



# Why Were More Than 200 Subjects Required to Demonstrate the Bioequivalence of a New Formulation of Levothyroxine with an Old One?

Didier Concordet<sup>1</sup> · Peggy Gandia<sup>1</sup> · Jean-Louis Montastruc<sup>2</sup> · Alain Bousquet-Mélou<sup>1</sup> · Peter Lees<sup>3</sup> · Aude A. Ferran<sup>1</sup> · Pierre-Louis Toutain<sup>1,3</sup> 

© The Author(s) 2019

## Abstract

At the request of French Regulatory Authorities, a new formulation of Levothyrox<sup>®</sup> was licensed in France in 2017, with the objective of avoiding the stability deficiencies of an existing licensed formulation. Before launching the new formulation, an average bioequivalence (ABE) trial was conducted, having enrolled 204 subjects and selected for interpretation a narrow a priori bioequivalence range of 0.90–1.11. Bioequivalence was concluded. In a previous publication, we questioned the ability of an ABE trial to guarantee the switchability within patients of the new and old levothyroxine formulations. It was suggested that the two formulations should be compared using the conceptual framework of individual bioequivalence. The present paper is a response to those claiming that, despite the fact that ABE analysis does not formally address the switchability of the two formulations, future patients will nevertheless be fully protected. The basis for this claim is that the ABE study was established in a large trial and analyzed using a stringent a priori acceptance interval of equivalence. These claims are questionable, because the use of a very large number of subjects nullifies the implicit precautionary intention of the European guideline when, for a Narrow Therapeutic Index drug, it recommends shortening the a priori acceptance interval from 0.80–1.25 to 0.90–1.11.

## Key Points

The use of a large number of subjects for an average bioequivalence (ABE) trial for Narrow Therapeutic Index drugs should be questioned.

Regarding the size of the within-subject variability, use of a large number of subjects for an ABE trial nullifies a precautionary intention implicit in the European Union bioequivalence guideline when it recommends shortening the a priori acceptance interval.

For Narrow Therapeutic Index drugs, if an ABE data analysis trial is planned, it is proposed that, as a minimal requirement, a fully replicated design be required to compare test and reference products using the within-subject variability.

## 1 Introduction

In France, almost 3 million patients are currently receiving levothyroxine, a Narrow Therapeutic Index (NTI) drug and, prior to 2017, most were prescribed Levothyrox<sup>®</sup>, a tablet formulation marketed by Merck Serono. In March 2017, at the request of French Authorities, a new formulation (NF) of Levothyrox<sup>®</sup> (Levothyrox<sup>®</sup>NF) was licensed in France to replace the old formulation (OF), for which stability deficiencies had been demonstrated. The single active drug (known as synthetic L-thyroxine, levothyroxine, or L-T4) was the same in both formulations; only the excipients were changed, with the replacement of lactose by mannitol and citric acid. Levothyrox<sup>®</sup>NF is marketed in France, Switzerland, and Turkey; it was launched in Germany in April 2019 and it is anticipated that the NF will receive marketing approval in 21 European Union (EU) member states [1].

✉ Pierre-Louis Toutain  
pltoutain@wanadoo.fr

<sup>1</sup> INTHERES, Université de Toulouse, INRA, ENVT, Ecole Nationale Vétérinaire, 23 Chemin des capelles, 31076 Toulouse, France

<sup>2</sup> Service de Pharmacologie Médicale et Clinique, Faculté de Médecine, Centre Hospitalier Universitaire de Toulouse, Université de Toulouse, Toulouse, France

<sup>3</sup> The Royal Veterinary College, University of London, London, UK

Whilst both formulations have been shown to be bioequivalent based on EU-recommended bioequivalence guidelines [2], adverse drug reactions were reported in several thousand patients using the NF [3, 4]. The number of reports, for a simple formulation substitution of the product with, quantitatively, the same active ingredient, was unprecedented. In June 2019, the French Agency released a pharmaco-epidemiological survey comparing 1,037,553 patients treated in 2016 with the OF versus 1,037,553 subjects treated in 2017 with the NF and it was concluded that approximately 20% of patients had ceased using the NF at the end of 2017 compared with 3% for the paired group treated with the OF in 2016 [5]. These figures are comparable to the rate of switchback, defined as the switching from a branded drug to a generic and then back to the branded drug reported for some antiepileptic drugs with an NTI. Examples are valproic acid and some other antiepileptic drugs versus antihyperlipidemics and antidepressants [6].

