Bioequivalence Studies for Levothyroxine

Submitted: February 7, 2005; Accepted: February 7, 2005; Published: March 30, 2005.

Sanford Bolton[1]

[1] University of Arizona, 5495 North Via Velazquez, Tucson, AZ 85750

ABSTRACT

The Food and Drug Administration (FDA) Guidance for Bioavailability and Bioequivalence Studies for Levothyroxine has been challenged by companies that manufacture brand-name products. Their contention is that the current guidance does not adequately address the endogenous background levels of the drug, and that the ratios of the PK parameters, a basis for approval of equivalence, are not assessed correctly. In particular, they conclude that products that have a potency differing by 12.5% cannot be differentiated using the present guideline and criteria for acceptance of bioequivalence. They claim that such a difference can be a public health hazard because of the perception among practitioners that levothyroxine is a narrow therapeutic index drug. This article describes the procedure recommended in the current Guidance for Levothyroxine and demonstrates that the methods recommended are adequate and will accept products that are therapeutically equivalent. To date, no generic product accepted as equivalent using FDA Guidances has been shown to result in a safety and efficacy profile different from its brand counterpart.

INTRODUCTION

The current bioequivalence standards to establish bioequivalence of generic drug products typically use the accepted 2 treatment-2 period-2 sequence crossover design and compare blood levels of drug for standard and reference products over time after a single dose to volunteer subjects. From these data, the maximum observed blood concentration (C_{max}) and the area under the blood level vs time curve (AUC) are calculated for each product using a logarithmic transformation. If the blood level curves are considered superimposable based on inspection of the curves and the calculated average C_{max} and AUC parameters, then the test product is deemed bioequivalent to the reference product. Because of variability owing to subject differences and possible small differences between the products, the FDA has recommended that 90% confidence intervals be placed on the ratio of test to reference for AUC and C_{max}, and that this interval be within 80% to 125% (based on the antilog of the log ratio) to obtain approval of bioequivalence. Although there has been some heated discussion concerning the adequacy of this interval, most knowledgeable people have accepted these limits as adequate. Most generic drugs show ratios close to 100% with confidence intervals varying depending on the variability of the data and the sample size used for the bioequivalence study. To this date, according to my understanding, there has been no documented public health hazard resulting from the implementation of this criterion. After all, most drugs are prescribed with little consideration for patient body size and other characteristics that may affect bioavailability and therapeutic effect. This practice has not been cause for alarm. Therefore, at the present time, in general, the scientific community is comfortable with the statistical criteria for approving studies for the majority of drugs on the market.

Corresponding Author: Sanford Bolton, University of Arizona, 5495 North Via Velazquez, Tucson, AZ 85750. Tel: (520) 529-9445; Fax: (520) 529-9445; E-mail: bolton@acninc.net

Some drugs that have special characteristics have been the subject of controversy regarding the Food and Drug Administration (FDA) recommendations for assessing bioequivalence. In particular, manufacturers of products containing drugs that may be considered as Narrow Therapeutic Index (NTI), and those that are natural products and have endogenous blood levels, have challenged the current guidelines. These manufacturers have asserted that NTI drugs need special considerations, such as specific study designs, and a narrower confidence interval requirement, (eg, 95% to 105%).

Several years ago, a concerted effort to impose such limits on warfarin generic products, clearly an NTI drug, was not successful. To this date, there have been no documented problems with NTI generic products, which have AUC and C_{max} ratios very close to 100%.

Endogenous blood levels that interfere with the assessment of drug blood levels resulting from exogenous dosing can usually be dealt with by subtracting baseline levels from the total levels measured, or an equivalent procedure. In these cases, the baseline levels are usually estimated by one or more pre-dose blood analyses. Also, in usual circumstances, the background level of drug is small compared with the levels due to dosing, so that the interference is minimal, and estimates of bioequivalence are not compromised.

The current FDA guidance for assessing the bioavailability of levothyroxine sodium (hereafter referred to as levothyroxine) products[1] uses the typical 2-period crossover design and uses a larger dose (600 μg) than is typically used therapeutically. These approaches are meant to overcome the potential problems of background endogenous interference and low levothyroxine levels from lower doses, which may be difficult to analyze with reasonable precision.

