ORIGINAL RESEARCH ARTICLE

Clin Pharmacokinet 2004; 43 (14): 1037-1053
0312-5963/04/0014-1037/$31.00/0

Assessment of Levothyroxine Sodium Bioavailability

Recommendations for an Improved Methodology Based on the Pooled Analysis of Eight Identically Designed Trials with 396 Drug Exposures

Ingeborg Walter-Sack,[1] *Christof Clanget*,[2] *Reinhard Ding*,[1] *Christoph Goeggelmann*,[1] *Vera Hinke*,[2] *Matthias Lang*,[1] *Johannes Pfeilschifter*,[2] *Yorki Tayrouz*[1] and *Karl Wegscheider*[3]

1 Department of Internal Medicine VI, Clinical Pharmacology and Pharmacoepidemiology, University Medical Center, Heidelberg, Germany
2 Department of Internal Medicine I, Endocrinology and Metabolism, University Medical Center, Heidelberg, Germany
3 Institute for Statistics and Econometry, University of Hamburg, Hamburg, Germany

Abstract

Background: Assessment of dosage form performance in delivering endogenous compounds, such as hormones, *in vivo* requires a specific approach.

Objectives: Assessment of relative bioavailability of levothyroxine sodium (L-T4) from eight solid preparations, compared with a liquid formulation, by using pharmacological doses, and critical evaluation of trial methodology based on the pooled analysis of individual data.

Design: Eight open-label, randomised, single-dose, crossover phase I studies using eight solid L-T4 dosage forms (25, 50, 75, 100, 125, 150, 175, 200μg per tablet; administered total doses 600, 625 or 700μg) and a liquid formulation; assessment of relative bioavailability by 90% confidence intervals for the relative area under the concentration-time curve (AUC) of total thyroxine (TT4), i.e. protein-bound plus free thyroxine, calculated by using the recommended log AUC four-way analysis of variance models for crossover designs. For the pooled analysis, general linear models were applied to assess the validity of model assumptions, to identify potential sources of effect modification, to discuss alternative modelling approaches with respect to endogenous hormone secretion and to give recommendations for future designs and sample sizes.

Participants: One hundred and sixty-nine healthy males; 29 of these individuals participating in two studies.

Interventions: Single oral doses of L-T4 tablets and the liquid formulation administered after fasting, separated by at least 6 weeks; a total of 396 drug exposures.

Main outcome measures: TT4 AUC from 0 to 48 hours and peak plasma concentration with and without baseline correction.

Results: Each study demonstrated equivalence of the tablets to the drinking solution, independent of the chosen analysis model. Sequence effects that could devalidate the chosen crossover approach were not found. Period effects with changing directions that could best be explained by seasonal variation were detected. While the pre-specified method of baseline correction of simply subtracting individual time-zero TT4 values was disadvantageous, the analysis of total AUC could be improved considerably by covariate adjustment for baseline TT4. With this approach, sample sizes could have been substantially reduced or, alternatively, the recommended equivalence ranges could be reduced to ±6%.
Conclusion: Using a single pharmacological dose of L-T4 in two-period crossover designs is a safe and reliable procedure to assess L-T4 dosage form performance. With an adequate statistical modelling approach, the design is efficient and allows general conclusions with moderate sample sizes.

Levothyroxine sodium (L-T4) is used for hormone replacement therapy and for suppression of thyrotropin (thyroid-stimulating hormone; TSH) secretion. Maintenance dosages of L-T4 in adults range from 25 to 200 μg/day, or even higher.[1-3] Thus, a broad range of oral L-T4 dosage forms is required. However, L-T4 effectiveness is influenced by formulation variables. It has long been recognised that variability of drug absorption from poorly designed thyroid hormone dosage forms may represent a therapeutic hazard in long-term treatment.[4-11] In addition, regular dose adjustments to individual needs can be attained more easily when the pharmaceutical properties of oral dosage forms are reliably stable and, thus, predictable.

Reliability of product characteristics is the relevant factor to be identified by bioavailability and bioequivalence studies. This includes quantification of not only the mean rate and extent of systemic availability of the active ingredient, but also the variability of these criteria, since large fluctuations in drugs with a narrow therapeutic range may be associated with therapeutic failure and intoxication, respectively.

In particular, analytical problems arise with lower doses of L-T4 as it becomes difficult or impossible to differentiate between endogenous hormone secretion and exogenous L-T4, except in athyroid patients after a withdrawal period of sufficient length. Therefore, it is not feasible to perform conventional bioavailability studies with hormone doses, as used for replacement therapy. Instead, a procedure is needed which reliably allows quantification of the administered drug. Using pharmacological doses on top of the endogenous thyroxine (T4) pool could be an appropriate approach for reliably obtaining increments of total serum thyroxine (TT4, i.e. protein-bound plus free thyroxine) over physiological levels.

A small pilot study in healthy individuals (four males and four females) using a 75μg and 100μg tablet strength,[12] and a bioequivalence study performed in healthy females with a single dose of two tablets of 300μg formulation[13] showed that after a single oral dose of L-T4 600μg it is possible to measure changes of serum concentrations against the background of endogenous T4. As single large doses of L-T4 (up to 3mg in healthy individuals) were reported to be safe and lack clinical toxicity,[14-17] the general use of single doses of 600μg as a tool in bioavailability and bioequivalence studies appears justifiable.

The main objective of our investigations was to assess relative bioavailability of L-T4 in a single brand series of oral L-T4 preparations, covering the whole range of dosage strengths, by using pharmacological (instead of replacement) doses of L-T4 as a standard procedure. The study design was chosen

in accordance with the guidelines[18] and general principles accepted at the time the study was performed, but is also in agreement with the latest version of the European guidelines on bioavailability and bioequivalence,[19] as well as the recent standards of the US FDA.[20,21] However, while the pre-specified statistical analysis for formal assessment of the relative bioavailability in each of the eight trials was well in line with the 'average bioequivalence approach' recommended for L-T4 bioavailability studies with nonreplicate designs,[20-22] the chosen method of baseline correction by simply subtracting the individual TT4 zero-time values (C_0) was, in the meantime, challenged by a national European regulatory agency in the process of obtaining drug approval, as well as by the recommendations given in the L-T4 specific US FDA guideline.[20] In reaction to this situation, we repeated the analyses using TT4 total area under the concentration-time curve from 0 to 48 hours (AUC_{48}) and peak plasma concentration (C_{max}) instead of the increments over baseline. Furthermore, we studied how these analyses change when the log AUC_{48} four-way analysis of variance (ANOVA) is extended by incorporation of log baseline TT4 as a covariate, i.e. covariate adjustment for baseline in the log-model by adding an independent variable to the model instead of pre-analysis baseline subtraction from the dependent variable, as initially intended.

In order to compare these three approaches, to further assess the appropriateness of study design, and to give recommendations with respect to sample sizes and evaluation methods of future bioavailability trials on L-T4, we pooled the individual data of the eight trials with identical design to increase the power of the analyses. For the resulting database of 338 applications in 169 individuals, we studied overall as well as trial-specific sequence effects (related to possible carry-over effects), period effects, treatment effects, seasonal effects and further covariate effects by using general linear model methodology.

Methods

Study Design

The single-centre study series consisted of eight separate investigations carried out consecutively during November 1998 through to August 1999. Each study was performed as an open-label, prospective, actively controlled, randomised trial with a two-way, two-period crossover of tablet and reference solution, separated by a washout phase of 6–8 weeks. For six of the eight tablet strengths, a single dose of L-T4 (L-Thyroxin Henning®)[1] 600μg, administered as the respective number of tablets (table I), was used compared with 600μg (6mL) of the reference solution supplied by the same manufacturer. For the two tablet strengths (125 and 175μg), in which multiplying the dose would not result in 600μg, the next higher total dose was chosen (625 and 700μg, respectively), compared with the corresponding amount of the liquid reference solution. Drug treatments were assigned in randomised order.

