

LEVOTHYROXINE BRAND AND WEIGHT-ADJUSTED DOSE REQUIREMENTS IN PRIMARY HYPOTHYROIDISM: REAL-WORLD EVIDENCE OF DOSE DIFFERENCES WITHOUT PHARMACODYNAMIC VARIATION

Nasazzi Doddi, Paula;¹ Feldman, Rodolfo Néstor;² Pustilnik, Estefanía;³ Bolzan, Diamela;⁴ Ramírez Stieben, Luis Agustín⁵

Servicio de Endocrinología del Grupo Gamma (Rosario, Argentina).

ARTICLE INFO

Keywords:

hypothyroidism, thyroxine, therapeutic equivalency, thyroid-stimulating hormone, pharmacokinetics

ABSTRACT

Levothyroxine (LT4) is commonly considered interchangeable across brands, yet small formulation differences may translate into different dose requirements in routine care. We compared three LT4 brands in adults with biochemically controlled hypothyroidism, focusing on weight-adjusted dose needed to maintain normal TSH and pharmacodynamic response. Materials and methods: observational cross-sectional study using a private outpatient database. Adults with primary hypothyroidism, TSH within the reference range, and etiology limited to Hashimoto's thyroiditis or post-thyroidectomy were included. Exposure was LT4 brand (A, B, C). Primary outcome was LT4 dose per kg ($\mu\text{g}/\text{kg}/\text{day}$). Pharmacodynamic outcome was log-transformed TSH modeled as a function of brand, dose per kg, age, sex, and etiology. Robust standard errors and propensity score-based inverse probability of treatment weighting (IPTW) were used to address confounding by non-random brand assignment. Brand $\times$ etiology interaction was explored. Results: we included 274 adults (A n=102, B n=152, C n=20). In adjusted models, brand was associated with dose per kg; Brand A required a higher dose than Brand B (adjusted difference $\approx 0.14 \mu\text{g}/\text{kg}/\text{day}$), while comparisons involving Brand C were imprecise and non-significant. Dose per kg decreased with age and was higher in post-thyroidectomy. In pharmacodynamic models, brand was not associated with log(TSH) after adjustment (global $p=0.58$), and results were consistent in IPTW-weighted analyses (global $p=0.68$). No significant brand $\times$ etiology interaction was detected. Discussion:

Received: February 10, 2026

Accepted: May 3, 2026

LT4 brand was associated with small differences in weight-adjusted dose requirements, but not with adjusted TSH response. These observational findings should be interpreted cautiously and support biochemical monitoring when treatment changes occur.

MARCA DE LEVOTIROXINA Y REQUERIMIENTOS DE DOSIS AJUSTADA POR PESO EN EL HIPOTIROIDISMO PRIMARIO: EVIDENCIA DEL MUNDO REAL DE DIFERENCIAS EN DOSIS SIN VARIACIÓN FARMACODINÁMICA

INFO ARTÍCULO

Palabras clave:

hipotiroidismo, tiroxina, equivalencia terapéutica, hormona estimulante de la tiroides, farmacocinética.

RESUMEN

La levotiroxina (LT4) suele considerarse intercambiable entre marcas; sin embargo, pequeñas diferencias de formulación podrían traducirse en distintos requerimientos de dosis en la práctica habitual. Comparamos tres marcas de LT4 en adultos con hipotiroidismo controlado bioquímicamente, enfocándonos en la dosis ajustada por peso necesaria para mantener la TSH en rango normal y en la respuesta farmacodinámica. Materiales y métodos: estudio observacional transversal utilizando una base de datos ambulatoria de un centro privado. Se incluyeron adultos con hipotiroidismo primario, TSH dentro del rango de referencia y etiología limitada a tiroiditis de Hashimoto o hipotiroidismo posttiroidectomía. La exposición fue la marca de LT4 (A, B, C). El desenlace primario fue la dosis de LT4 por kg ($\mu\text{g}/\text{kg}/\text{día}$). El desenlace farmacodinámico fue la TSH log-transformada, modelada en función de la marca, la dosis por kg, la edad, el sexo y la etiología. Se emplearon errores estándar robustos y ponderación por inversa de la probabilidad de tratamiento (IPTW) basada en puntaje de propensión para abordar la confusión por asignación no aleatoria de la marca. Se exploró la interacción marca $\times$ etiología. Resultados: se incluyeron 274 adultos (A n=102, B n=152, C n=20). En modelos ajustados, la marca se asoció con la dosis por kg: la Marca A requirió una dosis mayor que la Marca B (diferencia ajustada $\approx 0.14 \mu\text{g}/\text{kg}/\text{día}$), mientras que las comparaciones que involucraron a la Marca C fueron imprecisas y no significativas. La dosis por kg disminuyó con la edad y fue mayor en pacientes posttiroidectomía. En los modelos farmacodinámicos, la marca no se asoció con log(TSH) tras el ajuste (p global=0.58), y los resultados fueron consistentes en los análisis ponderados por IPTW (p global=0.68). No se detectó interacción significativa marca $\times$ etiología. Discusión: la marca de LT4 se asoció con pequeñas diferencias en los requerimientos de dosis ajustada por peso, pero no con la respuesta ajustada de TSH. Estos hallazgos observacionales deben interpretarse con cautela y respaldan el monitoreo bioquímico cuando se realizan cambios en el tratamiento.

