

## Research Article

# Lessons from Randomised Clinical Trials for Triiodothyronine Treatment of Hypothyroidism: Have They Achieved Their Objectives?

Rudolf Hoermann <sup>1</sup>, John E. M. Midgley,<sup>2</sup> Rolf Larisch,<sup>1</sup> and Johannes W. Dietrich <sup>3,4</sup>

<sup>1</sup>Department for Nuclear Medicine, Klinikum Lüdenscheid, Paulmannshöher Str. 14, 58515 Lüdenscheid, Germany

<sup>2</sup>North Lakes Clinical, 20 Wheatley Avenue, Ilkley LS29 8PT, UK

<sup>3</sup>Medical Department I, Endocrinology and Diabetology, Bergmannsheil University Hospitals, Ruhr University of Bochum, Buerkle-de-la-Camp-Platz 1, D-44789 Bochum, Germany

<sup>4</sup>Ruhr Center for Rare Diseases (CeSER), Ruhr University of Bochum and Witten/Herdecke University, Alexandrinenstr. 5, D-44791 Bochum, Germany

Correspondence should be addressed to Rudolf Hoermann; [rudolf.hoermann@gmail.com](mailto:rudolf.hoermann@gmail.com)

Received 25 May 2018; Revised 28 June 2018; Accepted 4 July 2018; Published 16 July 2018

Academic Editor: Massimo Tonacchera

Copyright © 2018 Rudolf Hoermann et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Randomised controlled trials are deemed to be the strongest class of evidence in evidence-based medicine. Failure of trials to prove superiority of T3/T4 combination therapy over standard LT4 monotherapy has greatly influenced guidelines, while not resolving the ongoing debate. Novel studies have recently produced more evidence from the examination of homeostatic equilibria in humans and experimental treatment protocols in animals. This has exacerbated a serious disagreement with evidence from the clinical trials. We contrasted the weight of statistical evidence against strong physiological counterarguments. Revisiting this controversy, we identify areas of improvement for trial design related to validation and sensitivity of QoL instruments, patient selection, statistical power, collider stratification bias, and response heterogeneity to treatment. Given the high individuality expressed by thyroid hormones, their interrelationships, and shifted comfort zones, the response to LT4 treatment produces a statistical amalgamation bias (Simpson's paradox), which has a key influence on interpretation. In addition to drug efficacy, as tested by RCTs, efficiency in clinical practice and safety profiles requires reevaluation. Accordingly, results from RCTs remain ambiguous and should therefore not prevail over physiologically based counterarguments. In giving more weight to other forms of valid evidence which contradict key assumptions of historic trials, current treatment options should remain open and rely on personalised biochemical treatment targets. Optimal treatment choices should be guided by strict requirements of organizations such as the FDA, demanding treatment effects to be estimated under actual conditions of use. Various improvements in design and analysis are recommended for future randomised controlled T3/T4 combination trials.

## 1. Introduction

Levothyroxine (LT4) substitution remains the standard treatment for patients with hypothyroidism [1]. The evidence base has been repeatedly reviewed by experts and authors of various guidelines [2–6]. Alternative treatments such as T3/T4 combination therapy have not been proven to be a superior option to LT4 monotherapy in randomised controlled trials (RCTs) including meta-analyses comparing the two different modes of treatment [7]. However, this has

not defused the situation, which continues to be one of the major controversies in thyroidology.

A disturbing fact has been the existence of a substantial number of patients with persistent symptoms on LT4 treatment. Complaints have been expressed in various forms on the internet and in surveys and have been substantiated by more rigorous retrospective and prospective studies [8–11]. Where subjective dissatisfaction with the received treatment has been acknowledged, opinions are divided about the cause. Nonthyroidal influences such as living with a chronic

disease, the burden of a debilitating or malignant disease, the autoimmune process, or obesity have frequently been cited [12–16]. Others have proposed dose inadequacy of LT4 rather than treatment modality to play an important role [2]. The European Thyroid Association (ETA) has issued guidelines supporting combination therapy for patients with persistent symptoms as a valid option to clinicians, but only on an experimental and closely monitored basis [12]. Mirroring the disagreement about the reference range for TSH, the controversy about optimum treatment has been exhaustively reviewed, but not resolved for a decade [17].

## 2. A Fresh Look at the Old Controversy

This raises the question whether there are any newer considerations that may contribute toward a resolution, which is clearly in the interest of patients and clinicians. Although no more RCTs have been recently conducted on the subject, novel insights have emerged from the examination of homeostatic equilibria and experimental treatment protocols in the rodent [19–23]. Generally, if a physiological chain of causal arguments leads up to the final clinical proof, issues can be resolved in a satisfactory way. This is however not the case with T3/T4 therapy. On the contrary, disagreements have widened between the supreme category of evidence in evidence-based medicine (EBM), which can only be derived from RCTs, and other experimental findings.