L-T4 tablets (batch numbers 18 002 and 18 003) were provided in individual pill boxes (per person and per study phase) by Henning Berlin GmbH, Berlin, Germany. Other ingredients were pregelatinised starch, maize starch, microcrystalline cellulose, anhydrous sodium carbonate, sodium thiosulfate × 5 H_2O, colloidal silicon dioxide and hydrogenated castor oil. L-T4 drinking solution (reference drug, batch number 18 159; additionally containing 1,2-propanediol, anhydrous glycerol, demineralised water and sodium thiosulfate pentahydrate) was also provided by Henning Berlin GmbH.

Each subject received two equivalent single-dose treatments with a washout period of at least 6 weeks. Each time, the study participants were admitted to the clinical research centre for the initial 12–13 hours. At approximately 8:00am, after fasting for at

1 The use of trade names is for product identification purposes only and does not imply endorsement.

Table I. Levothyroxine sodium dose, number of participants per study and population characteristics

Tablet strength (μg)	No. of tablets	Total oral dose (μg)	No. of participants	Age (y) [mean ± SD]	Weight (kg) [mean ± SD]	Height (cm) [mean ± SD]	Thyroid volume (mL) [mean ± SD]	Serum TT4 (pre-study; μg/L) [mean ± SD]	Serum TSH (mU/L) [mean ± SD]	Serum albumin (g/L) [mean ± SD]	Total protein (g/L) [mean ± SD]	Total cholesterol (mg/dL) [mean ± SD]
25	24	600	24	25.4 ± 4.4	78.2 ± 8.9	182.8 ± 5.9	14.85 ± 5.98	64.1 ± 10.2	1.71 ± 0.74	47.3 ± 2.8	73.2 ± 4.3	181.8 ± 28.5
50	12	600	24	27.4 ± 4.9	77.4 ± 8.1	182.3 ± 6.3	12.45 ± 3.61	69.2 ± 11.3	1.62 ± 0.45	46.7 ± 3.2	72.5 ± 3.8	170.7 ± 30.4
75	8	600	26	26.4 ± 3.1	77.2 ± 7.1	182.1 ± 6.4	13.03 ± 4.24	67.3 ± 12.0	1.54 ± 0.61	45.0 ± 2.3	71.2 ± 3.6	166.7 ± 37.3
100	6	600	24	24.9 ± 3.8	76.2 ± 9.4	179.9 ± 8.2	12.61 ± 5.17	65.3 ± 11.1	1.53 ± 0.74	46.9 ± 2.8	73.3 ± 4.5	173.8 ± 30.2
150	4	600	24	28.0 ± 6.6	77.7 ± 8.3	181.8 ± 5.8	13.54 ± 4.70	64.0 ± 16.0	1.73 ± 0.99	45.9 ± 2.3	73.6 ± 4.1	166.8 ± 33.7
200	3	600	24	26.3 ± 4.6	76.2 ± 7.1	181.4 ± 5.2	13.18 ± 5.15	67.1 ± 11.5	1.55 ± 0.58	46.6 ± 2.6	74.4 ± 5.1	169.0 ± 31.1
125	5	625	26	25.6 ± 4.0	76.5 ± 8.0	180.3 ± 7.0	12.89 ± 3.89	69.0 ± 09.6	1.62 ± 0.81	46.8 ± 2.8	74.2 ± 3.6	176.4 ± 27.6
175	4	700	26	24.8 ± 3.9	75.3 ± 9.6	179.5 ± 6.5	13.40 ± 5.42	71.1 ± 11.4	1.43 ± 0.73	46.1 ± 2.3	72.0 ± 4.3	174.0 ± 29.4

TSH = thyrotropin (thyroid-stimulating hormone); **TT4** = total thyroxine (protein-bound plus free fraction).

least 12 hours, the study drugs were administered together with 200mL of mineral water. During the hospitalisation period all subjects were provided with standard meals consisting of regular food. Two hours after dose administration a standardised breakfast was served. Alcoholic beverages were not allowed from 24 hours before, to 36 hours after administration. Methylxanthine-containing beverages were not permitted from 12 hours prior to administration until 24 hours after administration of the study drugs. Safety and tolerability of the study drugs were closely monitored and documented by recording frequency, intensity and severity of adverse events.

Prior to (time zero) and 15, 30, 45, 60 and 90 minutes, and 2, 3, 4, 5, 6, 8, 10, 12, 24 and 48 hours following drug administration, 7.5mL blood samples were collected. After clotting at room temperature for 45 minutes, the samples were centrifuged (at ≤4°C and 3000 rpm), and serum aliquots were stored at either 4°C or −20°C.

A 1.5mL serum aliquot was directly allocated to thyroid hormone measurements. For practical reasons this aliquot was kept at 4°C when measurements were performed on the day of, or the day following blood sampling. In all other cases, the aliquot was stored at −20°C until measurement. The reproducibility of hormone determinations after storage at 4°C was confirmed in stability experiments in which 96 ± 4% (mean ± SD; n = 14) of the first measurement was recovered. After short-term storage at −20°C for 1, 3, 6, 8 and 12 hours, recovery of total thyroxine (TT4) was 97, 101, 101, 99 and 99%, respectively (n = 6 for each time interval).

Measurements were performed by competitive immunoassay using automated direct chemoluminescent technology (Chiron Diagnostics ACS:180 Automated Chemiluminescence Systems; Chiron Diagnostics GmbH, Fernwald, Germany), as previously described.[23-25] The assay is able to discriminate between compounds of closely related structures (according to the manufacturer, the percentage of cross-reactivity is as follows: monoiodotyrosine

<0.03%, diiodothyrosine <0.03%, 3,5-diiodo-L-thyronine <0.03%, D-triiodothyronine 0.40%, L-triiodothyronine 1.2%, D-thyroxine 57.7%, reverse triiodothyronine 2.7%). To monitor system performance, three levels of quality control samples (Chiron Diagnostics GmbH, Fernwald, Germany) and identical aliquots of a serum sample from a euthyroid individual were assayed on each day. The same lots were used throughout the study phase. Assay calibrations were performed twice weekly. All TT4 determinations fell within the linear range (5–300 μg/L).

For determination of precision, three samples were assayed six times in each of 48 assays. The following results were obtained: at mean TT4 concentrations of 40.67/90.00/160.62 μg/L the within-run coefficient of variation (CV%) and total CV were 4.0/2.8/2.8 and 5.5/3.9/3.8, respectively. For assessment of interassay variability, Ligand A (low), B (medium) and C (high), and identical aliquots of a serum sample from a euthyroid individual were assayed in different runs during the study phase. The following results were obtained (mean ± SD μg/L, CV%): Ligand A 35.5 ± 2.6, 7.4; Ligand B 83.5 ± 3.0, 3.5; Ligand C 164.3 ± 7.9, 4.8; euthyroid individual 68.0 ± 2.78, 4.1. Accuracy was established during the study phase across the range of the assay by the use of quality control material. The obtained results were 89%, 100% and 97% of the expected values.

Study Population

One hundred and seventy-nine participants were admitted to the studies; ten withdrew owing to adverse events. One hundred and sixty-nine individuals completed their course and 29 of these individuals participated in two trials. The number of individuals who completed the study was 24 in each of the 25/50/100/150/200μg trials (participants who withdrew were replaced in these studies) and 26 in the trials with the 75/125/175μg tablets in order to ensure that late withdrawals would not expand the studies. Voluntarily signed informed consent was obtained from each study participant after full explanation of the study, verbally and in writing.