Recibido: 10 de febrero de 2026

Aceptado: 3 de mayo de 2026

INTRODUCTION

Primary hypothyroidism is a common condition and, consequently, a substantial proportion of adults receive lifelong replacement therapy. The therapeutic standard consists of daily oral levothyroxine (LT4) administration, with dose titration aimed at achieving physiological concentrations of thyroid-stimulating hormone (TSH), seeking stable biochemical control and symptom minimization [1].

In clinical practice, multiple LT4 formulations and brands (reference and generic) coexist, raising the issue of their interchangeability. Although regulatory approval is based on bioequivalence criteria, specific methodological limitations have been highlighted for LT4, partly due to the presence of endogenous hormone and the longstanding debate regarding its behavior as a “narrow therapeutic index” drug, which may amplify the clinical impact of small exposure variations [1,2]. In this context, “switching” between preparations (especially if repeated) has been proposed to potentially associate with undesirable TSH fluctuations and, therefore, require closer monitoring [1].

Clinical and real-world evidence remains heterogeneous. In adults, a large observational study, controlling confounding through matching propensity score, found similar results between generic and brand-name LT4 regarding the likelihood of maintaining TSH within range during follow-up [3]. However, a systematic review of real-world evidence concluded that available studies remain limited and yield mixed results, particularly for clinical and economic outcomes [4]. Likewise, controlled crossover trials have shown divergent findings depending on the population: in a classic adult study, no significant differences were observed between preparations in serum thyroid hormone parameters under trial conditions [5], whereas in pediatric populations with severe congenital hypothyroidism, biochemical differences were reported, with lower TSH levels on the

brand-name preparation, suggesting that patient vulnerability may modulate the clinical relevance of small inter-product variations [6].

The potential impact of brand switching also becomes evident in “natural” forced-substitution scenarios. In a population-based cohort, a dose-equivalent switch from a specific brand to other preparations was associated with an increased proportion of patients with out-of-range TSH at 6 weeks, a pattern compatible with overtreatment in a substantial fraction, particularly at higher doses [7]. Taken together, these data support the need to comparatively evaluate brand performance under real-world conditions, even among patients with TSH within range, incorporating analytical methods that reduce confounding by indication.

Within this framework, and considering that brand choice in practice is not random, it is pertinent to examine whether differences exist between brands in: (i) the weight-adjusted dose required to maintain biochemical euthyroidism and (ii) the pharmacodynamic response, assessed through TSH (preferably on a logarithmic scale) and, when available, free thyroxine (fT4), with adjustment for relevant clinical and anthropometric variables and sensitivity analyses using propensity score-based methods [1,3].

Accordingly, we aimed to compare three LT4 brands in adults with primary hypothyroidism and normal TSH by evaluating (1) the weight-adjusted dose required to maintain biochemical euthyroidism and (2) the pharmacodynamic effect of brand on TSH (log-transformed) and free T4 at a given dose per kg, adjusting for potential confounders. In addition, we will assess the robustness of findings using propensity score adjustment and explore effect heterogeneity by etiology (Hashimoto's thyroiditis vs post-thyroidectomy).

MATERIALS AND METHODS

Study design and population

We conducted an observational, cross-sectional study using a database of adult patients (≥ 18 years), of both sexes, with

a diagnosis of primary hypothyroidism evaluated at a private outpatient center. Eligible patients were receiving LT4 therapy, had been receiving the same LT4 brand at a stable daily dose for at least 6 months before the TSH measurement used for analysis, and had serum TSH within the reference range (0.24-4.2 $\mu\text{IU/mL}$). We included only patients whose underlying etiology was Hashimoto's thyroiditis or post-thyroidectomy hypothyroidism. Patients with missing data for the primary analyses (brand, dose per kg, TSH, age, sex, and etiology) were excluded. Free T4 was available in a subset and analyzed in complete cases.

All compared LT4 products were commercially available oral solid tablet formulations. Liquid formulations, soft-gel capsules, or compounded preparations were not included. Brands were anonymized as A, B, and C for reporting purposes. Brand selection reflected routine clinical practice and was not based on a formal institutional prescribing protocol.

Variables and definitions

We collected age, sex, hypothyroidism etiology, LT4 brand (three brands, coded as A, B, and C), daily LT4 dose, and anthropometric measures (weight, height, and BMI). The primary exposure was LT4 brand. The primary outcome was the weight-adjusted LT4 dose ($\mu\text{g/kg/day}$). The main pharmacodynamic outcome was TSH ($\mu\text{IU/mL}$), modeled on the logarithmic scale. Free T4 (ng/dL) was available in a subset of participants and was reported descriptively as part of the baseline characteristics. For secondary analyses, TSH was categorized as <2.50 versus ≥ 2.50 $\mu\text{IU/mL}$.