It is important to note that the FDA also considers the dissolution profile and the formulation as part of its global evaluation for approval of abbreviated new drug applications (ANDAs) for generic substitution. For levothyroxine products, the formulation is simple; there are no complicated ingredients or slow-release mechanisms. The dissolution is relatively rapid and uncomplicated. For a simple formulation, as is the case for levothyroxine products, if the exact same amount of active ingredient is in each dosage unit, the tablet can be expected to deliver the same amount of drug. All of this helps to ensure that products that meet the current FDA guidelines will perform the same as a reference-listed drug.

Studies performed for ANDA submissions of levothyroxine products as recommended by the FDA have shown moderate intrasubject variability. The results of studies also have shown dose proportionality for the 50, 100, and 300 μg tablets. These studies were performed using multiple tablets to obtain a total dose of 600 μg. The rapid dissolution of the tablets confirms that the formulations are uncomplicated (almost 100% dissolved in less than 30 minutes). The confidence intervals comfortably passed the FDA criterion. For example, in one such submission, although not an official requirement, between 80% and 100% of the individual ratios for C_{max} and AUC were between 75% and 125% (the old 75–75 rule), confirming the consistency and lack of variability for levothyroxine products (unpublished data, ANDA submission by Mova Laboratories to FDA [confidential]).

Nevertheless, the determination of bioequivalence of thyroid products has been controversial. The controversy is based on several issues:

1. The interference of the accurate analysis of plasma concentration of levothyroxine, following oral ingestion of tablets, owing to endogenous levels (which are themselves variable)
2. The designation of levothyroxine as an NTI drug
3. The relevance of levothyroxine compared with thyroid-stimulating hormone (TSH) as the bioequivalence "indicator"
4. The use of a high dose in bioequivalence studies
5. The potential for confounding carryover effects

These issues were raised, in the form of Citizens' Petitions to the FDA, by companies that market branded thyroid products.[2,3] These companies contend that the presently recommended protocol for determining bioequivalence for levothyroxine generic products is deficient. In particular, they claim that products that are not equivalent may be deemed to be equivalent using the design and bioequivalence criteria presently recommended by FDA.

This article presents a discussion of the adequacy of the current Bioequivalence Guidance for levothyroxine products. Again, the bases of arguments to the contrary, as defined in the Citizens' Petitions,[2,3] are related to the suggestion that levothyroxine is an NTI drug and that endogenous levels of levothyroxine interfere with, and bias, the assessment of bioequivalence. This is particularly relevant for levothyroxine, where endogenous levels are a significant portion of the total blood levels measured after dosing and may bias the assessment of bioequivalence. Also, arguments have been presented suggesting that levothyroxine may not be the appropriate marker for bioequivalence. Rather, a recommendation has been proposed for the possible use of TSH blood levels, a measure typically used in medical practice to monitor titration dosage.

Baseline Adjustment and Dosage

The evidence used as a basis for the claims that the present FDA guidance is inadequate largely depends on 2 items: (1) a simulation showing that baseline readings are needed to adjust blood levels to more accurately estimate relative potency (unpublished data, Globomax, Technical Report, Globomax Project #KNP00500, Submitted to Abbott Laboratories, April 25, 2002), and (2) a clinical bioequivalence study,[4] both sponsored by Abbott Laboratories.

The dose comparison paper[4] describes the results of a 3-period crossover design in 33 subjects, comparing pharmacokinetic (PK) parameters for levothyroxine doses of 400, 450, and 600 μg. The data were analyzed without a baseline correction and with 3 different baseline corrections (see Table 1). The purpose of the study was 2-fold: (1) to examine the effect of baseline corrections, and (2) to evaluate the sensitivity of the study to detect differences between products with known differences in potency.

There is little argument that a baseline correction is appropriate. The FDA recommends subtracting the average of 3 predose levothyroxine levels from each blood sampling value. When the baseline endogenous level is a significant part of the total drug measured, bioequivalence evaluation could be compromised. Consider the following situation. Suppose that the background level is a constant 8 μg/mL and that C_{max} for Product A is 16 μg/dL and for Product B, 18 μg/dL. Without a correction, the ratio A/B is 16/18 = 89%. Subtracting the baseline, the ratio is 8/10 = 80%. Thus, the relatively high endogenous level of drug results in a masking of the relative difference between the drug products. This result was also demonstrated in the simulation study submitted as part of Abbott's Citizens' Petition.[3] Because of the relatively high proportion of drug in blood attributable to endogenous levels, this is an important consideration. Clearly, increasing the dose and correcting for baseline appropriately should be a positive move toward the accurate estimation of relative potency. If the baseline is small relative to the total blood level, its effect on evaluating bioequivalence will be mini-