Eligible individuals had no clinically relevant findings in the medical history and in any of the investigations of the pre-study examination, including a thorough physical examination, safety laboratory screening (including total cholesterol and triglycerides, TT4, total triiodothyronine (TT3) and TSH in serum [table I]), 12-lead ECG and ultrasound examination of the thyroid gland. Exclusion criteria were any acute or chronic illness or allergy, any physical disorder that could interfere with the safety of the participant or with the study objectives, including thyroid disorders, administration of any other investigational drug during 2 months preceding study onset and any other regular drug treatment within 4 weeks prior to study onset. In addition, smoking, excessive drinking of alcoholic beverages (>60 g/day of alcohol), drug addiction or a positive result in drug screening, excessive coffee drinking (>6 cups/day) and excessive physical activities during the days prior to the study were further reasons for exclusion.

Ethical and Legal Standards

The study was carried out in accordance with the Declaration of Helsinki adopted by the World Medical Association in June 1964, Version of Somerset West, October 1996, the specific legal requirements in Germany and the Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95). Written approval by the Ethics Committee of the Medical Faculty of the University of Heidelberg was obtained prior to the beginning of the first study.

Assessment of Bioavailability in Each of the Eight Trials

As pre-specified in the study protocols, statistical analyses of each of the eight trials were performed in accordance with the respective version of the European Guideline "Investigation of Bioequivalence and Bioavailability", in particular Appendix III, "Technical Aspects of Bioequivalence Statistics".[18] For

the primary objective (demonstration of AUC-equivalence between the tablet and the reference solution) the 90% confidence intervals (CI) were calculated using the mean square error of the log AUC_{48} four-way ANOVA with the factors treatment (in this case tablets vs drinking solution), period, sequence and subject (random) nested within sequence. In order to assess the AUC of exogenous L-T4, AUC was calculated each time from the respective increments of TT4 (protein-bound plus free L-T4) over baseline levels (i.e. individual time-zero TT4 levels prior to drug administration, subsequently called baseline TT4) for 48 hours using the trapezoid method (baseline-corrected AUC [AUC-C]). Equivalence was accepted if the 90% CI of the ratio of the test drug AUC-C and the AUC-C of the reference drug were included in the equivalence range of 80–125%.[18] Baseline-corrected peak concentrations (C_{max}-C) were similarly studied using an equivalence range of 70–143% (as predefined in the study protocol) and 80–125% (as these narrower bands were suggested recently). In addition, we repeated these analyses by using TT4 total AUC_{48} and C_{max} instead of the increments over baseline. Furthermore, we studied how these analyses change when the log AUC_{48} four-way ANOVA is extended by incorporation of log baseline TT4 as a covariate, i.e. covariate adjustment for baseline in the log-model instead of pre-analysis baseline subtraction. Only significant effects are reported. In this study, significance is defined as $p < 0.05$ (two-sided).

Pooled Analysis: Model Comparisons

Since each single trial was too small to allow general conclusions on the validity and advantages of the three different modelling approaches, a pooled analysis of the eight trials was performed by pooling the individual data according to the recommendations of Stewart and Clarke.[26] To guarantee stochastic independence, only the first participation of the individuals (who were included in two trials) was admitted to the pooled analysis, resulting in a sample size of 169 subjects with 338 applications. At first, four general linear models (for log AUC-C, log AUC, log C_{max}-C and log C_{max}, respectively) were fitted that included each of the four factors used in the experiment-wise analyses, as well as the factors 'total dose' and 'dosage strength' (simultaneously representing trial identity one-to-one) nested within total dose. Furthermore, interaction terms combining tablet strength (corresponding one-to-one to trial identity) with sequence, period and treatment were added to the model to allow for differences between trials. If interaction terms did not contribute significantly to the model they were eliminated in a backward selection procedure, while the four main factors were always kept in the model, whether significant or not. In a second step, the resulting final models for log AUC and log C_{max} were extended by the covariate log baseline TT4.

For each of the resulting models, point estimates and 90% CI of significant factors or covariate effects are reported, along with an estimate of the residual standard deviation. The factor 'subject' is not listed, since it was always significant, as is to be expected by this kind of study. Since outcomes are on log-scales and are sufficiently small (≤0.15), factor effects and residual standard deviations multiplied by 100 can be approximately read as percent changes and coefficients of variation, respectively.

Pooled Analysis: Seasonal Variation

If period effects occurred, an attempt was made to explain these effects by introducing the month of the year into the models. As a result of the recruitment scheme, only 8 of the 12 calendar months were covered. The marginal monthly total AUC variation, as calculated from the final baseline-adjusted pooled analysis log-AUC model, is demonstrated by a time series chart.

Pooled Analysis: Influence of Subject Characteristics on Bioavailability

Individual variation in bioavailability may result from physiological variability, as well as from identifiable subject characteristics modifying absorp-

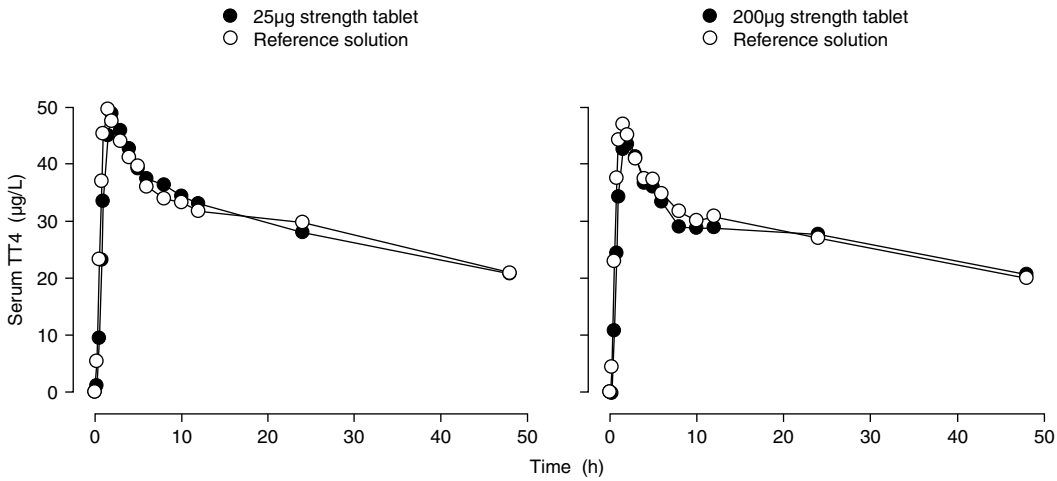

Fig. 1. Average increments of total serum thyroxine concentrations (TT4; protein-bound plus free fraction) over baseline levels following administration of a single dose of levothyroxine sodium 600µg, using the 25µg (**a**) and 200µg (**b**) tablet strength, and the respective dose of the reference solution.

tion. We considered both aspects in the second part of our pooled analysis, which is subject-based (not application-based as the first part) with a sample size of 169 subjects. Firstly, the distribution type of the residual crossover differences (tablet minus solution)[27] of log AUC-C, log AUC, log C_{max}-C and log C_{max} was studied by applying normal plots and Kolmogorov-Smirnov tests. Secondly, we fitted general linear models to the residual crossover differences[27] of the four outcome parameters that included the available subject characteristics, i.e. age, log weight, log height, log thyroid volume, log serum TSH, log serum albumin, total protein and log total cholesterol (see table I). Logarithmic transformations were performed if distributions were skewed. A backward selection procedure was applied and only variables contributing significantly to performance were kept in the model.

Calculations were performed using SPSS-10.0.5 (SPSS Inc., Chicago, IL, USA). The general linear models were fitted using the UNIANOVA module with type-III sums of squares, which is equivalent to the General Linear Models procedure (PROC GLM) in SAS software (SAS Institute, Inc., Cary, NC, USA).

Results

Assessment of Relative Bioavailability in Each of the Trials

The eight groups of participants were comparable with respect to all relevant population characteristics (table I). For each tablet strength, the TT4 concentrations rose within the initial half hour and reached a maximum around 2 hours after study drug administration. The average time courses of serum TT4 increments in the lowest and highest tablet strength groups are shown as examples in figure 1.