At least three physiological phenomena challenge the RCT-derived evidence. (1) FT3 concentrations tend to be invariantly lower in LT4-treated patients particularly in athyreotic patients where in the absence of a thyroid gland conversion of T3 from T4 is inefficient, compared to the healthy subject [23]. (2) Experiments performed some 20 years ago in rodents [24], when recently repeated [20], demonstrated that it may be virtually impossible for LT4 treatment alone to restore euthyroidism at the level of various tissues, an endpoint which cannot be readily studied in humans. (3) Thyroid regulation in LT4-treated patients, interrelationships, and homeostatic equilibria between FT3, FT4, and TSH differ from those in healthy subjects [25]. Confirming subjective dissatisfaction with LT4 treatment, hypothyroid patients treated with standard LT4 monotherapy failed in carefully conducted prospective clinical trials to restore their quality of life to a level observed in the healthy population [10].

These contradicting-rather-than-complementing results from different classes of evidence cause a major problem for clinicians and patients alike. Therefore, should we go so far in the formal pursuit of EBM guidelines to insist that RCTs should remain unchallenged and must prevail over physiologically based counterarguments? Is there an important gap in this logic?

## 3. Lessons Learned from RCTs on Combination Treatment

Historic RCTs on T3/T4 treatment have been subjected to meta-analysis and exhaustively reviewed [1–9, 12, 17]. Their inability to prove superiority of the addition of a T3

component to LT4 monotherapy has been used as a strong argument in favour of LT4 standard treatment by guidelines [3, 4]. However, this is more a choice of noninferiority, because superiority has not been established for the standard treatment either. Also, no formal clinical trials have been performed to evaluate the current standard treatment against other treatment modalities such as the original use of natural desiccated thyroid extract (NDT) before that decision was made. Only later was one such trial conducted [26]. Considering other physiologically derived evidence, the alternative option of T3/T4 would seem to be equally justified by RCT outcomes. Non-RCT criteria have strongly influenced preference, foremost a perceived fear of increased adverse T3 effects. Unlike the clinically infrequently used LT3, LT4 has a well-documented history of use and safety profile. As a drug, LT3 is generally more difficult to manage than the prohormone LT4, due to its higher biological activity, shorter half-life, and other pharmacological characteristics [27, 28]. A few available long-term studies suggest that T3 use may still be safe if administered with the necessary care [29].

Examining the strength of evidence and sources of disagreement with other findings, a number of problems can affect RCTs, related to the sensitivity of the QoL instrument, patient selection, effect sizes, statistical power and sample size, and response heterogeneity.

As for the QoL instrument used, historic trials did not rely on validated and thyroid-specific QoL questionnaires, which only became available more recently [30]. As QoL instruments vary considerably in their ability to detect relevant treatment effects in hypothyroidism and exhibit vast differences in effect sizes for individual items, this is an important issue [30]. Unlike ThyPRO, the SF-36 did not achieve even moderate effect sizes for any of its items in a validation trial by Watt et al. [30]. Asking less relevant and nonspecific questions diminishes the overall discriminatory power of the instrument, even more so when p values are adjusted for multiple testing. Targeted thyroid-specific questions may substitute more successfully for a panel of more widely engineered questions [16]. Strong emphasis should therefore be placed on specificity and relevance of questions, use of validated instruments, documented effect sizes, and mandatory realistic pretrial power estimates.

Estimated trial size using contemporary QoL assessment would be 120 patients to detect a marked change in thyroid function during follow-up from hypothyroidism to euthyroidism, based on two-sided t-test statistics and assuming two balanced groups, a moderate effect size (total score 0.59 [30]), alpha error of 5%, beta error of 20% (power 80%), and attrition rate of 30%. The QoL difference between two treatment modalities, such as LT4 and LT3/LT4 combination, can be expected to be much lower than that for a hypothyroid and euthyroid group. This suggests that many historic trials were underpowered, and careful preselection of patients may be necessary to achieve sufficient power for such trials. Post hoc subgroup analysis of patients dissatisfied with LT4 monotherapy, which has been reported in some studies, is statistically controversial and cannot substitute for randomised controlled treatment allocation [31].

Another potential source of loss of statistical power in trials involving the combined administration of T3/T4 relates to widely varying T3 conversion rates among both disease entities and individual subjects [19, 32]. A minor fraction of the peripheral T3 is thyroid-derived, but this amount varies due to the action of a regulatory TSH-T3 shunt [33]. Deiodination capacity is low in patients lacking functional thyroid tissue, e.g., after thyroidectomy, and may decline even further after initiation of LT4 treatment [23, 32]. Effects of adding on T3 may therefore be expected to be more pronounced and beneficial to patients with low T3 concentrations than to those with a higher endogenous or LT4-derived T3 supply. Aggregating patients with various aetiologies and stages of hypothyroidism and different treatment requirements in a single trial may render the expected statistical power of a substitution trial unpredictable [11, 34, 35]. From a clinical point of view, at least patients suffering from autoimmune thyroiditis and thyroid carcinoma should be treated as separate entities.

