceptance range. We commented at the time [4] that no established bioequivalence range exists for individual exposure ratio (IER) estimates. It is unclear how to interpret the percentage of observed IERs outside this range. Concordet et al. [5] did not refute this in their response; yet in the present publication, they again claim that the number of subjects with IER outside the 0.9–1.11 acceptance range for average bioequivalence (ABE) calls into question the results of the study. We maintain that it is misleading to apply the ABE acceptance range, intended for the confidence interval (CI) around the geometric mean ratio (GMR), to the IER in an ABE trial, and to draw conclusions about switchability from such calculations.

Concordet et al. [1] assert that a large sample size nullifies the effect of shortening the acceptance range. They ignore the fact that a shortened acceptance range of 0.9–1.11

ensures that for the future population, average exposure with the new formulation (NF) will deviate by no more than 10% from the exposure with the old formulation (OF). With the usual acceptance range (0.8–1.25), the difference to be ruled out is 20%. This is independent of the number of subjects included in the ABE trial: the conclusion of “bioequivalence” provides this guarantee, and it is therefore incorrect and deceptive to say (or to imply) that a large sample size provides any less protection for individuals than a small sample size. If the true GMR of two formulations is outside the 0.9–1.11 acceptance range, then the chance of falsely concluding bioequivalence is bounded by the type I error rate (5% for ABE, as required by regulators). The greater the deviation from equivalence, the smaller the chance of such an erroneous conclusion, especially with more precise estimates that come with larger sample sizes.

Concordet et al. [1] illustrate that a smaller sample size in the Levothyrox[®] trial would have led to a lower chance of demonstrating ABE. They use this undeniable and unsurprising fact to cast doubt on the conclusion of the original trial. The opposite is the case: it shows that the trial was planned appropriately, namely with a sufficiently large sample size to meet its objective of conclusively demonstrating presence or absence of bioequivalence. It should be noted that the sample size for any bioequivalence study must be justified statistically, as required in guidance documents by health authorities around the world. It is generally accepted (see, for example, Davit et al. [6], a reference also quoted by Concordet et al. [1]), that at least 80% or 90% power is required. If a trial is “under-powered”, participants are unnecessarily exposed to the study intervention and procedures, and this would be unethical. The Levothyrox[®] study was planned and executed to high scientific and ethical standards, and the sample size justification was no exception to this.

In this context, it is unnecessary and misleading to use a re-sampling approach. Unnecessary because the results can be obtained analytically—see Fig. 1 as an illustration.

This comment refers to the original article available at <https://doi.org/10.1007/s40262-019-00812-x>.

This comment refers to the reply available at <https://doi.org/10.1007/s40262-019-00848-z>.

✉ Claire Castello-Bridoux
claire.castello-bridoux@merckgroup.com

¹ Merck KGaA, Darmstadt, Germany

² Merck Institute for Pharmacometrics (An Affiliate of Merck KGaA, Darmstadt, Germany), Lausanne, Switzerland

³ Direction des Affaires Médicales, Merck Santé s.a.s, 37 rue Saint-Romain, 69008 Lyon, France

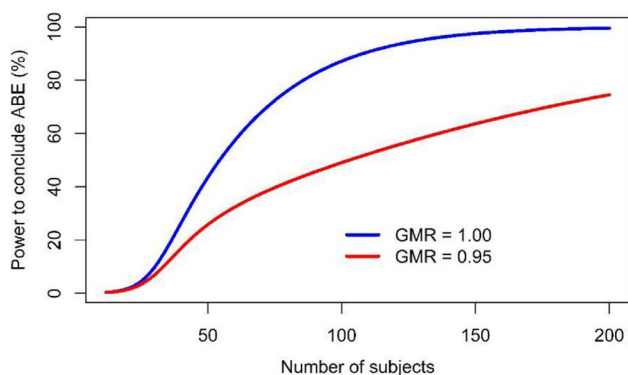


Fig. 1 Power to conclude ABE versus sample size in a 2×2 crossover study. The calculations assume a mixed intra-subject CV of 23.7%, and an ABE acceptance range of 0.9–1.11. The blue line shows the power assuming no difference between the formulations; the red line allows for a 5% difference in exposure. ABE average bioequivalence, CV coefficient of variation, GMR geometric mean ratio

Misleading because in their re-sampling exercise, the authors implicitly fix the true GMR as well as the variability so that they are known quantities. In an a priori sample size calculation, this information is not available, and the uncertainty around these unknown quantities must be accounted for. Specifically, the observed GMR in the Levothyrox[®] study was close to 1; in planning such a study, it is common practice (and explicitly required by the US Food and Drug Administration [FDA] [7]) to allow for a 5% deviation in exposure between the formulations. Further, by definition, a re-sampling exercise only considers study completers, whereas in planning an experiment, one must reflect the expected number of non-completers in the sample size. Especially in a crossover study with a relatively long washout period, not all participants will return for the second administration, and in accordance with guidelines, it is necessary to specify up front how many additional subjects will be recruited to compensate for this. Consequently, the statement that “it is with 150 subjects that the ABE would definitely have been demonstrated” is wrong in this generality. Furthermore, such reasoning is based on post hoc power considerations, a practice that is generally considered inappropriate (see, for example, Hoenig and Heisey [8] and Jirutek [9]).