In attempting to document the altered health status of the NF in some patients, we published a report [7] in which we questioned the ability of a classical bioequivalence trial, based on EU guidelines, to ensure the switchability between the two formulations of levothyroxine. Because introduction of the NF resulted in a major public crisis, the raw data of the bioequivalent dossier were placed in the public domain. From these data, we computed that almost 70% of the 204 healthy volunteers enrolled into this successful (from a regulatory perspective) average bioequivalence (ABE) trial were outside the a priori bioequivalence range selected of 0.90–1.11. We concluded that this very high number of subjects placed a question mark over switchability of the formulations. Indeed, in accordance with the original articles on this topic by Anderson and Hauck in 1990 [8] and 1994 [9], a clear distinction should be made between *prescribability* (the possibility of using the reference product or the generic at *initiation* of the treatment) and *switchability* (the possibility of switching one formulation with another in a given patient *already under treatment*). The current rule for ABE guarantees prescribability but not switchability. This is because only an individual bioequivalence (IBE) assessment formally aims to compare the exposure obtained with each formulation *within each individual subject*, thereby ensuring that each individual will be similarly exposed to the two formulations.

## 2 Why the Use of a Large Number of Subjects for an Average Bioequivalence (ABE) Trial should be Questioned

Following our earlier publication [7], some opinion leaders have claimed that the ABE for Levothyrox<sup>®</sup> was robustly established because (1) it was demonstrated with a very large number of subjects ( $n = 204$ ); and (2) the classical

bioequivalence acceptance interval was shortened from 0.80–1.25 to 0.90–1.11. Implicitly, the message conveyed was that any future patients will be appropriately protected, because bioequivalence was established in a large trial and the analysis was based on a stringent a priori bioequivalent acceptance interval.

In this article, we explain why these two lines of argument based on ABE are flawed, and we argue for the polar opposite opinion. Indeed, in an ABE analysis, a small a priori bioequivalence acceptance interval provides no supplementary protection to a future patient against the risk of an individual bio-unequivalence when, additionally, the demonstration of this required ABE had succeeded using an atypically very large number of subjects. It is essential to appreciate just what an ABE can and cannot guarantee and what, moreover, is the precise meaning and regulatory expectation when recommending a more *stringent* 0.90–1.11 bioequivalence interval.

By ABE definition, two products are deemed bioequivalent if the 90% confidence intervals of geometric mean test/reference exposure ( $\mu T/\mu R$ ) ratios for maximum concentration ( $C_{max}$ ) and area under the concentration–time curve (AUC) fall within the a priori bioequivalence limits of 80–125% (usual case) or 90–110% (exceptional case), following the EU guideline, which states: “In specific cases of products with a narrow therapeutic index, the acceptance interval for AUC should be tightened to 90.00–111.11%” [10]. This simply means that (1) an ABE can only guarantee that the  $\mu T/\mu R$  ratio of the median bioavailability is located, with a 5% statistical protection, within a regulatory pre-defined bioequivalence interval; and (2) that reducing this a priori bioequivalence from 0.8–1.25 to 0.90–1.11 merely requires, for a given number of patients, a smaller residual reflecting variability of both formulations that is implicitly desirable for a NTI drug. It should be understood that an ABE formally considers the subjects enrolled in a trial as ‘experimental material’ (they can be regarded as running chromatograph columns) which a priori shares exactly the same ratio of geometric mean  $\mu T/\mu R$  between the two formulations. Therefore, any individual departure from this value may be classified as ‘experimental noise’ and not as variability having any biological relevance. The consequence is that requiring a narrower interval to accept a bioequivalence range simply guarantees a low level of experimental noise and, in turn, increases confidence in the conclusion that the two formulations are, on average, bioequivalent.

To summarize, an ABE trial does not provide any guarantee on the individual status of each subject enrolled in the trial, regardless of the width of the a priori bioequivalence acceptance interval. All ABE ensures is protection against the risk of a large departure from unity of the  $\mu T/\mu R$  ratio in *in vivo* conditions.

A widely held view, which may be described as a ‘lax’ interpretation of what ABE actually establishes, amongst both healthcare professionals and regulatory authorities is that switchability is established between two formulations as a ‘by-product’ of an ABE trial. For example, in a retrospective analysis conducted by US Food and Drug administration (FDA) scientists, comparing generic and innovator product bioequivalence data from 2070 clinical bioequivalence studies, it was stated “The statistical approach used by the FDA to analyze BE study data is designed to minimize the risk in situations where the patient is *switched* to a generic version of a medication that he or she is currently taking” and “The robust performance of bioequivalence testing in generic drug approvals over many years lends strong support to the FDA’s *belief* that health professionals can *substitute* drug products determined to be *therapeutically equivalent* with the full expectation that the generic product will produce the same clinical effect and safety profile as the innovator product” (emphasis added) [10]. For the reasons outlined in the following sections, we do not concur with this opinion, for those drugs having a NTI, even when the a priori bioequivalence interval is tightened to 0.9–1.11