Symptom rates prior to the trial may give an indication if any improvement could be expected during the trial, irrespective of treatment modality. Averaged scores are less informative in this respect, and averaging techniques also tend to obscure occasions with low occurrence [34]. For symptoms at trial end, the question is how much improvement has occurred. However, a healthy control group was lacking in virtually all combination trials. Neither dose adequacy for LT4 nor LT3/LT4 has been ascertained in the T3/T4 trials. It remains therefore unclear how much either treatment option was able to raise QoL outcomes, compared to a healthy population. Only a control group defining the target level is able to ascertain adequacy of treatment outcome, protecting against an ineffective trial where neither treatment was properly administered, a real possibility indicated by recent prospective studies [10]. The trial must include a control group to address this issue adequately, because within-group comparisons over time are subject to regression to the mean, autocorrelation, and other statistical fallacies.

Inconsistencies between individual trials were assessed in a meta-analysis, although not specifically reported [7]. However, collective analysis may not compensate for shortcomings at the level of several single trials as design errors may distort outcomes in different directions and gravely reduce the legitimacy of an overall conclusion. When evaluating the trustworthiness of the bulk of systematic reviews and meta-analyses, a group of prominent authors and editors came to a devastating judgment and stated that “only about 3% of them are both well done and clinically useful” [35].

The unstratified inclusion of a heterogeneous mix of patients in the RCTs with different aetiologies of the disease, baseline biochemistry, and residual symptoms raises a concern of collider stratification bias [34, 35]. Physiologically, the biochemical equilibria between FT3, FT4, and TSH have been shown to be markedly altered by LT4-treatment, depending on disease aetiology, presence of residual thyroid tissue, and LT4 dose [19, 25, 32]. Treatment responses to LT4 may vary widely between patients even when the recorded biochemical parameter values are similar [11, 36]. This may result in

undertreatment of a substantial fraction of patients, because, as reviewed elsewhere, TSH-guided treatment may be compromised in certain circumstances [37]. A prospective study in patients with thyroid carcinoma, measuring biochemical surrogate markers of thyroid-influenced organ function, concluded that LT4-doses needed to be high enough to mildly suppress TSH levels in order to achieve satisfactory treatment outcomes close to euthyroidism, whereas patients with “normal” TSH levels remained mildly hypothyroid [38].

The role of T3 can be briefly expanded in this context. While uncorrelated with TSH in a healthy population over the euthyroid range, mainly due to a physiological TSH-FT3 shunt acting as a balancing tool, FT3 becomes strongly correlated in LT4-treated athyreotic patients lacking this modulatory device [23, 25]. Patients without a thyroid gland suffer a substantial loss of T3 conversion after thyroidectomy, on average approx. 25% in the same patient, compared to the rate prior to surgery [11, 37]. Low FT3 levels, in turn, were associated with persistent complaints in these patients [11]. Theoretical models suggest that the seemingly minor thyroid-derived T3 component has a more important physiological role in stabilising the system than merely directly contributing to the T3 pool [25, 33]. The symptom-FT3-FT4-TSH correlation chain may therefore vary widely, even on occasion inverting correlations.

As for trial design, amalgamation bias (Simpson's paradox [39]) arises when including heterogeneous study groups of patients who have different disease aetiologies, settle at different homeostatic equilibria, and display heterogeneous responses to treatment. Limited data are presently available on intraclass correlations (ICCs) and components of variance although this approach is essential for the interpretation of cluster-based studies [34, 35, 39, 40]. In addition, thyroid hormones are known for their high degree of individuality, with a low ratio of the intraindividual to interindividual variation of approximately 0.5 (low individuality index) [41]. For a person, their perceived “comfort zone” of response, like the intraindividual reference range, is also narrower than for the entire group, as evidenced by substantial intraclass correlation (0.3-0.5) between FT3 and patient complaints during follow-up of patients with thyroid carcinoma [11].