Table 1. Effect of Baseline Corrections on Estimation of Relative Potency of Different Doses of Levothyroxine Products*

Correction Method†	Dose Comparison	C_{max} Ratio (Confidence Ratio)	AUC (96-hour) Ratio (Confidence Interval)
1	450 vs 600 µg	0.783 (0.73–0.84)	0.680 (0.60–0.77)
	400 vs 600 µg	0.803 (0.75–0.87)	0.660 (0.58–0.75)
	450 vs 600 µg	0.975 (0.91–1.05)	1.031 (0.91–1.16)
2	450 vs 600 µg	0.793 (0.74–0.85)	0.816 (0.74–0.90)
	400 vs 600 µg	0.807 (0.75–0.87)	0.750 (0.68–0.82)
	450 vs 600 µg	0.982 (0.92–1.05)	1.088 (0.99–1.19)
3	450 vs 600 µg	0.820 (0.76–0.89)	0.693 (0.63–0.76)
	400 vs 600 µg	0.775 (0.72–0.84)	0.639 (0.58–0.70)
	450 vs 600 µg	1.058 (0.98–1.15)	1.084 (0.99–1.19)

*Data from Blakesley et al.[3]

†The corrections are designated as (1) the average of 3 baseline readings, (2) baseline levels diminishing according to a 7-day half-life, and (3) correction of each sampling point based on data from the day prior to dosing.

mal. The closer the baseline is to the total blood level of drug, the more serious is the effect on estimation of the true ratio of the PK parameters.

Another suggestion is to make a correction at each blood-sampling time point, based on a possible consistent diurnal pattern—correction method 3 in Blakesley et al.[4] Blakesley et al argue that the point-by-point correction is more effective because it takes into account diurnal variation, as opposed to the 3-point average prior to dosing. If the diurnal variation is real and consistent, it would certainly improve the analysis. However, ordinary variation in the diurnal levels, in addition to the fact that the correction would be based only on the results of a previous day's analysis, could increase the variability and potential bias. In fact, the results described in the Blakesley et al article[4] reveals very little difference in the conclusions based on the 2 approaches. The C_{max} ratio appears to be estimated more accurately with the point-by-point approach for the comparison of C_{max} for the 400 vs 450 µg doses (see Table 1).

It is my opinion that it is not obvious that the use of point-by-point baseline corrections would be of much advantage over the predose baseline correction, if any. It is certainly possible that such a correction could add more "noise" to the analysis, if the correction was inappropriate. Taking the average of 3 predose baseline values appears to be as reasonable and appropriate an approach as more complex methods of correction.

In a briefing document,[5] Abbott described the advantages and disadvantages of baseline corrections. They also agree that the baseline is variable.

Dosing in Bioequivalence Determinations

The use of a dose that is larger than that used clinically has been criticized as a deficiency in the current guidelines. However, as long as the larger dose does not result in a health hazard for the volunteers, a better alternative is not apparent. Blakesley et al[4] claim that products with clinically significant differing potency are not differentiated using the present Guidance. This claim is largely based on the comparison of the 400-µg vs 450-µg dose. However, the conclusion that levothyroxine products differing in dose by 12.5% could not be differentiated seems disingenuous.

There are problems with the comparison of the 2 lower doses, 400 and 450 µg, other than that such comparisons are not performed in bioequivalence studies for ANDA submissions. These doses are substantially lower than that recommended by the FDA for such studies, and low doses, with the relatively large endogenous levels, will yield less reliable blood levels than a bioequivalence comparison using the FDA recommended 600-µg dose. This is simply a matter of increasing the signal (concentration due to dosing) to noise (endogenous concentration) ratio. It would seem that if any comparisons were to be made in this study, the use of a larger dose such as 500 µg or larger to compare with the 600-µg dose would have been more meaningful. In fact, in a study using the baseline adjustment, method 1,[6] submitted in support of a levothyroxine product, a comparison of 500-µg and 600-µg doses had an observed ratio of almost exactly 1.2 (600/500) for both AUC and C_{max}. In the Blakesley et al[3] article, using their correction methods 1 and 2, the 450-µg dose actually showed a lower C_{max} than that for the 400-µg dose, and the confidence interval for C_{max} did not cover the true ratio (1.125). This result could be due to chance, or the correction based on only predose baseline values biased the result.