### 3 A Large Number of Subjects for an ABE Trial Nullifies the Precautionary Intention of the European Union (EU) Bioequivalence Guideline when it Recommends Shortening the A Priori Acceptance Interval from 0.80–1.25 to 0.90–1.11

Most ABE studies are typically conducted with 24–36 subjects [11] enrolled into a  $2 \times 2$  crossover design, whereas for the Levothyrox<sup>®</sup> study the number of subjects was 204. As for clinical trials, bioequivalence trial protocols must be submitted to and approved by an ethical committee, with a justification for the planned number of subjects (actually 216 for this trial) [2]. From the selection of this very large sample size, it seems likely that the company knew or surmised that a large number of subjects would be required to demonstrate bioequivalence based on ABE. Axiomatically, it follows that a much smaller number of subjects (e.g., typically 24–36) would have led to rejection of a bioequivalence conclusion on the basis of ABE. We explored this hypothesis using bootstrapping. After sampling 10,000 subsamples from the available raw data (i.e., 204 subjects taking the two formulations—as we had no access to the actual crossover design), we estimated the likelihood to conclude bioequivalence with a classical sample size ( $n = 24$  subjects) to be only 10.2%. With bootstrap samples of 48 and 98 subjects, the likelihood to

conclude bioequivalence increased to 42.3% and 87.8%, and it is with 150 subjects that the ABE would definitely be demonstrated (99.8%). When such a large number of subjects is required to demonstrate bioequivalence, it is appropriate to, first, identify the factors determining the large number and, second, reflect on the consequences of interpreting the results of the ABE.

It should be recognized that, when the average  $\mu T/\mu R$  ratio is equal or close to 1:1, as presently is the case for Levothyrox<sup>®</sup> (estimated average ratio of 0.993:1) [2], *it is always possible to demonstrate an ABE* merely by increasing the number of subjects. This is so, even when the NF has a very poor reproducibility in its performance, i.e., a large within- (intra-) subject variability (WSV). The WSV is generally expressed as the coefficient of variation (CV%) of the analysis of variance (ANOVA) residual and it is a matter of major importance from the patient perspective, especially when treated with an NTI drug, because it reflects the day-to-day variability of exposure to the formulation. Drugs for which this residual term is greater than 30% are classified as highly variable drugs [12]. According to the FDA, “it is believed that highly variable drugs generally have a wide therapeutic window; in other words, despite high variability, these products have been demonstrated to be both safe and effective” [11]. Axiomatically, it can be concluded that a product formulation of a highly variable drug is highly undesirable for any drug with an NTI and this is supported by FDA scientists, for whom one of the characteristics of an NTI drug is that it has a low-to-moderate WSV [13]. We note that, in an overview on the WSV of NTI drugs, it was reported by the FDA that the mean WSV for levothyroxine was only 9.3%, with a range of 3.8–15.5%, for the AUC in nine bioequivalence trials [13]. For the old and new Levothyrox<sup>®</sup> products, the very large number of planned subjects to demonstrate an ABE was likely attributable to an anticipated high WSV, rather than to deviation of  $\mu T/\mu R$  from unity, as confirmed by the actually measured value of 23.7% [2]. Hence, to propose tightening the a priori bioequivalence confidence intervals, in order to protect future patients, is unsound, as there is a trade-off between the number of subjects required to demonstrate an ABE and the width of the selected bioequivalence interval. This is because the width of the confidence interval is directly proportional to the intra-subject CV% and inversely proportional to the square root of the number of subjects. In other words, increasing the number of subjects in an ABE contradicts the implicit spirit of the international regulation, when it is used to tighten the bioequivalence interval for NTI drugs with a  $\mu T/\mu R$  ratio equal or close to 1:1.

## 4 For Narrow Therapeutic Index Drugs, it is Important to Place in Perspective the Number of Subjects in an ABE Trial and to Conduct a Bioequivalence Trial Measuring Intra-Subject Variability

In order to resolve these issues for the old and new Levothyrox<sup>®</sup> formulations, a new replicated bioequivalence trial, based on ad hoc analysis could be conducted. A first option would be to tentatively analyze the trial according to IBE rather than ABE concepts, as recently undertaken for gabapentin [14]. Like levothyroxine, gabapentin is a critical drug in terms of bioequivalence. Alternatively, as proposed by FDA, the same replicated trial could be analyzed following a reference-scaled ABE approach, as suggested for NTI drugs, i.e., scaling the bioequivalence limits of a test product to the WSV of the reference product to compare the mean as well as the WSV. This has been explored and discussed recently by others [15] and is proposed in draft FDA guidance specifically devoted to levothyroxine. Once finalized, this guidance will represent the FDA's thinking on this topic [16].