Consequently, multiple measures obtained from each subject are nested within that subject. Classical (single-level) methods cause issues related to the disaggregation of within-person and between-person effects over time, flattening the secondary level, and destroying the interrelationships within it. This is even further exacerbated by the fact that these two levels of influence can operate simultaneously and even in opposite directions [36, 39, 40]. This situation requires a multileveled instead of averaged statistical approach [42]. Personalised treatment targets may therefore be more appropriate than wider range considerations, especially for TSH [36]. In Figure 1, we demonstrate potential impact of conversion efficiency and shifted dose response curves on complaint rates. In such a situation, an error is made when applying the group-level effect to the individuals within the groups, and the patient case-mix may vary averaged outcomes in trials

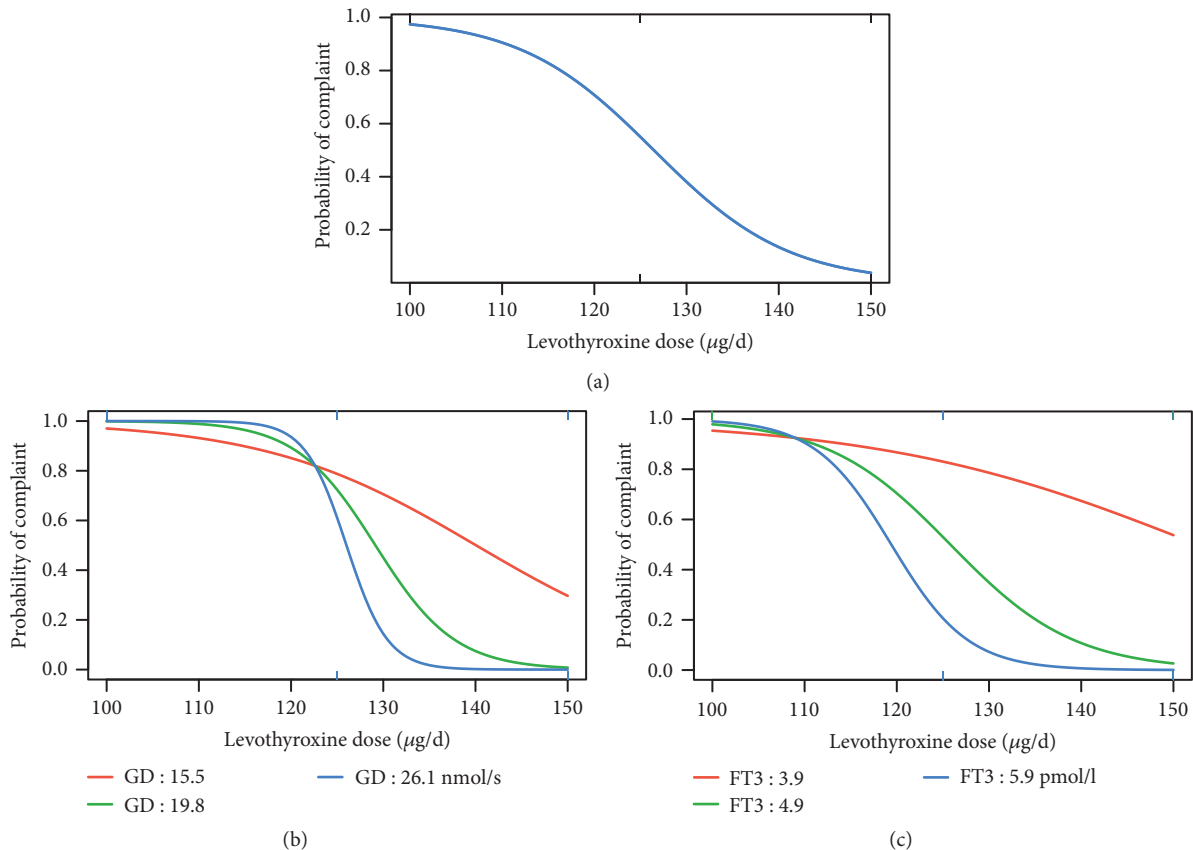


FIGURE 1: Demonstration of bias by data amalgamation in predicted probability of persisting hypothyroid complaints in response to LT4 dose changes depicted by a simulated trial. Group-level effect (a) is not predictive of individual responses with shifted response curves between patients displaying varying conversion rates (GD (b)), and FT3 concentrations (c). Variable patient individuality of response may impact the averaged outcome in a trial and should be accounted for in the analysis. For the purpose of this demonstration, complementing a study, which has been reported elsewhere [11], we resampled 60 follow-up visits from patients who reported relief of former complaints after LT4 dose increase within a range of 100 to 150  $\mu\text{g/d}$  to generate the structure of an extended generalised linear mixed-effects model [18]. GD refers to global deiodinase activity, a measure of T4 to T3 conversion similar to the molar T3-T4 ratio [11].

[40]. This may also affect statistical power of a trial (Figure 1, [18]), as discussed above. In sufficiently large samples, response heterogeneity to treatment can and should therefore be assessed, e.g., with latent class analysis [18, 39, 40, 42]. A statistical design and method must be chosen for trial analysis to guard against amalgamation bias, e.g., using multilevel models or cross-over designs.