Statistical analysis

Statistical analyses were conducted using linear and generalized regression models. Differences in weight-adjusted LT4 dose across brands were estimated using ANCOVA/regression models adjusting for age, sex, etiology, and TSH (modeled either

as a continuous variable or categorized as <2.50 vs ≥ 2.50 $\mu\text{IU/mL}$). Adjusted marginal means by brand and pairwise contrasts were reported with 95% confidence intervals, applying Tukey adjustment for multiple comparisons. To evaluate pharmacodynamic differences at a given dose, we fitted models with log-transformed TSH as the outcome and LT4 brand as the main predictor, adjusting for dose per kg, age, sex, and etiology. Because free T4 was available only in a subset of participants, it was reported descriptively and not used as a main pharmacodynamic endpoint. Effect modification by etiology was explored by including a brand $\times$ etiology interaction term when supported by the data. To address confounding by indication due to non-random brand assignment, propensity score-based approaches (inverse probability of treatment weighting [IPTW] and/or matching) were applied; the probability of receiving each brand was modeled using baseline covariates (age, sex, etiology, and available anthropometric variables), and primary analyses were replicated in the weighted/matched samples. Model assumptions (linearity, multicollinearity, and homoscedasticity) were assessed, and robust standard errors were used when appropriate. All analyses were performed in R (version 4.2.3), and statistical significance was defined as a two-sided p value <0.05 .

RESULTS

A total of 274 adults with primary hypothyroidism and TSH within the reference range were included. Etiology was Hashimoto's thyroiditis or post-thyroidectomy. Three LT4 brands were analyzed (A, B, and C) (Table 1).

Adjusted comparison of weight-based dose across brands (primary analysis)

In the ANCOVA model with dose per kg ($\mu\text{g/kg/day}$) as the outcome and TSH modeled as a continuous covariate, LT4 brand was independently associated with dose requirements after adjustment for age, sex, etiology, and TSH (Type II ANOVA $p=0.0216$)

Table 1. Baseline characteristics of adults with primary hypothyroidism and normal TSH, overall and stratified by levothyroxine brand.

Characteristic	Overall n=274	Brand A n=102	Brand B n=152	Brand C n=20	p value
Age, years	50.58 ± 12.43	51.26 ± 14.25	49.32 ± 10.90	56.65 ± 11.96	0.036
Female sex, n (%)	215 (78%)	84 (82%)	114 (75%)	17 (85%)	0.338
Etiology, n (%)					
Hashimoto's thyroiditis	224 (82%)	86 (84%)	123 (81%)	15 (75%)	0.500
Post-thyroidectomy	50 (18%)	16 (16%)	29 (19%)	5 (25%)	
TSH, µIU/mL	1.86 (1.13, 2.66)	1.79 (1.15, 2.22)	1.98 (1.14, 2.75)	2.20 (0.96, 2.96)	0.239
Daily LT4 dose, µg/day	95.49 ± 34.38	103.21 ± 38.21	91.06 ± 31.24	89.83 ± 30.93	0.016
Weight, kg	75.59 ± 17.07	75.83 ± 15.13	76.06 ± 18.73	70.87 ± 12.49	0.437
BMI, kg/m ²	28.08 ± 5.74	28.44 ± 5.55	27.94 ± 6.10	27.36 ± 3.69	0.674
Dose per kg, µg/kg/day	1.28 ± 0.42	1.37 ± 0.44	1.23 ± 0.41	1.28 ± 0.39	0.024
Free T4, ng/dL [†]	1.29 ± 0.19	1.35 ± 0.20	1.24 ± 0.16	1.28 ± 0.15	0.004

Values are mean ± SD, median (IQR), or n (%). * p values compare Brands A, B, and C and were obtained using one-way ANOVA for approximately normally distributed continuous variables, Kruskal-Wallis test for skewed continuous variables (TSH), and Fisher's exact test for categorical variables. Dose per kg (µg/kg/day) was calculated as daily LT4 dose (µg/day) divided by body weight (kg). † Free T4 was available in a subset of participants; analyses were performed using complete cases (no imputation). **Abbreviations:** LT4, levothyroxine; TSH, thyroid-stimulating hormone; BMI, body mass index; IQR, interquartile range; SD, standard deviation.

(Table 2). Age, etiology, and TSH were also independently associated with dose per kg, whereas sex was not.

Using adjusted marginal means, the estimated

mean dose per kg was 1.42 µg/kg/day (95% CI 1.33-1.52) for Brand A, 1.28 µg/kg/day (95% CI 1.20-1.36) for Brand B, and 1.38 µg/kg/day (95% CI 1.20-1.56) for

Table 2. ANCOVA (linear regression) for weight-adjusted daily LT4 dose (µg/kg/day), with TSH modeled continuously (main model).

Predictor	β (estimate)	SE	95% CI	p value
Intercept	1.826	0.119	1.591 to 2.060	<0.001
Brand: B vs A	-0.141	0.051	-0.241 to -0.040	0.006
Brand: C vs A	-0.043	0.097	-0.235 to 0.149	0.657
Age (per year)	-0.006	0.002	-0.010 to -0.002	0.003
Sex: Male vs Female	-0.039	0.058	-0.154 to 0.076	0.504
Etiology: Post-thyroidectomy vs Hashimoto	0.220	0.065	0.092 to 0.349	<0.001
TSH (per 1 µIU/mL)	-0.093	0.021	-0.135 to -0.051	<0.001

Outcome: LT4 dose (µg/kg/day). Reference levels: Brand A, Female, Hashimoto's thyroiditis. How to read β: it is the adjusted mean difference in µg/kg/day associated with each predictor, holding the others constant. For example, Brand B required on average 0.141 µg/kg/day less than Brand A after adjustment. *Robust SE sensitivity (HC1): results were consistent (e.g., Brand B vs A p=0.011; Age p=0.006; Etiology p<0.001; TSH p<0.001).*

Brand C (Table 3, Figure 1). In Tukey-adjusted pairwise contrasts, Brand A required a higher dose than Brand B (Δ 0.1405 $\mu\text{g/kg/day}$, 95% CI 0.0203 to 0.261, $p=0.0172$) (Table 3). No significant differences were observed between A vs C ($p=0.8968$) or B vs C ($p=0.5623$).