The fact that the comparison of C_{max} for the lower doses resulted in an anomalous outcome (higher C_{max} for the lower dose) reinforces the contention that the relatively low levels of levothyroxine due to the smaller dose (400 and 450 µg) results in a less reliable estimate of the true ratio of the products.

According to Dr Steven Johnson of the FDA, doses as low as 400 to 450 μg "yield concentrations that are closer to baseline [which] prevent an accurate evaluation of the true differences between the 400 and 450 μg doses."[7] Dr Johnson also noted that the data presented by Abbott showed a biased difference between the 400- and 450-μg doses, rather than the true 12.5% difference.

If one could exactly differentiate endogenous and exogenous levothyroxine levels, the blood level ratio should be unbiased. However, the baseline subtraction is variable and can be considered somewhat unreliable (ie, it is an estimate based on an assumption that endogenous levels are constant or that they vary in a predictable way throughout a 24-hour period). When we get close to the background T4 baseline levels, the variability of the blood level data increases dramatically. For example, the coefficient of variation based only on the assay will approximately double if we reduce the dose to 300 μg from 600 μg. Abbott noted that the variance for the 600-μg dose is less than that for the lower doses (0.0356 and 0.0336 for C_{max} and AUC, respectively, for 600 μg; 0.0563 and 0.0799 for 450 μg; and 0.0459 and 0.0574 for 400 μg).[8] The increased relative variability of baseline corrected blood level data and the relatively larger interference of endogenous levels with smaller doses are arguments for using the largest possible safe dose in levothyroxine bioequivalence studies.

There can be no dispute that using the largest dose possible, accounting for potential side effects, will give the most reliable estimate of relative potency.

Any disagreement that the larger dose does not reflect the behavior of smaller doses is specious, in my opinion. If the blood levels are "superimposable" for 2 products, then it can be concluded that the kinetics of drug absorption are also identical. Any contradiction to this conclusion would need both theoretical and clinical confirmation. Although the FDA gave no official written rationalization for the thinking behind the 600-μg dose recommendation, this was considered to be the highest single dose that would be considered safe for volunteers.[7]

A study using doses close to 600 μg to validate the usefulness of a baseline correction to more accurately compare levothyroxine products in bioequivalence studies would have been more informative.

Carryover

Another confounding factor in the Blakesley et al[4] study is the existence of a differential carryover effect; the larger dose resulting in higher baseline blood levels in subsequent periods. This effect would tend to bias the comparisons in a 2-period, 2-treatment design. In typical bioequivalence studies, the doses are identical, and the blood level curves reasonably similar. In these cases, we would anticipate inconsequential, or no differential carryover. This emphasizes the fact that the exaggerated difference in doses in the Blakesley et al study[4] does not simulate real conditions when testing bioequivalence, and the conclusions of their study may not be applicable to real life situations. The FDA has found no evidence of carryover in any of the studies submitted to them as new drug approvals (NDAs) or ANDAs.[9]

Thyroid Stimulating Hormone

Another proposed recommendation is to use TSH as a marker instead of levothyroxine. This recommendation is based on the fact that TSH is used clinically to adjust levothyroxine dosage.

However, TSH is more variable and is a secondary response to levothyroxine. The possible use of TSH as a marker has been dismissed by FDA personnel because of these problems.[7] An article by Carr et al[10] was presented as evidence for the possible use of TSH as the moiety in a bioequivalence study at the March 13, 2003, meeting of the Pharmaceutical Sciences Advisory Committee. This study was negatively critiqued by agency personnel, as well as by a member of the Advisory Committee. Dr Steve Johnson of the FDA staff cogently articulated the basis for the current standards and why the measurement of levothyroxine, rather than TSH, is more appropriate (unpublished data, Abbott request for FDA meeting, May 8, 2002). In particular, Dr Lesko of the FDA commented critically on the Carr et al study. Apparently, this was a case-control study that is "probably the lowest evidence of clinical studies…" not blinded or controlled.[10] In fact, this study showed that it was difficult to optimize dosage, and that variations in dosage did not result in different clinical response. Dr Lesko also noted that, according to the *British Medical Journal,* July 2001,[7] the TSH test "…is an unreliable test of thyroid function that has no proven scientific biochemical basis....Free levothyroxine is more reliable…."