The results of the pre-specified analyses using baseline-corrected outcomes are presented in table II; AUC-C and C_{max}-C values of the various tablet strength groups and the respective values after administration of the reference solution are listed in detail. AUC-C and C_{max}-C values were very similar within the 600µg groups. Increased total dose of the study drug for the 125 and 175µg tablet strengths resulted in an increase of the AUC as well as the C_{max} values. The time to reach maximum concentrations (t_{max}) was not influenced by the L-T4 dose. The mean ratios of the AUC-C and C_{max}-C of the test drugs, and the corresponding AUC-C and

Table II. Pharmacokinetic variables (baseline corrected) of total serum thyroxine

Tablet strength (µg)	Total dose (µg)	No. of participants	AUC-C tablets (µg • h/L) [mean ± SD]	AUC-C solution (µg • h/L) [mean ± SD]	Relative AUC-C (%; estimate) [90% CI]	C_{max}-C tablets (µg/L) [mean ± SD]	C_{max}-C solution (µg/L) [mean ± SD]	Relative C_{max}-C (%; estimate) [90% CI]	t_{max} tablets (h) [mean ± SD]	t_{max} solution (h) [mean ± SD]
25	600	24	1389.30 ± 249.60	1416.14 ± 262.50	98.1 [93.9, 102.6]	50.86 ± 8.94	51.72 ± 7.36	97.9 [92.4, 103.7]	2.17 ± 0.65	1.98 ± 1.42
50	600	24	1334.78 ± 286.61	1439.09 ± 207.35	90.9 [82.9, 99.7]	48.16 ± 8.04	50.38 ± 7.76	95.3 [89.0, 102.1]	2.42 ± 0.89	2.00 ± 0.94
75	600	26	1386.60 ± 310.77	1380.17 ± 280.25	100.4 [93.1, 108.1]	48.43 ± 11.30	50.70 ± 11.17	95.3 [87.5, 103.8]	2.55 ± 1.21	2.03 ± 1.29
100	600	24	1473.99 ± 317.34	1532.97 ± 277.59	95.2 [87.9, 103.0]	51.01 ± 10.38	54.53 ± 10.41	93.1 [85.6, 101.4]	2.60 ± 1.12	1.88 ± 1.11
150	600	24	1334.79 ± 319.26	1424.02 ± 251.48	92.5 [85.0, 100.7]	48.59 ± 10.29	50.60 ± 8.73	95.3 [88.9, 102.3]	2.40 ± 0.86	1.98 ± 1.07
200	600	24	1296.71 ± 286.95	1319.98 ± 276.99	97.9 [90.9, 105.4]	45.73 ± 9.67	49.48 ± 8.90	91.9 [86.0, 98.2]	2.21 ± 0.85	1.78 ± 1.14
125	625	26	1541.85 ± 347.43	1519.77 ± 426.25	103.1 [95.7, 111.0]	54.62 ± 10.71	54.83 ± 13.35	100.1 [93.3, 108.8]	2.63 ± 1.26	2.01 ± 0.86
175	700	26	1595.63 ± 336.03	1641.42 ± 418.17	98.4 [90.7, 106.7]	56.53 ± 10.23	57.41 ± 13.88	99.7 [93.1, 106.8]	2.40 ± 1.20	1.85 ± 1.04

AUC-C = baseline-corrected area under the concentration-time curve; **C_{max}-C** = baseline-corrected peak plasma concentration; **t_{max}** = time to reach C_{max}.

C_{max}-C of the reference solution (point estimates) and the corresponding ratios of AUC_{48} and C_{max}, as well as the respective 90% CI, were well included in the pre-specified ranges, independently of the total doses applied. Based on these results, equivalence of the tablet strengths to the corresponding drinking solution, according to the pre-defined criteria, can be accepted for each tablet strength (preliminary results were presented by Goeggelmann et al.[28]). Significant sequence effects with log AUC-C or log C_{max}-C were not observed in any of the trials. Period effects observed with log AUC-C and log C_{max}-C ranged from −0.106 to 0.070 (not significant; p = 0.012) and from −0.045 to 0.040 (not significant).

Table III shows the corresponding results for total AUC and C_{max} values. As can be seen, results were similar. Again, 90% CI were well included within the equivalence ranges, with and without adjustment for log baseline TT4. CI were considerably smaller than those for baseline-corrected outcomes, and even smaller after adjustment for baseline. With total AUC or total C_{max}, significant sequence effects were not observed. Period effects observed with log AUC ranged from −0.053 to 0.031 (p = 0.0001, p = 0.022, respectively) and for log C_{max} from −0.045 to 0.018 (p = 0.029, not significant, respectively). After adjustment for baseline, the estimates of the period effects were reduced to ranges of −0.034 to 0.024 (p = 0.025, p = 0.002, respectively) and from −0.020 to 0.012 (not significant, p = 0.003, respectively).

Pooled Analysis: Model Comparisons

Table IV shows the results of the pooled analyses based on the three different modelling approaches with respect to the handling of baseline TT4 measurements. The pre-specified method of baseline correction (i.e. subtraction) yielded the largest residual standard deviation, for TT4 AUC as well as for C_{max}. If uncorrected outcomes were applied, the residual standard deviations were reduced to around one-third and were further reduced by 10–20%

Table III. Pharmacokinetic variables (without baseline correction) of total serum thyroxine (TT4)

Tablet strength (μg)	Total dose (μg)	No. of participants	AUC tablets (μg • h/L) [mean ± SD]	AUC solution (μg • h/L) [mean ± SD]	Relative AUC (%; estimate) [90% CI]	Relative AUC with adjustment for baseline TT4 (%) [90% CI]	C_{max} tablets (μg/L) [mean ± SD]	C_{max} solution (μg/L) [mean ± SD]	Relative C_{max} (%; estimate) [90% CI]	Relative C_{max} with adjustment for baseline TT4 (%) [90% CI]
25	600	24	4598.30 ± 627.60	4663.52 ± 631.52	98.6 [96.5, 100.7]	99.1 [97.9, 100.3]	117.71 ± 16.34	119.38 ± 16.01	98.6 [95.6, 101.6]	99.1 [96.5, 101.6]
50	600	24	4699.19 ± 608.88	4775.48 ± 597.34	98.3 [96.4, 100.2]	98.1 [96.5, 99.6]	118.25 ± 16.01	119.88 ± 13.40	98.3 [95.7, 101.1]	98.1 [95.4, 100.7]
75	600	26	4620.50 ± 771.70	4681.46 ± 703.88	98.5 [95.8, 101.2]	99.6 [97.4, 101.8]	115.81 ± 19.73	119.47 ± 19.41	96.8 [93.2, 100.4]	97.7 [94.1, 101.3]
100	600	24	4707.98 ± 626.91	4702.36 ± 601.58	100.0 [97.4, 102.4]	99.2 [96.6, 101.6]	118.38 ± 15.83	120.56 ± 15.32	98.0 [94.2, 101.9]	97.1 [93.0, 101.1]
150	600	24	4401.79 ± 471.08	4527.62 ± 531.66	97.3 [94.9, 99.8]	97.5 [95.4, 99.7]	112.48 ± 12.20	115.26 ± 14.02	97.7 [94.8, 100.7]	97.9 [94.9, 100.7]
200	600	24	4555.31 ± 661.93	4605.98 ± 635.91	98.7 [96.4, 101.2]	99.1 [97.2, 100.9]	113.62 ± 17.29	117.93 ± 15.41	96.0 [92.7, 99.5]	96.2 [93.1, 99.3]
125	625	26	4733.47 ± 540.23	4734.48 ± 559.62	100.0 [97.6, 102.5]	100.3 [98.2, 102.4]	121.11 ± 14.34	121.81 ± 15.90	99.6 [96.3, 102.9]	99.8 [96.5, 103.0]
175	700	26	4882.15 ± 580.48	4980.55 ± 683.71	98.2 [96.4, 101.2]	98.7 [96.5, 100.8]	125.00 ± 14.45	126.98 ± 18.93	98.8 [95.6, 102.2]	99.4 [96.3, 102.5]

AUC = area under the concentration-time curve (without baseline correction); **C_{max}** = peak plasma concentration (without baseline correction).

when, additionally, an adjustment for log baseline TT4 (used as a covariate) was performed. The reductions were more pronounced in AUC than in C_{max} analyses. These observations are in good agreement with the observed width differences in the experiment-wise CI reported in table II and table III.