Covariate effects in the same model

indicated that dose per kg decreased with increasing age ($\beta= -0.00619 \mu\text{g/kg/day}$ per year, $p=0.0025$), was higher in post-thyroidectomy compared with Hashimoto's thyroiditis ($\beta= +0.220 \mu\text{g/kg/day}$, $p=0.0009$), and decreased with higher TSH within the normal range ($\beta= -0.0931 \mu\text{g/kg/day}$ per 1 $\mu\text{IU/mL}$, $p<0.001$).

Table 3. Adjusted marginal means and pairwise comparisons (emmeans).

Part A. Adjusted marginal means (Dose per kg, $\mu\text{g/kg/day}$)			
Brand	Adjusted mean	SE	95% CI
A	1.42	0.0492	1.33 to 1.52
B	1.28	0.0401	1.20 to 1.36
C	1.38	0.0930	1.20 to 1.56
Part B. Pairwise contrasts (Tukey-adjusted)			
Contrast	Δ ($\mu\text{g/kg/day}$)	95% CI	p (Tukey)
A – B	0.1405	0.0203 to 0.2610	0.0172
A – C	0.0433	-0.1863 to 0.2730	0.8968
B – C	-0.0972	-0.3209 to 0.1260	0.5623

Model: ANCOVA with TSH continuous (adjusted for age, sex, etiology, and TSH). Pairwise p values Tukey-adjusted.

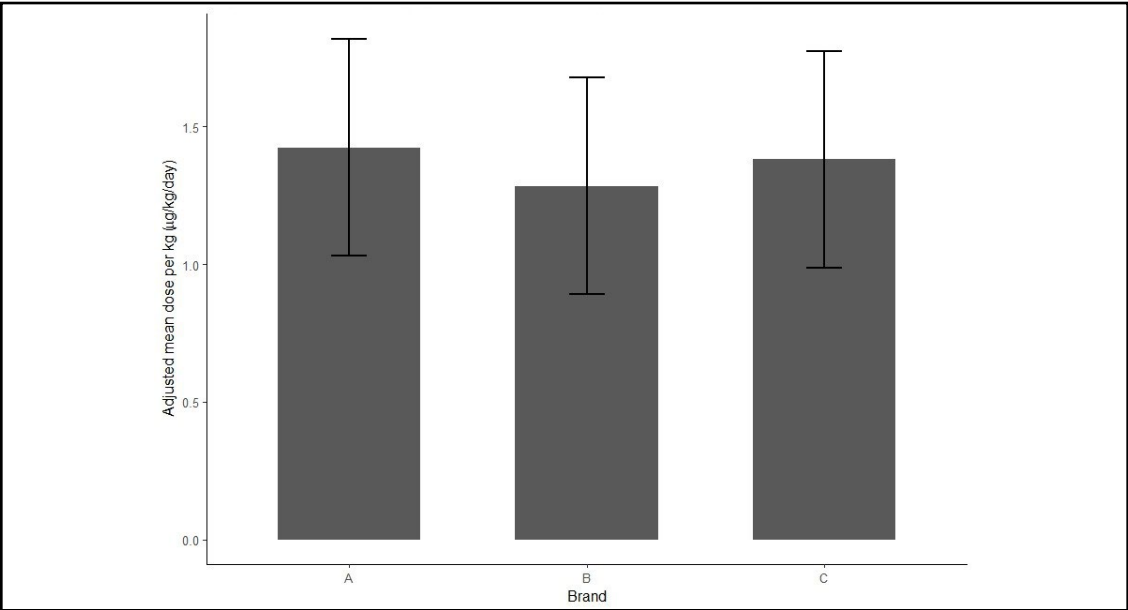


Figure 1. ANCOVA-adjusted weight-adjusted levothyroxine dose by brand (adjusted means with 95% confidence intervals). Bars show adjusted marginal means of LT4 dose ($\mu\text{g/kg/day}$) for each brand (A–C) estimated from a linear regression (ANCOVA) model adjusted for age, sex, etiology (Hashimoto's thyroiditis vs post-thyroidectomy), and serum TSH (continuous). Error bars indicate 95% confidence intervals derived from estimated marginal means. Pairwise comparisons of adjusted means were performed with Tukey correction (see Table 3).

Model diagnostics showed low multicollinearity (all VIF $\approx$ 1.0–1.1) and no evidence of heteroscedasticity ($p=0.408$). Results were consistent when using robust (HC1) standard errors, including the difference between Brands A and B ($p=0.0106$).

Sensitivity analysis using categorical TSH

When TSH was categorized as <2.50 vs ≥ 2.50 $\mu\text{IU/mL}$, brand remained significantly associated with dose per kg (Type II ANOVA $p=0.0285$). Adjusted means were $1.38 \mu\text{g/kg/day}$ (95% CI 1.28–1.49) for Brand A, $1.25 \mu\text{g/kg/day}$ (95% CI 1.17–1.33) for Brand B, and $1.35 \mu\text{g/kg/day}$ (95% CI 1.17–1.54) for Brand C. Again, Brand A required a higher dose than

Brand B ($\Delta 0.1347 \mu\text{g/kg/day}$, 95% CI 0.0139 to 0.256, Tukey-adjusted $p=0.0245$), with no significant differences for A vs C or B vs C.