#### 4. Treatment Efficiency and Adverse Effects

While well-suited to prove drug efficacy, RCTs are less able to predict the efficiency of a drug in clinical practice [43]. The latter deals with optimised drug use, as to which patients and conditions may benefit most from the intervention. This differs frequently from averaged results obtained under more artificial conditions in the RCT [34, 35, 40, 42]. While efficiency of T3 addition has not been rigorously studied, differential requirements of patients have been strongly suggested by homeostatic criteria. Mixed inclusion criteria and statistical outcome averaging can be expected to conceal differentiated treatment requirements as well as different risk profiles. Misclassification of true thyroid status by sole

TSH measurement and dichotomised analysis of a continuous variable is of concern. For instance, the association between hyperthyroidism and atrial fibrillation (AF) is well established [1], but 10-year AF risk correlated with FT4, not TSH concentrations in euthyroid subjects [44]. Few events were reported in this prospective study in thyroid hormone users (12/720), too few for separate analysis [44]. A high set point appears to be associated with a predetermined genetic risk for cardiac vulnerability [44]. We should note that a disintegration of the TSH-FT4 correlation toward the euthyroid range is unsurprising, given that set points are clustered [22, 45]. When assessing therapeutic risk profiles, this has to be taken into account as a confounder (collider stratification bias, Simpson's paradox), irrespective of study design including RCTs [34, 39, 40, 42]. FT3 concentrations, unlike TSH, correlated with heart rate within a more narrowly defined euthyroid TSH range [46]. A recent meta-analysis confirms that TSH measurements alone are unable to detect cardiac vulnerability with sufficient sensitivity and specificity [47]. In a large prospective study of euthyroid patients with AF undergoing catheter ablation, variations in concentration of thyroid hormones, but not TSH, were associated with



recurrence of arrhythmias [48]. Importantly, the risk association, while increasing linearly with FT4, was u-shaped for FT3; both high and low FT3 levels were associated with AF recurrence [48]. Possible pathophysiological explanations for this seemingly paradoxical behaviour involve the actions of nonclassical thyroid hormones, such as reverse-T3 and 3,5-T2 [49]. However, it is not known if raising low FT3 levels may improve the cardiac outcome. Both hyperthyroidism and hypothyroidism have also been shown to facilitate induced atrial fibrillation in rats [50]. Similarly, in a large unselected and predominantly euthyroid sample, we found a u-shaped relationship between the presence of anxiety/depression (Hospital Anxiety and Depression Scale) and serum FT3 levels, but no relationship of mood change with TSH and FT4 concentrations [51]. This association was masked when using standard linear logistic regression and only revealed by more advanced statistical methods (generalised linear-quadratic model) [51]. For bone health, the situation is also more complex in LT4-treated patients, with effects of low-TSH and low-FT3/TSH $\beta$ v on bone health opposing each other [52], although osteoporosis is a well-documented risk of hyperthyroidism [1].

Interestingly, patients treated with LT4 tend to display such a vulnerable constellation of relatively increased FT4 and diminished FT3 concentrations together with shifted equilibria for TSH [23]. This is uniquely different from endogenous hyperthyroidism, reminding us of the traditional distinction between hyperthyroidism and thyrotoxicosis.

From statistical principles, it follows that overlooking an unequally distributed confounding variable will bias the estimates of the effect (Simpson's paradox) [34, 35, 39, 53]. Optimised risk/benefit management requires more than a TSH value, namely, careful analysis of big data including all three thyroid hormones TSH, FT3, and FT4 and using modern statistical tools accounting for intraclass correlations. The wider diagnostic and potential therapeutic implications have been addressed elsewhere [36].

Importantly, while evidence-based medicine has traditionally based stronger emphasis on controlled outcomes achieved by RCTs, the treatment effect under actual "conditions of use" is the determinant of approval for a drug by the American Food and Drug Administration (FDA) [54]. RCTs may not actually meet this FDA criterium, as has been elegantly demonstrated by George et al. [55] from a statistical point of view. The very randomisation required to discriminate the treatment effect from other confounding influences may cause RCTs to poorly predict the actual treatment effect under conditions of intended use [55]. In RCTs, unlike under actual conditions of use where patients know about treatment, patients are uncertain whether they are receiving treatment. This introduces a bias, as patient expectations of improvement substantially impact outcomes and effect sizes of the treatment under actual conditions of use. Improvements in trial design (randomisation to randomisation probabilities, R2R) have been suggested to adjust for participant expectancy about receiving treatment and produce treatment estimates closer to those under actual conditions of use, while retaining core strengths of the RCT [55]. This may be particularly important to consider in QoL

trials where patient expectations may markedly influence the subjective treatment outcomes.

