

## Pourquoi a-t-il fallu plus de 200 sujets pour démontrer la bioéquivalence de la nouvelle et de l'ancienne formulation de lévothyroxine ?

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>

<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

### Résumé

À la demande des autorités réglementaires françaises, une nouvelle formulation de Lévothyrox® a été mise sur le marché en France en 2017, avec comme objectif la correction de défauts de stabilité de la formulation qui était jusqu'alors sur le marché. Avant de lancer la nouvelle formulation, un essai de bioéquivalence dite moyenne (BEM) a été réalisé, incluant 204 sujets et avec comme objectif la démonstration d'une bioéquivalence moyenne qui devait être réalisée avec un intervalle cible d'équivalence resserré entre 0,90 et 1,11. L'essai a conclu à la bioéquivalence. Dans une publication précédente, nous avons contesté la capacité d'un essai de BEM à garantir la substituabilité de l'ancienne par la nouvelle formulation pour un patient donné. Il a été suggéré que les deux formulations devraient être comparées en utilisant le cadre conceptuel de la bioéquivalence dite individuelle. Le présent texte est une réponse à ceux qui affirment que, bien que l'analyse de la bioéquivalence moyenne n'étudie pas directement le passage d'une formulation à l'autre, les futurs patients seront parfaitement protégés. L'affirmation repose sur le fait que l'étude de la bioéquivalence moyenne a inclus un grand nombre de sujets et qu'elle a été analysée en utilisant un intervalle d'équivalence cible plus strict. Ces arguments sont discutables parce que l'utilisation d'un grand nombre de sujets invalide l'intention de précaution implicite de la ligne directrice européenne lorsqu'elle recommande de réduire l'intervalle d'acceptation cible de 0,80 à 1,25 à 0,90 à 1,11 pour un médicament à marge thérapeutique étroite.

### Points clés

**L'inclusion d'un grand nombre de sujets dans un essai de bioéquivalence dite moyenne (BEM) sur des médicaments à marge thérapeutique étroite doit être discutée.**

**En ce qui concerne l'importance de la taille de la variabilité intra-sujet, l'inclusion d'un grand nombre de sujets dans un essai de bioéquivalence moyenne invalide l'intention de précaution qui est implicite dans la ligne directrice de l'Union européenne lorsqu'elle recommande de rétrécir l'intervalle d'acceptation cible.**

**Pour les médicaments à marge thérapeutique étroite la planification d'un essai avec une analyse des données en bioéquivalence moyenne, devrait exiger au minimum de faire un essai dans lequel chaque sujet reçoit deux fois chacune des deux formulations pour pouvoir comparer les deux formulations en prenant en compte la variabilité propre à chaque sujet.**

## 1. Introduction

En France, près de 3 millions de patients reçoivent actuellement de la lévothyroxine, un médicament à marge thérapeutique étroite (MTE) et, avant 2017, la plupart d'entre eux étaient traités par le Levothyrox® en comprimés, une formulation commercialisée par Merck Serono. En mars 2017, à la demande des autorités françaises, une nouvelle formulation (NF) de Levothyrox® (Levothyrox®NF) a été homologuée en France pour remplacer l'ancienne formulation (AF) qui présentait un défaut de stabilité. Le principe actif (connu sous le nom de L-thyroxine, de lévothyroxine ou de L-T4 synthétique) était le même dans les deux formulations; seuls les excipients étaient changés, le lactose étant remplacé par du mannitol et de l'acide citrique. Le Levothyrox®NF est commercialisé en France, en Suisse et en Turquie; il a été mis sur le marché en Allemagne en avril 2019 et on s'attend à ce qu'il soit autorisé dans 21 des États membres de l'Union européenne (UE) [1].