In the analyses of baseline-corrected outcomes, no significant trial-specific factor effects (interactions) were seen for the factors sequence and treatment. A significant trial*period interaction was seen in log AUC-C, indicating that the observed different trends in the trials were not random fluctuations. No significant sequence effects were observed in log AUC-C or C_{max}-C. As was expected, based on the results presented in figure 1 and table II, total dose was a significant confounder in the baseline-corrected analyses. After adjustment for all other factors, an increase of total dose from 600 to 625μg (4.2%) or to 700μg (16.6%) was estimated to increase log AUC-C by 0.076 (90% CI 0.034, 0.119) or 0.108 (90% CI 0.061, 0.154), respectively. The corresponding C_{max} effects were similar. A significant treatment effect was observed for log C_{max}-C (−0.039, 90% CI −0.066, −0.012). However, this small effect was in line with the equivalence claims, since the CI is well covered by the equivalence limits.

In the analyses of total outcomes without baseline TT4 adjustment, the AUC-related trial*period interaction was still present; however, this time accompanied by a significant global period effect (0.011, 90% CI −0.019, −0.002). Again, neither significant sequence effects nor significant total dose effects were observed. Significant treatment effects were detected, this time in log AUC (−0.015, 90% CI −0.023, −0.006) as well as in log C_{max} (−0.021, 90% CI −0.033, –0.009).

When log baseline TT4 was added to the models, it turned out to be an important confounder for log AUC as well as for log C_{max}. Log baseline TT4 explained 35.1% of the log AUC variation and 17.4% of the variation in the log C_{max} values. Ac-

Table IV. Residual variation, significant factors and covariates according to choice of model

Outcome	Residual standard deviation of final model	Significant interaction terms	Significant factors/covariates	Significant subject characteristics
Log AUC-C	0.157	Trial*period (p = 0.031)	Total dose (p = 0.015)	Nil
Log C_{max}-C	0.150	Nil	Total dose (p = 0.009) Treatment (tablet vs solution; p = 0.019)	Nil
Log AUC (without adjustment for baseline TT4)	0.048	Trial*period (p = 0.012)	Treatment (tablet vs solution; p = 0.005) Phase (p = 0.048)	Age (p = 0.001) Log thyroid volume (p = 0.042)
Log C_{max} (without adjustment for baseline TT4)	0.068	Nil	Treatment (tablet vs solution; p = 0.006)	Nil
Log AUC (with adjustment for baseline TT4)	0.039	Trial*period (p = 0.009)	Log baseline TT4 (p < 0.001) Total dose (p = 0.003) Treatment (tablet vs solution; p = 0.004)	Nil
Log C_{max} (with adjustment for baseline TT4)	0.062	Nil	Log baseline TT4 (p < 0.001) Total dose (p = 0.004) Treatment (tablet vs solution; p = 0.006)	Nil

AUC = area under the concentration-time curve (without baseline correction); **AUC-C** = baseline-corrected area under the concentration-time curve; **C_{max}** = peak plasma concentration (without baseline correction); **C_{max}-C** = baseline-corrected peak plasma concentration; **TT4** = total thyroxine.

cording to the resulting parameter estimates, an increase of log baseline TT4 by 0.1 (approximately 10%) is associated with a log AUC increase of 0.040 (90% CI 0.032, 0.047) and a log C_{max} increase of 0.040 (90% CI 0.029, 0.052). The AUC-related trial*period interaction and the total dose effect continued to be significant despite adjustment for baseline. Log AUC increased by 0.198 (90% CI 0.135, 0.261) or 0.217 (90% CI 0.153, 0.281) for 625μg or 700μg doses, respectively, log C_{max} by 0.037 (90% CI 0.021, 0.053) or 0.053 (90% CI 0.035, 0.071), respectively. Similar treatment effects were observed as previously, in log AUC (−0.012, 90% CI −0.019, −0.005) as well as in log C_{max} (−0.019, 90% CI −0.030, −0.008), resulting in a pooled relative bioavailability estimate of 98.8% (90% CI 98.1%, 99.5%).

Pooled Analysis: Seasonal Variation

As soon as month of year was added as an additional factor to one of the models, the trial*period interaction term and the period term lost significance. If the factor period and the trial*period interaction terms were replaced by the factor 'month of year', the model fit was comparable. Month was not significant if log AUC-C was the outcome, but significant with the outcome log total AUC, without (p = 0.005) and with adjustment for baseline TT4 (p = 0.046). Figure 2 demonstrates the marginal distribution of log AUC in the latter model. The study period did not cover more than 8 calendar months, therefore the seasonal variation is incompletely demonstrated. Baseline adjustment did not change the monthly distribution to a recognisable extent.

Pooled Analysis: Influence of Subject Characteristics on Bioavailability

Normal plots of the residual crossover differences were almost perfectly linear in each case. All the distributions were symmetric and unimodal. Only with log AUC-C residuals were a few outliers present. In the corresponding Kolomogorov-

Smirnov tests, normality was rejected for log AUC-C only ($p = 0.017$).

Log AUC-C or log C_{max}-C residual crossover differences were not associated with any of the subject characteristics considered. While log C_{max} crossover differences were not associated with any of the characteristics, log AUC differences were associated with age and log thyroid volume (table IV) as long as no baseline TT4 adjustment was performed. An increase of 10 years in age was associated with a log AUC difference decrease of 0.040 (90% CI 0.016, 0.063), i.e. a reduction of the relative bioavailability of L-T4 from tablets by approximately 4%. An increase of log thyroid volume of 0.1 (i.e. an approximately 10% increase of thyroid volume) was associated with a log AUC difference increase of 0.007 (95% CI 0.0003, 0.014), i.e. an increase of relative bioavailability of <1%. Both associations were not significant after baseline TT4 adjustment.

Drug Safety and Tolerability

During the entire study series comprising 179 participants, a total of three serious adverse events occurred; none of these was related to the study drugs. Of the 56 nonserious adverse events, a causal relation to the study drug appeared possible, or could not be excluded, in only 14 adverse events. All of these were mild in intensity and included problems of falling asleep, feeling uncomfortably hot or increased sensitivity to heat, increased sweating, increased resting heart rate up to 108 beats/minute, headache, soft stools, problems passing urine, feeling nervous for 1 or 2 days after study drug administration and increased exercise tachycardia for 4 days in one individual. None of the symptoms recurred in the same person.

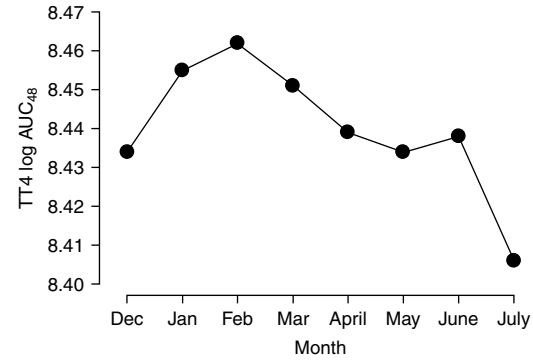

Fig. 2. Seasonal variation of log area under the concentration-time curve from 0 to 48 hours (AUC_{48}) of total thyroxine (TT4) after adjustment for baseline (marginal distribution).