## 5. A Need for New Functional Disease Markers

While acting as a generally important outcome measure, QoL may not be an ideal marker for a functional disease. Objective markers of hypothyroidism, which truly reflect the functional thyroid status for various organs, would be required to reconcile RCT results in humans with contradictory experimental evidence from animal studies [20, 24, 56]. TSH can at best serve as a marker of pituitary activity [37]. Even in that role, it provides an adaptive and variable control mode rather than a fixed reference point [36]. Furthermore, peripheral tissues differ from regulatory central organs in important ways, among others in enzyme expression of deiodinases and cellular sensitivity to thyroid hormones [57]. In agreement with physiologically altered equilibria between TSH and FT3, as well as TSH and surrogate tissue markers for hypothyroidism, in LT4-treated patients, animals remained hypothyroid after LT4 administration despite achieving normalisation of their TSH concentrations [19–21, 24, 32, 38, 56]. While generally recognised as serum markers, superiority of surrogate markers such as serum concentrations of methylhistidine, ceruloplasmin, copper, and sex hormone binding globulin (SHBG) over TSH and thyroid hormone concentrations has so far been shown in athyreotic patients and other rare disorders of thyroid homeostasis [38, 58]. Genetic fingerprints and mitochondrial RNA species emerged as novel organ markers, although they are not yet ready for clinical use [59, 60]. Nonclassical thyroid hormones such as reverse-T3 and thyronamines, long believed to be bioinactive, are currently reevaluated [61]. 3,5-T2 correlated with atrial remodelling in the presence of low T3 [49]. Calculated parameters of thyroid homeostasis may serve as functional biomarkers, particularly in sub-clinical thyroid disease or low-T3 and high-T3 syndromes [62]. These markers may help distinguishing scenarios of different pathophysiology and superficially similar endocrine phenotype and guide stratification of subjects in prospective trials.

## 6. Conclusion

Several areas of recommended improvements in trial design have been identified (Table 1), cautioning against uncritical clinical interpretation of outcomes from previous RCTs on T3/T4 combination therapy. Until better evidence becomes available, the reliance on dated RCTs should be questioned, giving more weight to other forms of valid evidence that have accumulated in recent years and contradict key assumptions from these trials. In addition, strict regulations by organizations such as the FDA demand treatment effects to be estimated under actual conditions of use. To achieve this, inherent shortcomings of RCTs, such as the expectancy bias caused by uncertainty about the treatment in RCTs, as opposed to treatment certainty under conditions of use, must be addressed.

TABLE 1: Suggestions for improvement in trial design for randomised controlled T3/T4 combination studies.

Criterion	Issue of Previous Trials	Improvement
QoL instrument	lack of sensitivity and specificity of older methods	use of validated thyroid-specific methods
Detectable effect on QoL	small effect size	moderate effect size
Statistical power	very low	low
Sample size requirement	very large	large
Patient selection	selection bias due to inclusion of heterogenous patient groups by etiology and prognosis	inclusion of homogeneous diagnostic categories, use of stratified randomisation
Proportion of symptomatic patients	dilution of the true effect	randomized controlled designs for subgroups with persistent symptoms
Treatment-related improvement	healthy control group lacking	inclusion of a healthy control group
Dose adequacy	TSH targets may be misguided.	Treatment-related altered equilibria have to be considered.
Response heterogeneity	wide variation in the treatment response	physiologically based categorisation
Specific confounders	T4 to T3 conversion efficiency	identify conversion issues and apply strata
Statistical analysis	presence of unknown hierarchies and latent groups	latent class analysis
Statistical method	amalgamation bias (Simpson's paradox), disaggregation of within-group and between group effects over time	multilevel models, cross-over design
Patient expectancies	expectancy bias from treatment uncertainty in RCTs vs treatment certainty under actual conditions of intended drug use	randomization to randomization probabilities (R2R) adjusting for differences in patient expectancies
Tissue effects	not addressed by RCTs due to lack of differential markers for organ-specific effects	limited usefulness of surrogate markers, requirement for novel markers
Actions of non-classical thyroid hormones	not addressed	improvement of assay technology, evaluation as possible additional treatment targets
Safety profile	not addressed by RCTs	prospective acquisition and analysis of big data, especially from T3 users
Drug-related issues of LT4	generally reduced and variable T4 to T3 conversion rates	measuring conversion efficiency and targeted T3 addition
Drug-related issues of LT3	pharmacological properties, among others short half-life, high peak levels	slow-release preparations
Drug-related issues of natural desiccated thyroid extracts	popular choice among patients, but few studies	effective large-scale trials

These novel developments should encourage appropriately redesigned and methodologically improved trials to evaluate treatment outcomes of T3/T4 combination therapy, compared with standard LT4 treatment. A clinical focus should be on distinct disease strata, such as patients with autoimmune hypothyroidism or with thyroidectomy for malignant disease and a heterogenous treatment response in identifiable subgroups.