Pharmacodynamic effect (TSH at a given weight-adjusted dose)

In the multivariable model with log-transformed TSH as the dependent variable, LT4 brand was not associated with TSH after adjustment for dose per kg, age, sex, and etiology (global test for brand: $p=0.579$). Consistently, observed data and model-predicted TSH across the range of weight-adjusted LT4 dose showed near-overlapping brand-specific prediction curves from the multivariable (unweighted) model (Figure 2).

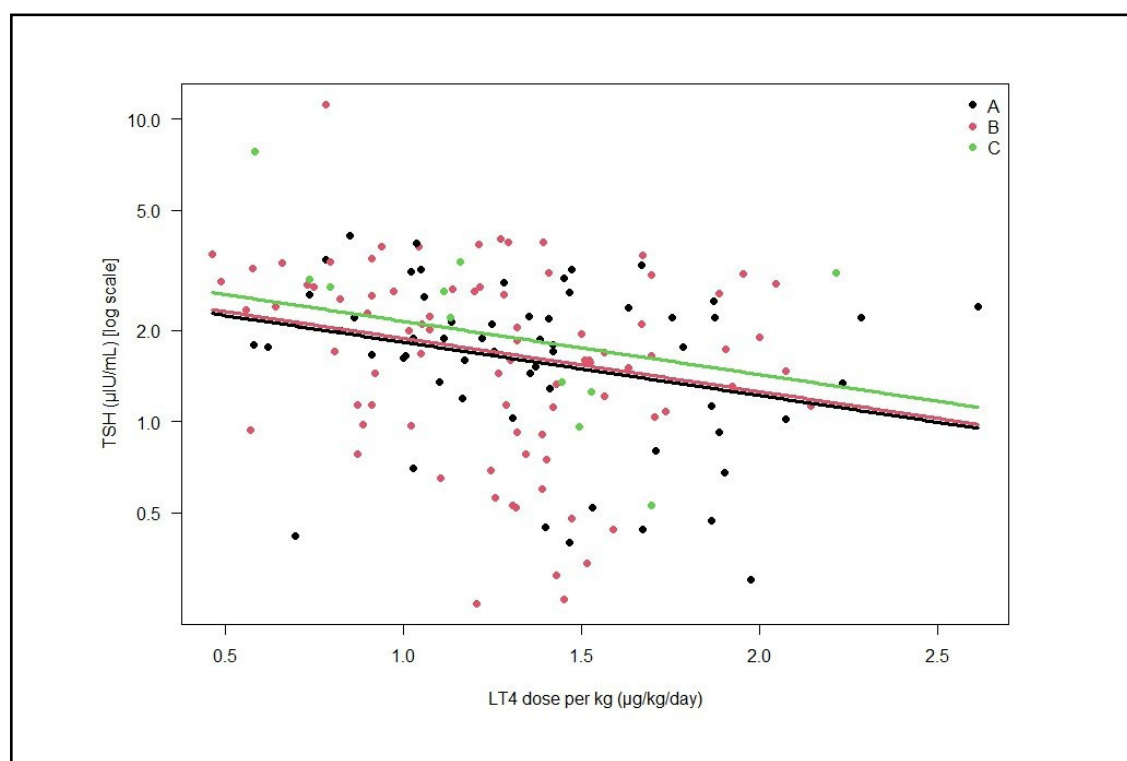


Figure 2. Observed and model-predicted TSH across weight-adjusted levothyroxine dose by brand. Points represent individual TSH measurements ($\mu\text{IU/mL}$) displayed on a logarithmic scale versus weight-adjusted LT4 dose ($\mu\text{g/kg/day}$), stratified by brand (A–C). Solid lines show model-based predictions from the multivariable regression with log-transformed TSH as the outcome, including LT4 dose per kg, age, sex, and etiology (Hashimoto's thyroiditis vs post-thyroidectomy). Predicted values were back-transformed to the TSH scale for display; near-overlapping curves indicate no meaningful between-brand differences in TSH at a given dose per kg after adjustment.

Using robust standard errors, neither Brand B nor Brand C differed from Brand A. In contrast, higher dose per kg was independently associated with lower TSH, and age was inversely associated with TSH.

IPTW / propensity-score-weighted pharmacodynamic analysis

To address confounding due to non-random brand selection, we conducted a propensity score-based IPTW analysis and refitted the pharmacodynamic model using survey-weighted regression. In the IPTW-weighted model with log-transformed TSH as the outcome and LT4 brand as the main predictor, adjusted for dose per kg, age, sex, and etiology, brand was again not associated with log(TSH) (global Wald test for brand: $F=0.385$; $df = 2$ and 267 ; $p=0.681$). Consistent with this, neither Brand B nor Brand C differed significantly from Brand A (Brand B vs A: $\beta=0.015$, $p=0.853$; Brand C vs A: $\beta=0.143$, $p=0.382$). In contrast, higher weight-adjusted LT4 dose was independently associated with lower log(TSH) ($\beta=-0.376$, $p<0.001$), and age was also inversely associated with log(TSH) ($\beta=-0.0086$, $p=0.0067$), while sex and etiology were not statistically significant predictors in the weighted model.