Bien que les deux formulations aient été démontrées comme étant bioéquivalentes sur la base des lignes directrices de l'Union Européenne [2], des effets indésirables ont été rapportés chez des milliers de patients utilisant la NF [3, 4]. Le nombre de notifications a été sans précédent pour un simple changement de formulations gardant le même principe actif en même quantité. En juin 2019, l'agence française a publié en ligne une enquête pharmaco-épidémiologique comparant 1 037 553 patients traités en 2016 avec l'AF à 1 037 553 patients traités en 2017 avec la NF et concluait qu'environ 20% des patients avaient cessé d'utiliser la NF à la fin de 2017 contre 3% pour le groupe traité avec l'AF en 2016 [5]. Ces chiffres sont comparables aux taux de retour (défini comme le passage d'un médicament de marque à un générique suivi par le retour au médicament de marque) rapportés pour certains antiépileptiques à MTE par exemple l'acide valproïque et certains autres antiépileptiques dans une comparaison avec des antihyperlipidémiques et des antidépresseurs [6].

Avec l'intention de comprendre l'altération des états de santé induits par la NF chez certains patients, nous avons publié un rapport [7] dans lequel nous mettions en doute la capacité d'un essai de bioéquivalence classique conduit selon la ligne directrice de l'UE, à assurer le passage d'une formulation de lévothyroxine à l'autre. Parce que l'introduction de la NF a donné lieu à une crise de santé publique majeure, les données brutes du dossier de bioéquivalence ont été mises en ligne. À partir de ces données, nous avons calculé que près de 70% des 204 volontaires sains inclus dans cet essai de BEM - considéré comme un succès du point de vue réglementaire-étaient en dehors de la zone de bioéquivalence cible comprise entre 0.90–1.11. Nous avons conclu que ce très grand nombre de sujets permettait de douter de la sécurité du changement de formulation. En effet, conformément aux articles originaux sur ce sujet par Anderson et Hauck en 1990 [8] et 1994 [9], une distinction claire doit être faite entre le caractère prescriptible d'un nouveau médicament c'est-à-dire la possibilité d'utiliser sans problème un médicament princeps ou un générique lors d'un nouveau traitement et la substituabilité qui est la possibilité de passer d'une formulation à une autre chez un patient déjà sous traitement. La ligne directrice actuelle garantit la bioéquivalence moyenne mais pas la possibilité de passage d'une formulation à l'autre. Seule l'évaluation de la bioéquivalence dite individuelle permet de comparer formellement les résultats obtenus avec chaque formulation, sujet par sujet, garantissant ainsi que chaque individu sera exposé de façon semblable avec les deux formulations.

## 2. Pourquoi faut-il remettre en question l'utilisation d'un grand nombre de sujets dans un essai de bioéquivalence moyenne ?

Suite à notre précédente publication [7], certains leaders d'opinion ont prétendu que la bioéquivalence moyenne des deux Levothyrox® était solidement établie parce que (1) elle avait été établie dans une étude ayant inclus un grand nombre de sujets ( $n = 204$ ); et (2) le classique intervalle cible d'acceptation de la bioéquivalence avait été rétréci de 0,80-1,25 à 0,90-1,11. Le message implicite était que les futurs patients seraient protégés de manière appropriée parce que la bioéquivalence avait été établie lors d'un grand essai et parce que son analyse était basée sur un intervalle cible d'acceptation de la bioéquivalence qui était étroit.

Dans cet article, nous expliquons pourquoi ces deux arguments fondés sur la BEM sont erronés et défendons une position diamétralement opposée. En effet, dans une analyse de la BEM, retenir un intervalle cible d'acceptation de la bioéquivalence étroit n'apporte aucune garantie supplémentaire à un futur patient contre le risque de non bioéquivalence individuelle quand, simultanément, la démonstration de cette BEM est obtenue en utilisant un grand nombre de sujets. Il faut comprendre ce que la BEM démontre exactement et ce qu'elle ne peut pas démontrer, et comprendre en plus quelle est l'intention réglementaire quand elle recommande un intervalle cible plus stricte : 0.90–1.11 pour la bioéquivalence.