As in their previous opinion paper, Concordet et al. [1] make a connection between a narrower bioequivalence acceptance range, and individual bioequivalence (IBE). It is not our place to comment on the usefulness of IBE as a concept; we do, however, take issue with the implication that the Levothyrox[®] ABE trial was planned to hide a deficiency of the NF. Both the choice of design and the justification of the sample size met all regulatory requirements and were scientifically and ethically sound.

Without wanting to appear petty, we should point out that the paper contains several inaccuracies and misrepresentations. For instance, the authors confuse median and mean at one point, and they imply that Levothyrox[®] is a “highly variable drug” (HVD; according to the definition set by regulatory authorities [10, 11], a drug is an HVD if its within-subject variability is greater than 30%—this is not the case for Levothyrox[®]). Finally, we would like to distance ourselves from the statement that “ABE formally considers the subjects enrolled in a trial as ‘experimental material’ (they can be regarded as running chromatograph columns)”. This implies that sponsors and planners of ABE trials inherently display a callous attitude to participants in such a trial, an accusation we must strongly protest. Any protocol must be reviewed and approved by an ethical committee and the very choice of a sufficiently large sample size to ensure a sufficiently high power (> 90%) is proof of our high consideration to volunteers as pointed out earlier.

Funding No external funding was used in the preparation of this letter.

Compliance with Ethical Standards

Conflict of Interest Axel Krebs-Brown, Alain Munafo, Sumedh Gaikwad, Bogumila Urgatz, and Claire Castello-Bridoux are all employees of Merck KGaA, Darmstadt, Germany, or of one of its affiliates.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Concordet D, Gandia P, Montastruc JL, Bousquet-Mélou A, Lees P, Ferran AA, et al. Why were more than 200 subjects required to demonstrate the bioequivalence of a new formulation of levothyroxine with an old one? *Clin Pharmacokinet*. 2019. <https://doi.org/10.1007/s40262-019-00812-x> (Epub 2019 Aug 21).
2. Gottwald-Hostalek U, Uhl W, Wolna P, Kahaly GJ. New levothyroxine formulation meeting 95–105% specification over the whole shelf-life: results from two pharmacokinetic trials. *Curr Med Res Opin*. 2017;33:169–74.
3. Concordet D, Gandia P, Montastruc JL, Bousquet-Mélou A, Lees P, Ferran A, et al. Levothyrox[®] new and old formulations: are they switchable for millions of patients? *Clin Pharmacokinet*. 2019;58(7):827–33. <https://doi.org/10.1007/s40262-019-00747-3>.
4. Munafo A, Krebs-Brown A, Gaikwad S, Urgatz B, Castello-Bridoux C. Comment on “Levothyrox[®] new and old formulations: are they switchable for millions of patients?” [letter]. *Clin Pharmacokinet*. 2019;58(7):969–71. <https://doi.org/10.1007/s40262-019-00785-x>.

5. Concordet D, Gandia P, Montastruc J-L, Bousquet-Mélou A, Lees P, Ferran AA, et al. Authors' reply to Castello-Bridoux et al.: "Comment on levothyrox[®] new and old formulations: are they switchable for millions of patients?" [letter]. *Clin Pharmacokinet*. 2019;58(7):973–5. <https://doi.org/10.1007/s40262-019-00786-w>.
6. Davit BM, Nwakama PE, Buehler GJ, Conner DP, Haidar SH, Patel DT, et al. Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration. *Ann Pharmacother*. 2009;43:1583–97.
7. US FDA. Guidance for Industry: Statistical approaches to establishing bioequivalence. Final guidance. 2001. <https://www.fda.gov/media/70958/download>. Accessed 7 Sep 2019.
8. Hoening JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am Stat*. 2001;55(1):19–24.
9. Michael R. Jiroutek: why it is nonsensical to use retrospective power analyses to conduct a postmortem on your study. *J Clin Hypertens*. 2018;20:408–10.
10. US FDA. Guidance for Industry: Bioavailability and bioequivalence studies submitted in NDAs or INDs—general considerations. Draft guidance. 2014. <https://www.fda.gov/media/88254/download>. Accessed 30 Aug 2019.
11. Committee for Medicinal Products for Human Use (CHMP). Guideline on the investigation of bioequivalence. European Medicines Agency. 2010. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf. Accessed 30 Aug 2019.