Subgroup and effect-modification analyses (etiology)

Cell counts supported estimation across all brand-etiology strata (Hashimoto: A $n=86$, B $n=123$, C $n=15$; post-thyroidectomy: A $n=16$, B $n=29$, C $n=5$). In the Hashimoto subgroup, the overall adjusted difference in weight-adjusted LT4 dose across brands did not reach statistical significance (Type II test $p=0.103$). Nonetheless, adjusted marginal means suggested a trend toward higher dose for brand A compared with brand B (A $1.30 \mu\text{g/kg/day}$ [95% CI 1.20 - 1.40] vs B 1.17 [1.09 - 1.25]; Tukey-adjusted $p=0.0838$), while comparisons involving brand C were imprecise given the smaller sample.

In the full cohort, there was no evidence of effect modification by etiology (Brand×Etiology

interaction $p=0.755$). Etiology-stratified adjusted marginal means showed the same qualitative ordering ($A > B$) in both etiologic groups, but pairwise contrasts within strata were not statistically significant after Tukey adjustment (Hashimoto: A–B $p=0.051$; post-thyroidectomy: A–B $p=0.338$), consistent with limited power in the smaller post-thyroidectomy group and sparse data for brand C.

DISCUSSION

In this cohort of adults with primary hypothyroidism and TSH values within the reference range, we observed two main findings. First, the LT4 brand was independently associated with the weight-adjusted dose ($\mu\text{g/kg/day}$) required to maintain biochemical euthyroidism: after adjustment for age, sex, etiology, and TSH, Brand A required a higher dose than Brand B (adjusted difference $\approx 0.14 \mu\text{g/kg/day}$), whereas comparisons involving Brand C were less precise and did not reach statistical significance. Second, when evaluating the pharmacodynamic component—TSH (log scale) at an equivalent dose/kg—no brand-related differences were identified in the multivariable model. This lack of association persisted in the propensity score-weighted analysis (IPTW), suggesting that, within the TSH range considered “normal,” the overall biochemical response is comparable across brands at a given weight-adjusted dose.

Interpreting why one brand was associated with a higher average weight-adjusted dose is complex in an observational design. The observed difference was modest and should be interpreted cautiously. Although formulation-related factors such as potency, dissolution characteristics, excipients, or bioavailability have been discussed in the broader LT4 literature, the present study was not designed to evaluate these mechanisms and cannot determine whether the observed association reflects true pharmaceutical differences between products [1,2,8-12]. Alternative explanations, including residual

indication bias, adherence patterns, previous titration history, prescribing preferences, comedications, and clinical follow-up practices, remain plausible and cannot be fully excluded. At the same time, controlled clinical evidence in adults has shown, in certain contexts, no significant differences between preparations, suggesting that any observed dose differences in routine care should not be interpreted as evidence of therapeutic inequivalence without longitudinal or interventional confirmation [5,13].

Another, non-mutually exclusive explanation is indication/brand-selection bias: it is plausible that certain patient profiles (e.g., more complex cases, those with greater variability, or with a history of frequent dose adjustments) are preferentially channeled toward a given brand, which could distort “crude” comparisons. In this regard, real-world studies have highlighted heterogeneity in outcomes and the need for analytical designs that mitigate confounding when interpreting differences between products [3,4,14,15]. Moreover, forced-substitution scenarios have shown that switching between dose-equivalent brands may be associated with a higher proportion of out-of-range TSH values during follow-up, reinforcing that clinical equivalence may depend on context and patient characteristics [6,7,15]. In vulnerable populations, such as severe congenital hypothyroidism, biochemical differences between brand and generic formulations have also been described [6]. In our study, propensity score-weighted analysis (IPTW) adds value by replicating the main pattern ($A > B$ in dose/kg) and suggests robustness to non-random assignment; however, it should be acknowledged that balance across groups does not always reach ideal thresholds for all covariates, and residual confounding may persist.

The fact that brand was not associated with $\log(\text{TSH})$ at an equivalent dose/kg is relevant from the perspective of “fine clinical equivalence.” In practical terms, it suggests that once the dose is adjusted (as occurs in

routine practice), final biochemical control—at least as measured by TSH within the normal range—does not differ systematically across brands. This is consistent with a scenario in which the main difference between brands is expressed as titration requirement (i.e., how much needs to be prescribed to reach target), rather than as persistent discrepancies in the pharmacodynamic marker once euthyroidism has been achieved [16-18].

Beyond the effect attributable to brand, the model allowed exploration of clinical determinants of LT4 requirement and biochemical response, providing elements of internal validity to the analysis. The covariates showed clinically plausible associations. In the dose/kg model, post-thyroidectomy etiology was associated with higher weight-adjusted dose requirements, whereas age was associated with lower dose/kg. Likewise, even within the normal range, an inverse association between TSH and dose/kg was observed, consistent with routine titration practice (higher doses tending to accompany lower TSH values). These relationships align with the use of TSH as the principal biomarker of therapeutic adequacy and with the stepwise LT4 dose-adjustment schemes described in clinical guidelines and specialized literature, also functioning as “internal controls” that reinforce the plausibility of the model [12,17,19,20].