As made clear in 21 CFR 320.24(b),[11] for determining bioavailability and bioequivalence, the primary variable in order of accuracy, sensitivity, and reproducibility is the concentration of active ingredient in plasma (levothyroxine). The acute pharmacological effect of the active moiety (TSH) is third on the list, following urinary excretion of the active moiety. Thus, TSH is more variable and may bias the ratio estimate. TSH may also show more variable background levels, and accounting for the baseline correction introduces more problems. Clearly, this is not something one would want to use to establish equivalence when there are no such problems with the measurement of levothyroxine.

The variability of baseline measurements of levothyroxine and TSH based on baseline levels and prescreening data in a bioequivalence study (unpublished data, ANDA submission by Mova Laboratories to FDA [confidential]) is ~25% to 26% for TSH and 10% to 11% for levothyroxine.

Consider the consequences of using TSH as a marker and narrowing the confidence interval to 90% to 110% as suggested in the Citizens' Petition noted above.[2] If the Coefficient of Variation (CV) for TSH is 26% and the confidence interval narrowed from ±20% to ±10%, and if products were perfectly identical, a bioequivalence study with 80% power would require more than 100 subjects. If the products were only 2% different, for example, more than 130 subjects would be needed to demonstrate equivalence.[12] A significant increase in variability alone would make the use of TSH as a marker less desirable than levothyroxine for establishing bioequivalence.

The suggestion by Abbott in their Citizen's Petition that special populations be used in a clinical trial also seems farfetched, especially when bioequivalence studies yield reliable data.[2] For example, they recommend using athyrotic subjects for these studies, a contentious recommendation that has no valid scientific rationale in a bioequivalence setting (unpublished data, Abbott request for FDA meeting, May 8, 2002).

Levothroxine as a Narrow Therapeutic Index Drug

The question of levothyroxine being an NTI drug is controversial. In practice, levothyroxine is perceived as an NTI drug[4]; "…levothyroxine has a narrow therapeutic index."[4] Arguments have been put forth supporting this contention, pointing to the multiple tablet potencies that are available to aid in the titration, and how these tablets are very close in dosage. Nevertheless, according to my knowledge, there is no definitive study that has verified the notion of levothyroxine as an NTI drug, as presently defined. Drugs that are unquestionably NTI, such as sodium warfarin and carbamazepine, do not have any special restrictions for acceptance of bioequivalence, and have not been the subject of any safety or efficacy problems with regard to generic substitution.[9] Other official agencies, such as the Health Protection Branch (HPB) in Canada, specifically do not consider levothyroxine as NTI[13] and have no special bioequivalence requirements.

According to the Code of Federal Regulations 21 CFR 320.33, narrow therapeutic ratio (NTR) is defined as follows: "There is less than 2 fold difference in median lethal dose (LD50) and median effective dose (ED50), or there is less than 2 fold difference in minimum toxic concentrations and (LD50) and median effective concentrations in the blood."[11] However, there is no scientific evidence for this assertion with regard to levothyroxine.

In addition, an NTR drug is further defined as a product where "safe and effective use of drug products requires careful titration and patient monitoring."[11]

The petitioners' argument that doses with as little a difference as 12.5% could have serious health consequences has never been documented. In fact, a single dose missed in 1 week would lower blood levels by more than 10%. If such differences were meaningful, very restricted prescribing and labeling would be indicated.

At this time, there are no FDA recommendations that any drug product has a requirement of a confidence interval for AUC and C_{max} narrower than 85% to 125%. Therefore, until it can be demonstrated scientifically that levothyroxine is unique, necessitating a narrower confidence interval, such a requirement should not be imposed.

Discussion and Summary

The object of a bioequivalence study as recommended by the FDA is to ensure that there is a high degree of confidence (90%) that the ratios of key PK parameters (AUC and C_{max}) for comparative products are within the bioequivalence interval of 0.8 to 1.25. This recommendation has not changed for many years and has not been modified for individual drugs. In particular, all drugs, whether low-dose drugs or NTI drugs, have this same criteria for acceptance. As of this date, these criteria have seemed reasonable and have not resulted in any documented public health issues. It would be an arduous task to assign different criteria for different drugs, especially in the absence of any scientifically demonstrated need to do so.