Discussion

Rationale for the Chosen Study Design

For treatment with L-T4, solid oral preparations are generally used because medication errors can be avoided more readily than with a liquid formulation. However, in drugs with a narrow therapeutic range, such as L-T4, unwanted fluctuations of drug delivery as a result of poor dosage form design are a problem affecting both drug safety and effectiveness, and may represent a therapeutic hazard.[4-11] Thus, the performance of the dosage form in delivering the active compound to the systemic circulation, and thereby to the site of action, needs to be well controlled; assurance of adequate bioavailability may be vital.

When measuring serum hormone levels it is difficult to differentiate between endogenous secretion and the exogenous compound, especially when low (i.e. therapeutic) doses of the drug are administered. Therefore, former bioavailability and bioequivalence studies usually restricted direct determination of TT4 to a small number of values at variable timepoints,[29-33] or attempted to evaluate T4 concentrations indirectly by measuring the action of the drug as TSH suppression over time.[17,34] However, this is quite an insensitive method, especially with lower doses of L-T4. Another strategy to perform L-T4 bioavailability/bioequivalence studies that has been

applied[31,32] or suggested[35-38] is to investigate hypothyroid or athyroid patients with no endogenous hormone levels. However, this does not solve the problem of background thyroxine levels, since it is not possible to withdraw patients from their vital L-T4 supplements for study purposes. Therefore, baseline correction remains an issue in patients, just as in healthy individuals. Further flaws of study design, such as insufficient washout periods between consecutive study drug administrations,[12,39] add to the lack of valuable information to be drawn from the available published studies. As a consequence, the variables used to characterise systemic availability of L-T4 need to be refined.

Despite appreciation of the importance of formulation factors for drug effectiveness, it is only very recently that regulatory agencies have started to set official standards to ensure adequate L-T4 release in humans.[20] In 1998, at the time of study planning, no such recommendations were available. A bioequivalence study in females based on a single dose of L-T4 600μg, administered as two 300μg tablets, had shown a reasonable drug-induced increment of TT4 concentrations over baseline.[13] This principle of applying a pharmacological dose (instead of the lower, more physiological doses used for replacement therapy) in our study series was extended to the entire range of tablet strengths in order to generally avoid analytical problems. Close monitoring, particularly of initial hormone levels, confirmed that it is possible to assess drug absorption. The results support the concept of using a pharmacological L-T4 dose. However, for assessing relative bioavailability, the selection of the reference formulation may be crucial. A liquid formulation is regarded as the preferred reference.[20] This also proved to be a suitable procedure in our investigation. L-T4 absorption from the reference solution was prompt.

With the exception of the information discussed in the section on Adjustment for Baseline Levels, the present US FDA guidelines[20-22] as well as the new European guideline[19] consistently recommend (or at least allow the use of) the study design chosen in our investigation as the design of first choice for the demonstration of relative bioavailability. Our pooled data analysis confirms many of the detailed recommendations given in these guidelines and adds some new insight into the mechanisms that govern TT4 concentrations in healthy individuals.

Validity of the Crossover Design

The European guidelines valid at the time the trials were planned[18] suggested a two-period, crossover design as first choice for bioavailability and bioequivalence trials. Crossover designs are still generally recommended. They require smaller sample sizes than parallel-group designs since inter-individual variability is usually considerably higher than intraindividual variability. However, crossover designs can be devalidated by sequence effects, e.g. caused by a carry-over effect or a rebound in the second trial period. While L-T4 washout periods of 35 days are considered to be sufficient to prevent carry-over effects,[20] unfortunately the absence of sequence effects cannot usually be studied directly. For a single trial, the corresponding tests of sequence effects are discouraged for statistical reasons, as the power is too low with 24 subjects or less.[22] In our pooled data analysis we were able to test sequence effects and trial*sequence interactions with sufficient power; no sequence effects or trial*sequence interactions were found, regardless of which kind of baseline correction or adjustment was chosen. Therefore, we conclude that crossover designs should continue to be the first choice of design, as recommended in the guidelines.

Period Effects and Seasonal Variation

Crossover designs are ideally suited to control period effects. By design, period effects are balanced between random groups and, thus, do not bias treatment comparisons. Moreover, in the standard four-way ANOVA, period effects are explicitly modelled; therefore, their contribution to the residual variation in the trial is minimised. However, we found it useful to study period effects, since they

may indicate the presence of calendar effects that have to be taken into account if a parallel-group design has to be chosen for some reason, comparisons between trials are desired, or recruitment plans for future studies are to be developed.

In our studies, at least as far as AUC endpoints were concerned, period effects were indeed present; the directions of these effects differed between trials (significant trial*period interactions). These period effects could be almost completely explained by seasonal variation. The observed monthly distribution is in agreement with the findings in healthy individuals of various geographical regions[40-42] that showed highest TT4 values in winter and lowest values in summer, and seasonal variation of thyroid size in healthy males.[43] It is of note that in our studies C_{max} values were generally not subject to period effects and did not depend on season, probably because seasonal variation is predominantly represented in baseline TT4 levels that contribute less to a single maximum than to the total AUC.

In summary, seasonal variation will be present in L-T4 trials with TT4 AUC-related outcome, and should be taken into account in all comparisons between measurements taken at different times of the year. Crossover trials control these effects sufficiently. Otherwise, baseline TT4 adjustment will mitigate but not remove these effects, while explicit modelling of seasonal variation seems to have the best potential to remove any kind of time effects.

Adjustment for Baseline Levels

The general target of L-T4 bioavailability studies is "to detect T4 above baseline levels".[20] For this purpose, pharmacological doses of 600µg are given. However, it is an open issue as to whether this approach allows sufficient independence of endogenous L-T4 secretion or whether some kind of baseline adjustment is advantageous.

During the planning phase of the eight trials it was assumed that a simple subtraction of baseline levels, i.e. time zero TT4 values before drug application, may be an easy-to-do and sufficient way of coping with the problem of endogenous secretion. However, this notion was challenged by the consideration that endogenous secretion may decline to a relevant extent as a consequence of systemic effects of exogenous L-T4. While the assumed reduction of endogenous secretion could not be directly assessed, it would result in an overcorrection if time zero levels were simply subtracted. Moreover, the correction may cause statistical problems, since an additive subject effect is introduced before taking logs that cannot be adequately handled in the subsequent log model.

The results of our pooled analysis reveal another problem that is indispensably connected with the chosen baseline correction approach – the precision of the resulting estimate is insufficient. This is unavoidable since the determination of the baseline level is in itself imprecise. The standard error of the baseline-corrected AUC in the original scale will, thus, be approximately 1.4 (square root of two) times the standard error of the total AUC estimate. Since, at the same time, the mean baseline-corrected AUC values will be lower than the mean total AUC values, the coefficients of variation and the corresponding standard errors of the log AUC-C values will be increased substantially by the chosen method of baseline correction, as demonstrated in table II and table III. In summary, the pre-specified method of baseline correction increased the random error and may have overcorrected for baseline values and, thus, cannot be recommended for future studies.

However, from the deficiencies of the simple substraction method it cannot be concluded that it is best to ignore the baseline measurements completely in the statistical evaluation. Our analyses demonstrate that even after application of pharmacological doses that raise TT4 concentrations substantially over endogenous levels, AUC values are closely associated with baseline TT4. Furthermore, if baseline TT4 was ignored, log AUC crossover differences depended on season, age and thyroid volume. If a term for baseline TT4 adjustment was added to the model, these dependencies were substantially

reduced, and so too was the residual variation in the model. While the baseline adjustment by adding an additive term to the log model is certainly not a precise mathematical description of the relationship between baseline TT4 and total AUC observed, it seems to work well as an approximation to the underlying feedback process, while simple subtraction of baseline levels does not. This is in line with the general notion that adjustment by linear modelling is advantageous, even if the model is only a rough description of the underlying process.[44] Adjustment for baseline TT4 was less effective in C_{max} than in AUC analyses, presumably because, in general, maxima are less sensitive to baseline levels than areas.