Selon la définition de la BEM, un produit test est réputé bioéquivalent à un produit de référence si les intervalles de confiance à 90% du rapport des moyennes géométriques test/référence ( $\mu T / \mu R$ ) de la concentration ( $C_{max}$ ) et de l'aire sous la courbe de concentration en fonction du temps (AUC) se situent dans les limites cibles de la bioéquivalence de 80–125% (cas habituel) ou de 90–110% (cas exceptionnel). Ceci est conforme à la ligne directrice de l'UE, qui stipule: «Dans le cas spécifique des produits à index thérapeutique étroit, l'intervalle cible pour l'AUC doit être réduit à 90,00–111,11% » [10]. Cela signifie simplement - (1) qu'une BEM garantit seulement que le rapport des médianes des biodisponibilités est situé avec un risque statistique d'erreur de 5%, dans un intervalle cible de bioéquivalence prédéfini réglementairement; et (2) que la réduction de cet intervalle de bioéquivalence cible de 0,8–1,25 à 0,90–1,11 exige simplement, pour un nombre donné de patients, une variabilité résiduelle plus faible des deux formulations ce qui est implicitement souhaitable pour un médicament à marge thérapeutique étroite. Il faut comprendre que dans un essai de BEM, les sujets inclus sont considérés comme du "matériel expérimental" (ils peuvent être considérés comme des colonnes de chromatographie ambulantes) ayant *a priori* exactement le même rapport des moyennes géométriques avec les deux formulations. Par conséquent, tout écart individuel autour de cette valeur peut être qualifié de «bruit expérimental» et non comme une variabilité ayant une pertinence biologique. La conséquence est que l'exigence d'un intervalle plus étroit comme cible de la bioéquivalence garantit simplement un faible niveau de bruit expérimental et donc augmente la confiance en la conclusion que les deux formulations sont, en moyenne, bioéquivalentes.

En résumé, un essai de BEM ne fournit aucune garantie sur l'état de chacun des sujets inclus dans l'essai, quelle que soit la largeur de l'intervalle cible d'acceptation de la bioéquivalence. Tout ce que la BEM garantit c'est la protection contre le risque d'un écart important par rapport à l'unité du rapport  $\mu T / \mu R$  dans des conditions in vivo.

De nombreux professionnels de santé et certaines autorités de régulation pensent, dans une interprétation plus permissive que la substituabilité est une conséquence potentielle d'une BEM et qu'un essai qui établit la BEM, établit en conséquence la possibilité de passage d'une formulation à l'autre. Par exemple, des scientifiques de la « Food and Drug Administration » aux Etats-Unis, ont fait une analyse rétrospective de 2 070 études de bioéquivalence clinique comparant un médicament générique au médicament princeps, et écrivent : «L'approche statistique utilisée par La FDA pour analyser les données d'études de bioéquivalence est conçue de façon à réduire les risques d' un patient qui passe d'un traitement princeps au générique correspondant» et «La robustesse des performance des tests de bioéquivalence dans l'approbation de médicaments génériques depuis plusieurs années apporte un argument solide à la position de la FDA selon laquelle les professionnels de la santé peuvent substituer les médicaments jugés équivalents sur le plan thérapeutique en toute confiance et que le générique produira le même effet clinique et le même profil d'innocuité que le médicament princeps» (nous soulignons) [10]. Pour les raisons exposées dans les sections suivantes, nous ne partageons pas cet avis pour les médicaments ayant une MTE, même lorsque l'intervalle de bioéquivalence cible est réduit à 0.9–1.11

**3. Inclure un grand nombre de sujets dans un essai de BEM invalide l'intention de précaution de la ligne directrice de l'Union européenne sur la bioéquivalence lorsqu'elle recommande de rétrécir l'intervalle d'acceptation cible de 0,80-1,25 à 0.90–1.11**