Levothyroxine is no exception. There have been efforts to designate this drug as an NTI drug, and to introduce theoretical arguments why this drug is different and special, needing special methods of design and analysis. In particular, arguments have been presented critical of the current bioequivalence guidance. This is not the first time that such unfounded arguments have been made to introduce special methodology. For the most part, these proposals have been politically and economically motivated and have not had a demonstrable scientific basis.

Levothyroxine may need to be treated differently because it is an endogenous drug with varying blood levels depending on time as well as other circumstances. This is not unique to levothyroxine. However, because of the relatively large endogenous levels, it is a problem that needs to be addressed so that conclusions of bioequivalence are not compromised.

As part of an effort to investigate the adequacy of levothyroxine submissions, the FDA reanalyzed 4 submissions for levothyroxine[10] in which 16 comparisons of doses (12 × 50 μg, 2 × 300 μg, 6 × 100 μg) were made for AUC and C_{max}. These comparisons were expected to show equivalence, as the tablets were shown to release the same amount of drug at the same rate based on solubility and formulation characteristics. The 90% confidence intervals in all of these studies passed the 80% to 125% criterion. Fourteen of 16 failed a 95% to 105% confidence interval, and 8 of 16 failed a 90%

to 110% confidence interval. (Of course, the 90% confidence intervals could have been met for any of these criteria by using a larger sample size.) The average of all of the point estimates in these studies was 100.5%, strongly suggesting that the present Guidance gives unbiased estimates of equivalence. Two of the 16 confidence intervals in the comparisons did not include 100%. From strictly statistical considerations, we would expect that if the true ratio is 100%, that 10% of the intervals would not cover 100%. This again indicates that the procedure is behaving in a reliable manner.

The contention that a difference of ~10% in dosage can cause serious therapeutic failures or toxicity is contradicted by the fact that patients frequently miss their dose or take a dose more than prescribed (patient compliance issues). The predicted incidence of serious problems has not been observed.

When evaluating bioequivalence, in addition to the PK of the drug, it is important that the formulation of the drug be considered, particularly when small changes in bioequivalence could be a significant issue. If a drug were a problem drug (eg, poorly soluble, poorly absorbed) or in a complex formulation, the problems of evaluating formulations would be greatly compounded. Fortunately, levothyroxine has no formulation problems or difficulties, except, perhaps, for instability, which has been mostly overcome in recent years. Thus, we would expect that simple formulations containing the same dose of drug would not differ in their bioavailabilities. This has been the case for levothyroxine bioequivalence studies recently submitted as ANDAs. Generic NTI drugs with simple formulations are presently on the market, without any indication of therapeutic failures or adverse effects.

The use of a high dose of levothyroxine to obtain a valid estimate of bioavailability is necessary at this time and does not compromise the comparison of 2 products. It should be clearly acknowledged that such studies use humans as an "in vivo device" that measures drug PK. We assume that when comparing multiple tablets of each product applied to the human apparatus, an equivalent PK response for equal total dose would result in the conclusion that administration of single tablets would also result in bioequivalence. For levothyroxine formulations, this is certainly a very reasonable assumption. There is little argument in the case of levothyroxine that blood levels resulting from a single small dose would result in very fuzzy data owing to a relatively large background noise from endogenous levels.

The recommendation to use TSH rather than levothyroxine as the bioequivalence marker has too many problems to be taken seriously. It is a secondary measure of levothyroxine blood levels and is much more variable than levothyroxine.