Considering that LT4 requirements are not uniform across etiologies—with potentially different pharmacodynamic profiles and therapeutic targets—it was clinically relevant to explore whether the equivalence observed between brands in the overall analysis remained consistent within specific etiologic subgroups. Subgroup analyses and brand $\times$ etiology interaction testing did not demonstrate statistically significant heterogeneity. In the Hashimoto subgroup, a trend ($A > B$) was observed but did not reach significance after adjustment for multiplicity, while in the post-thyroidectomy subgroup estimates were less precise, particularly due to the small size of some cells (notably Brand

C). This suggests that, if true heterogeneity of the brand effect by etiology exists, this study may have been underpowered to detect it with precision, leaving the question open for larger series or multicenter analyses. This interpretation is consistent with literature showing that the clinical performance of different LT4 formulations may vary according to pathophysiologic context and the population evaluated, particularly in settings of greater biological vulnerability or extreme requirements [6,11], as well as with data underscoring the complexity of extrapolating pharmacokinetic bioequivalence to therapeutic equivalence across diverse clinical populations [9,21].

This study has several methodological strengths: the use of adjusted models (ANCOVA/regression) with estimation of marginal means and multiplicity-corrected comparisons, assessment of model assumptions, and sensitivity analyses with robust standard errors, as well as a confounding-control component through propensity score methods. However, inherent limitations must be acknowledged. First, the cross-sectional design does not allow capture of longitudinal changes, titration processes, or brand-switching events. Second, because inclusion was restricted to patients with TSH within the reference range, the cohort represents individuals who had already achieved biochemical control under their current LT4 regimen. Therefore, our findings should not be extrapolated to patients with uncontrolled hypothyroidism, unstable TSH values, recent brand switching, poor adherence, or difficulties achieving target TSH. This restriction was consistent with the study objective —estimating dose requirements among biochemically controlled patients— but limits inference regarding the comparative performance of LT4 brands in the broader treated population. Third, there is potential for residual confounding from unmeasured variables such as treatment adherence, comedication, dosing schedule, pharmacologic interactions, recent dose or

brand changes, comorbidities, individualized clinical targets, prescribing preferences, and previous titration history. Moreover, the observational, real-world nature of the study affects the quality and completeness of available information, with possible heterogeneity in clinical records and measurement conditions. Preanalytical and analytical variability in TSH determination constitutes an additional potential source of uncontrolled dispersion. Finally, free T4 was available only in a subset of participants and was therefore reported descriptively rather than analyzed as a formal adjusted pharmacodynamic endpoint. In addition, the relatively modest sample size, particularly in the Brand C group, reduced the precision of some estimates and the power to detect small differences or interactions.

Taken together, the findings raise a clinically pertinent “real-world” signal: differences between brands may exist in the average weight-adjusted dose required to maintain normal TSH, even when the final pharmacodynamic response (TSH) at an equivalent dose/kg does not show detectable differences. This pattern is consistent with the longstanding debate on the interchangeability of LT4 formulations and with the possibility of small variations in bioavailability or pharmaceutical performance within accepted bioequivalence margins [9,21,22]. These observations carry practical implications for follow-up: when brand switches occur, it may be prudent to reinforce biochemical monitoring within timeframes consistent with guidelines and clinical practice, particularly in subpopulations with a narrower therapeutic margin [15,17].

CONCLUSION

In adults with primary hypothyroidism and TSH within the reference range, LT4 brand was statistically associated with small differences in weight-adjusted dose requirements, mainly driven by a higher adjusted dose for Brand A compared with Brand B. However, given the observational cross-sectional

design, this finding should be interpreted as hypothesis-generating and does not establish pharmacological or therapeutic inequivalence between formulations. No between-brand differences were observed in log-transformed TSH at a given dose per kg, including in IPTW analyses. These results support careful biochemical monitoring in routine practice, particularly when brand changes occur, while underscoring the need for longitudinal or interventional studies to determine whether such dose differences reflect formulation-related effects or residual confounding.

Conflictos de intereses

Los autores declaran no presentar conflictos de intereses.

Financiamiento

No se recibieron fondos ni apoyos económicos institucionales, públicos ni privados para la realización de este trabajo.

Contribuciones de los autores

Todos los autores contribuyeron de manera sustancial a la atención clínica y/o la recolección e interpretación de datos, así como a la redacción y revisión crítica del manuscrito, y aprobaron la versión final para su publicación.

Consideraciones éticas

Este trabajo se llevó a cabo conforme a los principios de la Declaración de Helsinki. El estudio fue evaluado y aprobado por el Comité de Ética en Investigación de Grupo Gamma.