In summary, it appears advisable that while the primary outcome of L-T4 bioavailability study should be log total AUC, the usual four-way ANOVA model should be supplemented by baseline TT4 as a covariate in order to minimise the influence of potential sources of bias and to reduce the residual variation as much as possible.

The issue of the appropriate method of baseline correction in L-T4 bioavailability studies was discussed recently during a meeting of the US FDA Advisory Committee for Pharmaceutical Science.[45] While the US FDA currently recommends using the mean of three pre-dose measurements for subtractive baseline correction, other more sophisticated methods of subtractive correction were suggested and controversially discussed. However, while some of these methods will certainly improve the precision and adequacy of baseline determination, the subsequent subtraction will inevitably increase the standard errors of the corrected log AUC values and, thus, result in a loss of precision. On the other hand, if the more sophisticated baseline determination methods are included in the models as covariates, they could potentially further improve the evaluation method suggested above. Whichever method turns out to be best, it is clear from the calculations above that baseline adjustment will be preferable to baseline subtraction as far as precision and study economy is concerned.

Assessment of Relative Bioavailability and Bioequivalence Claims

According to the pre-specified analysis method, equivalence was claimed for each of the respective dosage forms in the eight trials with respect to the reference solution, and the complete L-T4 product series met the standards required for drug approval. The ex post-analyses corroborated these findings. From this result, bioequivalence can be postulated for the six tablet strengths that were studied at the same total dose of 600μg.

In our pooled analyses, after adjustment for total dose, trial*treatment interactions were never significant, indicating that the bioequivalence claim may well be extended to the two remaining tablet strengths (125 and 175μg). Furthermore, the different tablet strengths can, thus, be combined to a pooled estimate of 98.8% relative bioavailability of tablet versus reference solution, significantly different from 100% but well within the acceptance region, because of the large sample size of the pooled data analysis. While this figure suggests the existence of a very small difference in bioavailability between tablets and drinking solution in general, it is of no concern for the equivalence claims and only of little concern for future sample size considerations. Thus, the L-T4 tablets in all eight dosage strengths, and the drinking solution, can be judged to be equivalent for all practical purposes.

Sample Size Considerations

Sample sizes in bioavailability and equivalence trials have to be minimised for ethical and economic reasons. With respect to L-T4, the requirement to minimise study burden was made more urgent by the fact that there was no agreement with the agencies to reduce the number of tablet strengths to be investigated by applying a dosage form proportionality argument. Thus, eight trials had to be performed for drug approval, multiplying the number

of individuals to be included and examinations to be performed. Based on the assumption of a residual standard deviation of 0.156 derived from the available literature, a sample size of at least 24 subjects per trial was determined to yield a power of 93% for each trial and 56% for the complete series. Hence, at least 384 single dose applications had to be performed (in fact, 396 were performed). From our pooled analysis, the residual standard deviation for the model corresponding to the pre-defined outcome log AUC-C was 0.157 (table IV) and, therefore, almost perfectly met the assumption.

Based on our estimates of total variances we found that if a parallel-group design was to be applied, $2 \times 46 = 92$ subjects, with 92 observation periods, per trial were required to yield the same power. Thus, this approach would approximately double the efforts, and 4-fold as many participants would have to be recruited. However, if baseline-corrected AUC had been replaced by total AUC with baseline TT4 adjustment, the residual standard deviation of the crossover model would have been reduced to approximately one-quarter of the assumed standard deviation (table IV). Thus, the sample size could have theoretically been reduced to one-sixteenth; however, for general considerations of robustness of design it should not be reduced to <12 participants. After all, with a sample size of 12 there seems to be sufficient efficiency reserves to generalise the study design to less homogeneous study populations and to cope with the small tablet versus drinking solution effect described in the earlier section on Assessment of Relative Bioavailability and Bioequivalence Claims. Alternatively, with sample sizes of 24, using the method of baseline adjustment, the recommended equivalence range could be reduced by up to one-quarter, allowing exclusion of AUC differences as low as 6% between drugs if bioequivalence is stated. Therefore, the suggested method of baseline adjustment seems to be "the more precise method for examining bioequivalence", called for recently in an editorial on oral L-T4 bioavailability testing.[46]

With 12 subjects per trial, study efforts would be reduced by one-half. Further reductions would have been possible if a perspective of dosage form proportionality had been taken that allowed the reduction of the set of tablet strengths studied to three (e.g. 25, 100 and 200μg to study the extremes and the most common tablet strengths). This would have further reduced the number of participants to a total of 36, with 72 observation periods, i.e. approximately only 20% of the actual efforts.

Conclusions

Our approach to develop a standard procedure for bioavailability assessment of hormone formulations by applying pharmacological doses instead of hormone replacement doses not only met the official standards of the German regulatory agency, but is also in agreement with recent standards published by the US FDA[20] and was successful in demonstrating equivalence between eight dosage forms of L-T4 and drinking solution. Nevertheless, a pooled analysis of the eight trials revealed that design and statistical evaluation should be modified in future studies. The two-way crossover design with washout periods of at least 35 days will continue to be the design of first choice, since it is efficient and takes care of the period effects to be expected in L-T4 trials. However, the primary outcome should be log total AUC. A subtractive baseline correction is not recommended. Instead, baseline TT4 determined directly before drug intake should be added to the standard crossover evaluation model as a covariate for the purpose of baseline adjustment. With this approach, sample sizes can be reduced to 12 volunteers per trial, even if less homogeneous groups that are representative of the general population are admitted for inclusion. If a series of dosage forms is studied, further reductions in sample size are only possible if a dosage form proportionality consideration is accepted that reduces the number of tablet strengths to be studied directly.

Acknowledgements

The study series was funded by Henning Berlin GmbH, succeeded by Sanofi-Synthelabo, now part of the Aventis Group. The pooled analysis was investigator initiated. The sponsor had no influence on the study conduct or on the analysis and publication of the data.

Dr J. Pfeilschifter reports having received lecture fees from Henning Berlin GmbH, succeeded by Sanofi-Synthelabo, now part of the Aventis Group.

Drs C. Clanget, R. Ding, C. Goeggelmannm, V. Hinke, M. Lang, Y. Tayrouz, I. Walter-Sack and K. Wegscheider have no conflicts of interest directly relevant to the content of this study.

The authors thank Mrs Karin Arnold, Mrs Andrea Deschlmayr, Mrs Jutta Kocher, Mrs Magdalena Longo and Mrs Gisela Schwahn for their excellent technical assistance, and Walter E. Haefeli, Professor of Medicine and Clinical Pharmacology, for his helpful discussions and assistance in reviewing the manuscript.