When 2 products are compared with the same dose of the same very soluble drug, and with a simple formulation, expectations are that differences will be minimal. Levothyroxine is very soluble and rapidly dissolved from the comparative formulations. If there is or has been any doubt that these conditions will lead to similar blood level curves, or that products differing by 25% will be differentiated, a study using sufficiently large doses (~600 μg) of formulations having the same proportion of active and inactive ingredients would have been appropriate. The study by Blakesley et al[3] attempted to investigate this question and showed that products differing by 12.5% in potency passed the FDA confidence interval criterion for bioequivalence. What would have been surprising is that such a study, properly powered, would have failed the FDA confidence interval criterion. The ratio estimate in the Blakesley et al study was biased to the low side, showing less than a 12.5% difference. In fact, in 2 of the 3 baseline correction methods that were proposed, the lower dose had a higher average C_{max}. This result may have been due to the low dose used for the comparison (400 vs 450 μg), resulting in levothyroxine levels too close to the baseline. In any event, as noted above, a better design would be to compare 600 μg to 525 μg or 675 μg, for example. The experimentally derived ratios would likely have been closer to their theoretical values. In fact, in a study submitted to the FDA for approval of a levothyroxine product, the observed ratios of AUC and C_{max} of a 600-μg dose compared with a 500-μg dose were very close to the theoretical ratio of 1.2.[6] Again, Blakesley et al's results for the 400- and 450-μg comparison suggested that these doses are too low to yield reliable estimates of drug potency.[4]

In order to challenge the adequacy of the present guideline, much more and better scientific evidence than that provided to date is necessary. This would require, among other things, comparison of doses that are in the range of 600 μg, and demonstration that small differences in dosage have actual clinical consequences. Thyroid products have been used for many years, with patients receiving the same product year after year. It is known that these products differ somewhat from tablet to tablet and batch to batch. Ordinary variation, plus overage and stability problems did not deter the usefulness of these products. Now, many of these problems have been corrected. Allegations of deficiency, without any verifiable scientific evidence of a deficiency in a proven methodology and that are possibly based on ulterior motives, cannot be taken seriously.

In conclusion, there is no hard, scientific evidence to suggest that the current guideline for bioequivalence for levothyroxine products does not perform as it purports. Evidence, so far, demonstrates that properly powered studies, performed and analyzed as recommended by the FDA, should and do establish bioequivalence for products that differ by less than 20% and differentiate and reject as inequivalent those products that differ by more than 20%.

REFERENCES

1. Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for Industry Levothyroxine Sodium Tablets - In Vivo Pharmacokinetic and Bioavailability Studies and In Vitro Dissolution Testing. Rockville, MD: February 2001.

2. Citizen's Petition. Jones Pharma, Inc. March 28, 2003.

3. Citizen's Petition. Abbott Laboratories. August 25, 2003.

4. Blakesley V, Awni W, Locke C, Ludden T, Granneman GR, Braverman LE. Are bioequivalence studies of levothyroxine sodium formulations in euthyroid volunteers reliable? *Thyroid*. 2004;14:191-200.

5. Food and Drug Administration. FDA Advisory Committee for Pharmaceutical Science. *Briefing Document From Abbott Laboratories.* Rockville, MD; May 12-13, 2003. Available at: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3926B1_02_A-Abbott%20.htm. Accessed March 23, 2005.

6. Mylan Laboratories, Inc. *Levothyroxine Sodium Tablets.* Canonsburg, PA: Mylan Laboratories, Inc; 2000:76-187.

7. Food and Drug Administration. FDA Advisory Committee for Pharmaceutical Science. Record of Proceedings. March 13, 2003. Available at: http://www.fda.gov/ohrms/dockets/ac/cder03.html#PharmaceuticalScience. Accessed March 23, 2005.

8. Citizen's Petition. Abbott Laboratories. August 25, 2003.

9. Food and Drug Administration. FDA Response to Abbott Petition, Citizen's Petition #2003P-0126/CP1, Document PDN 1. June 29, 2004.

10. Carr D, McLeod DT, Parry G, Thornes HM. Fine adjustment of thyroxine replacement dosage: comparison of the thyrotrophin releasing hormone test using a sensitive thyrotrophin assay with measurement of free thyroid hormones and clinical assessment. *J Clin Endocrinol*. 1988;28:325-333.

11. Food and Drug Administration. Code of Federal Regulations. Title 21, Volume 5: CFR 320.33. Available at: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=320.24. Accessed: March 23, 2005.

12. Bolton S, Bon C. *Pharmaceutical Statistics, Practical and Clinical Applications.* 4th ed. New York, NY: Marcel Dekker; 2003.

13. Health Canada. Expert Advisory Committee on Bioavailability. Record of Proceedings. April 16, 2003. Available at http://www.hc-sc.gc.ca/hpfb-dgpsa/tpd-dpt/2003-04-16_e.html. Accessed March 23, 2005.