### Data Availability

No original data were used for this study, a systematic review. Previously reported original data were reused in part for Figure 1 and are available at [doi:10.1055/s-0043-125064,

doi:10.1371/journal.pone.0187232]. These prior studies and datasets are cited at relevant places within the text as [11, 22]. Data have been deposited according to the process of the respective journal together with the original publication.

### Conflicts of Interest

Johannes W. Dietrich received funding and personal fees by Sanofi-Henning, Bristol-Myers Squibb, Hexal AG, and Pfizer and is coowner of the intellectual property rights for the patent "System and Method for Deriving Parameters for Homeostatic Feedback Control of an Individual" (Singapore Institute for Clinical Sciences, Biomedical Sciences Institutes, Application no. 201208940-5 and WIPO no.

WO/2014/088516). All other authors declare that there are no conflicts of interest that could be perceived as prejudicing the impartiality of the research reported.

## References

- [1] B. Biondi and L. Wartofsky, "Treatment with thyroid hormone," *Endocrine Reviews*, vol. 35, no. 3, pp. 433–512, 2014.
- [2] L. Wartofsky, "Combination L-T3 and L-T4 therapy for hypothyroidism," *Current Opinion in Endocrinology, Diabetes and Obesity*, vol. 20, no. 5, pp. 460–466, 2013.
- [3] J. Jonklaas, A. C. Bianco, A. J. Bauer et al., "Guidelines for the treatment of hypothyroidism: prepared by the American thyroid association task force on thyroid hormone replacement," *Thyroid*, vol. 24, no. 12, pp. 1670–1751, 2014.
- [4] O. Okosieme, J. Gilbert, P. Abraham et al., "Management of primary hypothyroidism: Statement by the British Thyroid Association Executive Committee," *Clinical Endocrinology*, vol. 84, no. 6, pp. 799–808, 2016.
- [5] W. M. Wiersinga, "Therapy of endocrine disease: T4 + T3 combination therapy: is there a true effect?" *European Journal of Endocrinology*, vol. 177, no. 6, pp. R287–R296, 2017.
- [6] J. V. Hennessey and R. Espaillet, "Current evidence for the treatment of hypothyroidism with levothyroxine/levotriiodothyronine combination therapy versus levothyroxine monotherapy," *International Journal of Clinical Practice*, vol. 72, no. 2, p. e13062, 2018.
- [7] S. Grozinsky-Glasberg, A. Fraser, E. Nahshoni, A. Weizman, and L. Leibovici, "Thyroxine-triiodothyronine combination therapy versus thyroxine monotherapy for clinical hypothyroidism: Meta-analysis of randomized controlled trials," *The Journal of Clinical Endocrinology & Metabolism*, vol. 91, no. 7, pp. 2592–2599, 2006.
- [8] L. F. Michaelsson, B. B. Medici, J. L. la Cour et al., "Treating Hypothyroidism with Thyroxine/Triiodothyronine Combination Therapy in Denmark: Following Guidelines or Following Trends?" *European Thyroid Journal*, vol. 4, no. 3, pp. 174–180, 2015.
- [9] S. J. Peterson, A. R. Cappola, M. R. Castro et al., "An Online Survey of Hypothyroid Patients Demonstrates Prominent Dissatisfaction," *Thyroid*, vol. 28, no. 6, pp. 707–721, 2018.
- [10] K. H. Winther, P. Cramon, T. Watt et al., "Disease-specific as well as generic quality of life is widely impacted in autoimmune hypothyroidism and improves during the first six months of levothyroxine therapy," *PLoS ONE*, vol. 11, no. 6, Article ID e0156925, 2016.
- [11] R. Larisch, J. Midgley, J. Dietrich, and R. Hoermann, "Symptomatic Relief is Related to Serum Free Triiodothyronine Concentrations during Follow-up in Levothyroxine-Treated Patients with Differentiated Thyroid Cancer," *Experimental and Clinical Endocrinology & Diabetes*, 2018.
- [12] W. M. Wiersinga, L. Duntas, V. Fadeyev, B. Nygaard, and M. P. Vanderpump, "2012 ETA guidelines: the use of L-T4 + L-T3 in the treatment of hypothyroidism," *European Thyroid Journal*, vol. 1, no. 2, pp. 55–71, 2012.
- [13] J. Ott, R. Promberger, F. Kober et al., "Hashimoto's thyroiditis affects symptom load and quality of life unrelated to hypothyroidism: a prospective case-control study in women undergoing thyroidectomy for benign goiter," *Thyroid*, vol. 21, no. 2, pp. 161–167, 2011.