## 5 Conclusion

In the future, it would be wise to explore the risk of declaring a highly variable drug with an NTI as bioequivalent simply by enrolling a large total number of subjects in an ABE trial, in order to mitigate an unduly large WSV. More broadly, there is a need to establish scientifically sound bioequivalence standards for NTI drugs in Europe. Finally, it should be reiterated that, even if the ABE approach and the two-way crossover constitute the general rule in the EU, in particular circumstances a company should not be disallowed from exploring alternative replicated designs and data analysis, such as population bioequivalence to document prescribability and IBE to support switchability between formulations.

**Author Contributions** PG retrieved and validated the raw data, AF validated the raw data, DC performed the statistical analysis, PLT and DC drafted the article. All co-authors critically reviewed several drafts of the manuscript.

## Compliance with Ethical Standards

**Funding** No sources of funding were used in the preparation of this manuscript.

**Conflict of interest** Didier Concordet, Peggy Gandia, Jean-Louis Montastruc, Alain Bousquet-Mélou, Peter Lees, Aude A. Ferran, and Pierre-Louis Toutain declare that they have no competing interests.

**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. Merck. Merck receives recommendation for approval in 21 EU countries for the new formulation of Euthyrox<sup>®</sup>. <https://www.merckgroup.com/en/news/euthyrox-approval-eu-26-07-2018.html>. Accessed 16 Aug 2018.
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. “Enquête de pharmacovigilance Levothyrox<sup>®</sup>” levothyroxine. Laboratoire Merck. Enquête officielle (3eme presentation). Rapport du 04/07/2018. ANSES. [https://ansm.sante.fr/var/ansm\\_site/storage/original/application/73c1a2f8f97a5fd351f4196ed1d68b1a.pdf](https://ansm.sante.fr/var/ansm_site/storage/original/application/73c1a2f8f97a5fd351f4196ed1d68b1a.pdf). Accessed 16 Aug 2018.
4. Mouly S, Roustit M, Bagheri H, Perault-Pochat M-C, Molimard M, Bordet R. The French Levothyrox<sup>®</sup> crisis: we did the best we could but... *Therapie.* 2019;74:431–5.
5. Levothyrox et médicaments à base de lévothyroxine: rapport final de l'étude de pharmaco-épidémiologie à partir des données du Système National des Données de Santé (SNDS) - Point d'Information - ANSM : Agence nationale de sécurité du médicament et des produits de santé. 2019. <https://www.ansm.sante.fr/S-informer/Actualite/Levothyrox-et-medicaments-a-base-de-levothyroxine-Rapport-final-de-l-etude-de-pharmaco-epidemiologie-a-partir-des-donnees-du-Systeme-National-des-Donnees-de-Sante-SNDS-Point-d-Information>. Accessed 24 Jun 2019.
6. Andermann F, Duh MS, Gosselin A, Paradis PE. Compulsory generic switching of antiepileptic drugs: high switchback rates to branded compounds compared with other drug classes. *Epilepsia.* 2007;48:464–9.
7. Concordet D, Gandia P, Montastruc JL, Bousquet-Mélou A, Lees P, Ferran A, et al. Levothyrox<sup>®</sup> new and old formulations: are they switchable for millions of patients? *Clin Pharmacokinet.* 2019;58:827–33.
8. Anderson S, Hauck WW. Consideration of individual bioequivalence. *J Pharmacokinet Biopharm.* 1990;18:259–73.
9. Hauck WW, Anderson S. Measuring switchability and prescribability: when is average bioequivalence sufficient? *J Pharmacokinet Biopharm.* 1994;22:551–64.
10. 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](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf). Accessed 10 Aug 2019.
11. 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.
12. Endrenyi L, Tothfalusi L. Bioequivalence for highly variable drugs: regulatory agreements, disagreements, and harmonization. *J Pharmacokinet Pharmacodyn.* 2019;46(2):117–26.

13. Yu L, Jiang W, Zhang X, Lionberger R, Makhoul F, Schuirmann D, et al. Novel bioequivalence approach for narrow therapeutic index drugs. *Clin Pharmacol Ther.* 2015;97:286–91.
14. Van Lancker G, Van Bortel L, Delafontaine B, Boussey K, Swart E, Chahbouni A, et al. Switchability of gabapentin formulations: a randomized trial to assess bioequivalence between neurontin and gabasandoz on the individual subject level. *Clin Pharmacol Ther.* 2019;106(1):195–203.
15. Jayachandran P, Okochi H, Frassetto LA, Park W, Fang L, Zhao L, et al. Evaluating within-subject variability for narrow therapeutic index drugs. *Clin Pharmacol Ther.* 2019;105:411–6.
16. US FDA. Draft guidance on levothyroxine sodium. Food & Drug Administration; 2018. p. 2. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM428208.pdf>. Accessed 4 Oct 2019.