References

1. Cooper DS. Subclinical hypothyroidism. N Engl J Med 2001; 345: 260-5
2. Toft AD. Thyroxine therapy. N Engl J Med 1994; 331: 174-80
3. Fish LH, Schwartz HL, Cavanaugh J, et al. Replacement dose, metabolism, and bioavailability of levothyroxine in the treatment of hypothyroidism. N Engl J Med 1987; 316: 764-70
4. Olveira G, Almaraz MC, Soriguer F, et al. Altered bioavailability due to changes in the formulation of a commercial preparation of levothyroxine in patients with differentiated thyroid carcinoma. Clin Endocrinol 1997; 46: 707-11
5. Lee M. Equivalence of levothyroxine tablets [letter]. Drug Intell Clin Pharm 1983; 17: 265
6. Ramos-Gabatin A, Jacobson JM, Young RL. *In vivo* comparison of levothyroxine preparations. JAMA 1982; 247: 203-5
7. Jacobson JM, Ramos-Gabatin A, Young RL, et al. Nonequality of brand name thyroxine preparations [letter]. JAMA 1980; 243: 733
8. Stoffer SS. Drug substitution [letter]. JAMA 1979; 242: 1257
9. Stoffer SS. Substitution of levothyroxine products [letter]. JAMA 1979; 241: 1229
10. Drug Bioequivalence Study Panel. Drug bioequivalence: recommendations to the Office of Technology Assessment, Congress of the United States. J Pharmacokinet Biopharm 1974; 2: 433-66
11. Barr WH. Bioavailability of oral solid dosage forms and clinical response to drug therapy. In: Swarbrick J, editor. Current concepts in the pharmaceutical sciences: dosage form design and bioavailability. Philadelphia (PA): Lea & Febiger, 1973: 31-75
12. Ball P, Klinger W, Strubelt O. Untersuchungen zur relativen Bioverfügbarkeit von Laevothyroxin-Na nach oraler Gabe. Nuc-Compact 1987; 18: 92-6
13. Berg JA, Mayor GA. A study in normal human volunteers to compare the rate and extent of levothyroxine absorption from Synthyroid® and Levoxine®. J Clin Pharmacol 1993; 33: 1135-40
14. Spencer CA, LoPresti JS, Nicoloff JT, et al. Multiphasic thyrotropin responses to thyroid hormone administration in man. J Clin Endocrinol Metab 1995; 80: 854-9
15. Wenzel KW, Kirschsieper HE. Aspects of the absorption of oral L-thyroxine in normal man. Metabolism 1977; 26: 1-8
16. Wenzel KW, Meinhold H. Evidence of lower toxicity during thyroxine suppression after a single 3-mg l-thyroxine dose: comparison to the classical l-triiodothyronine test for thyroid suppressibility. J Clin Endocrinol Metab 1974; 38: 902-5
17. Wallack MS, Adelberg HM, Nicoloff JT. A thyroid suppression test using a single dose of L-thyroxine. N Engl J Med 1970; 283: 402-5
18. European Union guideline. Investigation of bioavailability and bioequivalence [online]. Available from URL: http://pharmacos.eudra.org/F2/eudralex/vol-3/pdfs-en/3cc15aen.pdf [Accessed 2004 Oct 7]
19. The European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products (CPMP). Note for guidance on the investigation of bioavailability and bioequivalence [online]. Available from URL: http://www.emea.eu.int/pdfs/human/ewp/140198en.pdf [Accessed 2004 Oct 7]
20. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Guidance for industry: levothyroxine sodium tablets: *in vivo* pharmacokinetic and bioavailability studies and *in vitro* dissolution testing [online]. Available from URL: http://www.fda.gov/cder/guidance/3645fnl.pdf [Accessed 2004 Oct 7]
21. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Guidance for industry: bioavailability and bioequivalence studies for orally administered drug products: general considerations [online]. Available from URL: http://www.fda.gov/cder/guidance/4964dft.pdf [Accessed 2004 Oct 7]
22. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Guidance for industry: statistical approaches to establishing bioequivalence [online]. Available from URL: http://www.fda.gov/cder/guidance/3616fnl.htm [Accessed 2004 Oct 7]
23. Piketty ML, d'Herbomez M, Le Guillouzic D, et al. Clinical comparison of three labelled-antibody immunoassays of free triiodothyronine. Clin Chem 1996; 42: 933-41
24. Price A, Bennett K, Giles M, et al. The ACS: 180 as a thyroid function test analyzer. Clin Chem Enzym Comm 1995; 6: 373-84
25. Liewendahl K, Melamies L, Helenius T, et al. Automated and manual serum free thyroxine assays evaluated with equilibrium dialysis. Scand J Clin Lab Invest 1994; 54: 347-51
26. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Stat Med 1995; 14: 2057-79
27. Senn S. Cross-over trials in clinical research. Chichester: Wiley, 1993
28. Goeggelmann C, Clanget C, Ding R, et al. Bioequivalence assessment of endogenous compounds like hormones: eight

levothyroxine (L-T4) formulations [abstract]. Eur J Clin Pharmacol 2002; 58: S94

29. Vaisman M, Spina LDC, Eksterman LF, et al. Comparative bioavailability of two oral L-thyroxine formulations after multiple dose administration in patients with hypothyroidism and its relation with therapeutic endpoints and dissolution profiles. Arzneimittel Forschung 2001; 51: 246-52
30. Förster G, Hansen C, Mörsch F, et al. Bioäquivalenz eines Kombinationspräparates aus Levothyroxin und Jodid im Vergleich zum Monopräparat Levothyroxin. Med Klin (Munich) 1998; 93: 401-6
31. Dong BJ, Hauck WW, Gambertoglio JG, et al. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. JAMA 1997; 277: 1205-13
32. Hennessey JV, Evaul JE, Tseng YC, et al. L-thyroxine dosage: a re-evaluation of therapy with contemporary preparations. Ann Intern Med 1986; 105: 11-5
33. Hennessey JV, Burman KD, Wartowsky L. The equivalency of two L-thyroxine preparations. Ann Intern Med 1985; 102: 770-3
34. Wenzel KW, Mehrländer U. Wirksamkeitsunterschiede von Levothyroxin-Präparaten. Dtsch Med Wochenschr 1988; 113: 53-8
35. Klein I. Evaluation of bioequivalence and efficacy of L-thyroxine preparations in the treatment of human thyroid disease [online]. Available from URL: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3926B2_03_King%20Pharmaceutical-Backgrounder.pdf [Accessed 2004 Oct 7]
36. DeGroot L. Letter presented at the FDA Advisory Committee for Pharmaceutical Science Meeting scheduled for March 13, 2003 [online]. Available from URL: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3926OPH2_10_DeGroot.pdf [Accessed 2004 Oct 7]
37. Hershman JM. Letter presented at the FDA Advisory Committee for Pharmaceutical Science Meeting scheduled for March 13, 2003 [online]. Available from URL: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3926OPH2_11_Hershman.pdf [Accessed 2004 Oct 7]
38. Baskin HJ. Letter presented at the FDA Advisory Committee for Pharmaceutical Science Meeting scheduled for March 13, 2003 [online]. Available from URL: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3926OPH2_16_Baskin.pdf [Accessed 2004 Oct 7]
39. Trantow T, Herzog R, Fuder H, et al. A pilot study on the determination of the relative bioavailability of levo-thyroxine. Methods Find Exp Clin Pharmacol 1995; 17: 333-43
40. Behall KM, Scholfield DJ, Hallfrisch JG, et al. Seasonal variation in plasma glucose and hormone levels in adult men and women. Am J Clin Nutr 1984; 40: 1352-6
41. Levine M, Duffy L, Moore DC, et al. Acclimation of a non-indigenous sub-artic population: seasonal variation in thyroid function in interior Alaska. Comp Biochem Physiol 1995; 111A: 209-14
42. Smals AGH, Ross HA, Kloppenborg PWC. Seasonal variation in serum T3 and T4 levels in man. J Clin Endocrinol Metab 1977; 44: 998-1001
43. Hegedüs L, Rasmussen N, Knudsen N. Seasonal variation in thyroid size in healthy males. Horm Metab Res 1987; 19: 391-2
44. The European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products (CPMP). Points to consider on adjustment for baseline covariates [online]. Available from URL: http://www.emea.eu.int/pdfs/human/ewp/286399en.pdf [Accessed 2004 Oct 7]
45. Food and Drug Administration, Advisory Committee for Pharmaceutical Science. Transcript of meeting on bioequivalence/bioavailability of endogenous products, March 13, 2003 [online]. Available from URL: http://www.fda.gov/ohrms/dockets/ac/03/transcripts/3926T2.htm [Accessed 2004 Oct 7]
46. Stockigt J. Testing the bioavailability of oral l-thyroxine by studying its absorption: smoke or mirrors? Thyroid 2004; 14: 167-8

Correspondence and offprints: Dr *Ingeborg Walter-Sack*, Department of Internal Medicine VI, Clinical Pharmacology and Pharmacoepidemiology, Medizinische Universitätsklinik, Im Neuenheimer Feld 410, Heidelberg, D-69 120, Germany.
E-mail: ingeborg_walter-sack@med.uni-heidelberg.de