La plupart des essais de BEM utilisent un plan expérimental dans lequel chaque sujet est son propre témoin et incluent généralement entre 24 et 36 sujets [11] alors que l'essai de Levothyrox®, a inclus 204 sujets. Comme pour les essais cliniques, les protocoles d'essais de bioéquivalence doivent être soumis et approuvés par un comité d'éthique, avec une justification pour le nombre de sujets prévu (216 pour l'essai de Lévothyrox) [2]. Pour avoir planifié un si grand échantillon, il semble probable que l'entreprise savait ou supposait qu'un grand nombre de sujets seraient nécessaires pour démontrer la bioéquivalence basée sur la BEM. A contrario on peut en déduire qu'un nombre beaucoup plus réduit de sujets (par exemple, un nombre plus habituel entre 24 et 36) auraient conduit au rejet de la bioéquivalence sur la base de la BEM. Nous avons exploré cette hypothèse en utilisant une méthode bootstrap. Après échantillonnage de 10 000 sous-échantillons obtenus à partir des données brutes disponibles (c'est-à-dire 204 sujets prenant les deux formulations sans information sur l'ordre des traitements), nous avons estimé que la probabilité de conclure à la bioéquivalence n'était que de 10,2% avec un échantillon de 24 sujets. Avec des échantillons de 48 et 98 sujets, la probabilité de conclure à la bioéquivalence atteignait respectivement 42,3% et 87,8%, et c'est avec 150 sujets que la BEM était établie de façon quasi certaine (99,8%). Quand un si grand nombre de sujets est nécessaire pour démontrer la bioéquivalence, il faut 1) identifier les facteurs conduisant à ce grand nombre et, 2) réfléchir à ses conséquences dans l'interprétation des résultats de la BEM.

Il faut bien comprendre que si le rapport moyen  $\mu_T / \mu_R$  est égal ou proche de 1 (comme pour les deux Levothyrox® pour lesquels ce rapport est égal à 0,993) [2], on peut toujours démontrer la BEM simplement en augmentant le nombre de sujets. C'est vrai, même quand la nouvelle formulation a une très faible reproductibilité dans ses performances, c'est-à-dire, une grande variabilité intra sujet. Cette variabilité est généralement exprimée par le coefficient de variation de la variance résiduelle dans

l'analyse de variance et c'est une donnée d'une importance capitale pour les patients, surtout quand le traitement est à MTE, car elle reflète la variabilité au jour le jour de l'exposition à la formulation. Des médicaments pour lesquelles ce terme résiduel est de plus de 30% sont considérés comme hautement variables [12]. Selon la FDA, "on pense que les médicaments hautement variables ont généralement une large fenêtre thérapeutique; autrement dit, malgré une grande variabilité, ces produits ont été démontrés à la fois sûrs et efficaces » [11]. A contrario, on peut conclure qu'une formulation hautement variable d'un médicament à MTE est à éviter et cette position est défendue par les experts de la FDA, pour qui l'une des caractéristiques d'un médicament à MTE est qu'il a une variabilité intra-sujet faible ou modérée [13]. A noter, dans une revue de la variabilité intra sujet des médicaments à MTE la FDA a indiqué que la variabilité intra sujet de la lévothyroxine pour l'AUC n'était que de 9,3%, avec une étendue de 3,8% à 15,5%, d'après neuf essais de bioéquivalence [13]. Pour les deux Lévothyrox®, le très grand nombre de sujets prévus pour démontrer la BEM était probablement justifié par l'attente d'une variabilité intra-sujet élevée, plutôt qu'à l'attente d'une déviation du rapport  $\mu_T / \mu_R$  de la valeur nominale de 1, et comme cela a été confirmé dans l'essai avec un coefficient de variation estimé de 23,7% [2]. Il n'est donc pas valable de simplement proposer de rétrécir les intervalles de cible de la bioéquivalence pour protéger les futurs patients à cause d'un effet de balançoire entre le nombre de sujets nécessaire pour démontrer la BEM et la largeur de l'intervalle de bioéquivalence. En effet la largeur de l'intervalle de confiance (NDT : qui doit être inclus dans l'intervalle cible d'équivalence) est proportionnelle à la variabilité intra-sujet et inversement proportionnelle à la racine carrée du nombre de sujets. En d'autres termes, l'augmentation du nombre de sujets dans une étude de BEM contredit l'esprit implicite de la réglementation internationale, quand cette augmentation est utilisée pour rétrécir l'intervalle de bioéquivalence pour les médicaments à MTE avec un rapport  $\mu_T / \mu_R$  égal ou proche de 1.