REFERENCES

1. Benvenga S, Carlé A. Levothyroxine Formulations: Pharmacological and Clinical Implications of Generic Substitution. *Adv Ther.* 2019; 36(Suppl 2):59-71. doi: 10.1007/s12325-019-01079-1.
2. Bolton S. Bioequivalence studies for levothyroxine. *AAPS J.* 2005; 7(1):E47-E53. doi: 10.1208/aapsj070106.
3. Brito JP, Ross JS, Sangaralingham L, Dutcher SK, Graham DJ, Wang Z, et al. Comparative Effectiveness of Generic vs Brand-Name Levothyroxine in Achieving Normal Thyrotropin Levels. *JAMA Netw Open.* 2020; 3(9):e2017645. doi: 10.1001/jamanetworkopen.2020.17645.
4. Qian J, Tanni KA. Real world evidence in effectiveness, safety, and cost savings of generic levothyroxine: a systematic review. *Endocrine.* 2021; 74(2):228-34. doi: 10.1007/s12020-021-02833-8.
5. Dong BJ, Hauck WW, Gambertoglio JG, Gee L, White JR, Bubp JL, Greenspan FS. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. *JAMA.* 1997; 277(15):1205-13.
6. Carswell JM, Gordon JH, Popovsky E, Hale A, Brown RS. Generic and brand-name L-thyroxine are not bioequivalent for children with severe congenital hypothyroidism. *J Clin Endocrinol Metab.* 2013; 98(2):610-617. doi: 10.1210/jc.2012-3125.
7. Flinterman LE, Kuiper JG, Korevaar JC, van Dijk L, Hek K, Houben E, et al. Impact of a Forced Dose-Equivalent Levothyroxine Brand Switch on Plasma Thyrotropin: A Cohort Study. *Thyroid.* 2020; 30(6):821-8. doi: 10.1089/thy.2019.0414.
8. Morais JA, Lobato Mdo R. The new European Medicines Agency guideline on the investigation of bioequivalence. *Basic Clin Pharmacol Toxicol.* 2010; 106(3):221-5. doi: 10.1111/j.1742-7843.2009.00518.x.
9. Blakesley VA. Current methodology to assess bioequivalence of levothyroxine sodium products is inadequate. *AAPS J.* 2005; 7(1):E42-6. doi: 10.1208/aapsj070105.
10. Burman K, Hennessey J, McDermott M, Wartofsky L, Emerson C. The FDA revises requirements for levothyroxine products. *Thyroid.* 2008; 18(5):487-90. doi: 10.1089/thy.2008.0109.
11. Hennessey JV. Generic vs name brand L-thyroxine products: interchangeable or still not? *J Clin Endocrinol Metab.* 2013; 98(2):511-4. doi: 10.1210/jc.2012-4310.
12. Garber JR, Hennessey JV. Generic levothyroxine: What is all the fuss about? *Endocr Pract.* 2005; 11(3):205-7. doi: 10.4158/EP.11.3.205.
13. Colucci P, D'Angelo P, Mautone G, Scarsi C, Ducharme MP. Pharmacokinetic equivalence of a levothyroxine sodium soft capsule manufactured

- using the new food and drug administration potency guidelines in healthy volunteers under fasting conditions. *Ther Drug Monit.* 2011; 33(3):355-61. doi: 10.1097/FTD.0b013e318217b69f.
14. Ross JS, Rohde S, Sangaralingham L, Brito JP, Choi L, Dutcher SK, et al. Generic and Brand-Name Thyroid Hormone Drug Use Among Commercially Insured and Medicare Beneficiaries, 2007 Through 2016. *J Clin Endocrinol Metab.* 2019; 104(6):2305-14. doi: 10.1210/jc.2018-02197.
15. Hennessey JV, Malabanan AO, Haugen BR, Levy EG. Adverse event reporting in patients treated with levothyroxine: results of the pharmacovigilance task force survey of the american thyroid association, american association of clinical endocrinologists, and the endocrine society. *Endocr Pract.* 2010; 16(3):357-70. doi: 10.4158/EP0362.OR.
16. Hennessey JV. The emergence of levothyroxine as a treatment for hypothyroidism. *Endocrine.* 2017; 55(1):6-18. doi: 10.1007/s12020-016-1199-8.
17. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al; American Thyroid Association Task Force on Thyroid Hormone Replacement. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. *Thyroid.* 2014; 24(12):1670-751. doi: 10.1089/thy.2014.0028.
18. Collier JW, Shah RB, Gupta A, Sayeed V, Habib MJ, Khan MA. Influence of formulation and processing factors on stability of levothyroxine sodium pentahydrate. *AAPS PharmSciTech.* 2010; 11(2):818-25. doi: 10.1208/s12249-010-9434-8.
19. Ramírez Stieben LA, Pustilnik E, Feldman R, Paladini L, Mancinelli L, Pellizzón NA, Schwarzstein D. Optimal levothyroxine dose to achieve euthyroidism in patients with primary hypothyroidism: analysis according to etiology. *Rev Fac Cien Med Univ Nac Cordoba.* 2022; 79(4):353-7. doi: 10.31053/1853.0605.v79.n4.35157.
20. Pustilnik E, Schwarzstein D, Feldman R, Mancinelli L, Paladini L, Pellizzón NA, Ramírez Stieben LA. The influence of age and body weight on levothyroxine replacement dosage to achieve euthyroidism in patients with primary hypothyroidism. *Endocrinol Diabetes Nutr (Engl Ed).* 2021; 68(8):542-47. doi: 10.1016/j.endien.2021.11.014.
21. American Thyroid Association; Endocrine Society; American Association of Clinical Endocrinologists. Joint statement on the U.S. Food and Drug Administration's decision regarding bioequivalence of levothyroxine sodium. *Thyroid.* 2004; 14(7):486. doi: 10.1089/1050725041517138.
22. Lipp HP, Hostalek U. A new formulation of levothyroxine engineered to meet new specification standards. *Curr Med Res Opin.* 2019; 35(1):147-50. doi: 10.1080/03007995.2018.1545635.