**4. Pour les médicaments à marge thérapeutique étroite, il est important de mettre en perspective le nombre de sujets dans un essai de BEM et de faire un essai de bioéquivalence en mesurant la variabilité intra-sujet.**

Afin de résoudre ces problèmes pour l'ancienne et la nouvelle formulation de Lévothyrox®, on pourrait faire un nouvel essai de bioéquivalence en faisant deux séries de mesures par formulation pour chaque sujet, et faire une analyse ad hoc. Une première option serait d'essayer d'analyser la bioéquivalence individuelle au lieu d'analyser la bioéquivalence en moyenne, comme cela a été récemment fait pour la gabapentine [14]. Comme la lévothyroxine, la gabapentine est un médicament pour lequel la bioéquivalence est une donnée critique. Dans une autre option, proposée par la FDA, le même essai pourrait être analysé selon une BEM dite normalisée par rapport à la formulation de référence. Cette approche préconisée pour les médicaments à MTE, consiste à définir les limites cibles de l'intervalle de bioéquivalence en fonction de la variabilité intra-sujet du produit de référence afin de comparer en même temps moyenne et variabilité intra-sujet. Cela a été exploré et discuté récemment par d'autres [15] et est présenté dans un projet de ligne directrice de la FDA spécialement consacré à la lévothyroxine. Une fois finalisée, cette ligne directrice représentera la position de la FDA sur ce sujet [16].

## 5. Conclusion

À l'avenir, il serait sage d'explorer le risque de déclarer bioéquivalent un médicament très variable avec une MTE simplement en incluant un grand nombre de sujets dans un essai de BEM, afin d'atténuer une variabilité intra-sujet trop grande. De manière plus générale, il est nécessaire d'établir des normes de bioéquivalence scientifiquement valables pour les médicaments à MTE en Europe. Enfin, il convient de rappeler que, même si l'approche en BEM et l'étude de chaque sujet comme son propre témoin sans réplication constitue la règle générale dans l'UE, dans des circonstances particulières, un industriel ne devrait pas s'interdire d'explorer des types d'essais alternatifs notamment avec réplication et des analyses alternatives des données, comme les études de la bioéquivalence dite de population pour documenter le caractère prescriptible d'un nouveau médicament et de la bioéquivalence dite individuelle pour documenter le passage d'un produit à un autre chez des patients en cours de traitement.

Contributions des auteurs PG a récupéré et validé les données brutes, AF validé les données brutes, DC a effectué l'analyse statistique, PLT et DC ont rédigé l'article. Tous les coauteurs ont révisés plusieurs versions du manuscrit.

### **Conformité aux règles éthiques**

Financement Aucune source de financement n'a été utilisée dans la préparation de ce manuscrit.

Conflit d'intérêts Didier Concordet, Peggy Gandia, Jean-Louis Montastruc, Alain Bousquet-Mélou, Peter Lees, Aude A. Ferran et Pierre-Louis Toutain déclare ne pas avoir de conflits d'intérêts.

### **Références**

1. Merck. Merck reçoit un accord de mise sur le marché dans 21 pays de l'UE pour la nouvelle formulation d'Euthyrox®. <https://www.merckgroup.com/en/news/euthyrox-approval-eu-26-07-2018.html>. Consulté le 16 août 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®" levothyroxine. Laboratoire Merck. Enquête officielle (3eme présentation). Rapport du 04/07/2018. ANSM. [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 ® 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/Points-d-information-Points-d-information/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 ® 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-bioequivalencerev1\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalencerev1_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, Makhlof 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.accessdata.fda.gov/drugsatfda\\_docs/psg/Levothyroxine\\_Sodium%20capsules\\_NDA%20021924\\_RC%20Oct%202018.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/psg/Levothyroxine_Sodium%20capsules_NDA%20021924_RC%20Oct%202018.pdf) (NDT : lien mis à jour pour cette traduction)