- [14] N. Kelderman-Bolk, T. J. Visser, J. P. Tijssen, and A. Berghout, "Quality of life in patients with primary hypothyroidism related to BMI," *European Journal of Endocrinology*, vol. 173, no. 4, pp. 507–515, 2015.
- [15] M. K. Applewhite, B. C. James, S. P. Kaplan et al., "Quality of life in thyroid cancer is similar to that of other cancers with worse survival," *World Journal of Surgery*, vol. 40, no. 3, pp. 551–561, 2016.
- [16] C. Hedman, T. Djärv, P. Strang, and C. I. Lundgren, "Effect of Thyroid-Related Symptoms on Long-Term Quality of Life in Patients with Differentiated Thyroid Carcinoma: A Population-Based Study in Sweden," *Thyroid*, vol. 27, no. 8, pp. 1034–1042, 2017.
- [17] J. Jonklaas, "Persistent hypothyroid symptoms in a patient with a normal thyroid stimulating hormone level," *Current Opinion in Endocrinology & Diabetes and Obesity*, vol. 24, no. 5, pp. 356–363, 2017.
- [18] M. P. Kain, B. M. Bolker, and M. W. McCoy, "A practical guide and power analysis for GLMMs: detecting among treatment variation in random effects," *PeerJ*, vol. 3, p. e1226, 2015.
- [19] R. Hoermann, J. E. M. Midgley, A. Giacobino et al., "Homeostatic equilibria between free thyroid hormones and pituitary thyrotropin are modulated by various influences including age, body mass index and treatment," *Clinical Endocrinology*, vol. 81, no. 6, pp. 907–915, 2014.
- [20] J. P. W. De Castro, T. L. Fonseca, C. B. Ueta et al., "Differences in hypothalamic type 2 deiodinase ubiquitination explain localized sensitivity to thyroxine," *The Journal of Clinical Investigation*, vol. 125, no. 2, pp. 769–781, 2015.
- [21] E. A. McAninch and A. C. Bianco, "New insights into the variable effectiveness of levothyroxine monotherapy for hypothyroidism," *The Lancet Diabetes & Endocrinology*, vol. 3, no. 10, pp. 756–758, 2015.
- [22] R. Hoermann, J. E. M. Midgley, R. Larisch, and J. W. C. Dietrich, "Advances in applied homeostatic modelling of the relationship between thyrotropin and free thyroxine," *PLoS ONE*, vol. 12, no. 11, Article ID e0187232, 2017.
- [23] R. Hoermann, J. E. M. Midgley, R. Larisch, and J. W. Dietrich, "Relational Stability in the Expression of Normality, Variation, and Control of Thyroid Function," *Frontiers in Endocrinology*, vol. 7, 2016.
- [24] H. F. Escobar-Morreale, F. E. Del Rey, M. J. Obregón, and G. M. De Escobar, "Only the combined treatment with thyroxine and triiodothyronine ensures euthyroidism in all tissues of the thyroidectomized rat," *Endocrinology*, vol. 137, no. 6, pp. 2490–2502, 1996.
- [25] R. Hoermann, J. E. M. Midgley, R. Larisch, and J. W. Dietrich, "Integration of Peripheral and Glandular Regulation of Triiodothyronine Production by Thyrotropin in Untreated and Thyroxine-Treated Subjects," *Hormone and Metabolic Research*, vol. 47, no. 9, pp. 674–680, 2015.
- [26] T. D. Hoang, C. H. Olsen, V. Q. Mai, P. W. Clyde, and M. K. M. Shakir, "Desiccated thyroid extract compared with levothyroxine in the treatment of hypothyroidism: A randomized, double-blind, crossover study," *The Journal of Clinical Endocrinology & Metabolism*, vol. 98, no. 5, pp. 1982–1990, 2013.
- [27] J. Jonklaas and K. D. Burman, "Daily administration of short-acting liothyronine is associated with significant triiodothyronine excursions and fails to alter thyroid-responsive parameters," *Thyroid*, vol. 26, no. 6, pp. 770–778, 2016.



- [28] J. Jonklaas, "Risks and safety of combination therapy for hypothyroidism," *Expert Review of Clinical Pharmacology*, vol. 9, no. 8, pp. 1057–1067, 2016.
- [29] G. P. Leese, E. Soto-Pedre, and L. A. Donnelly, "Liothyronine use in a 17 year observational population-based study - the tears study," *Clinical Endocrinology*, vol. 85, no. 6, pp. 918–925, 2016.
- [30] T. Watt, P. Cramon, L. Hegedüs et al., "The thyroid-related quality of life measure ThyPRO has good responsiveness and ability to detect relevant treatment effects," *The Journal of Clinical Endocrinology & Metabolism*, vol. 99, no. 10, pp. 3708–3717, 2014.
- [31] S. J. Pocock, J. J. V. McMurray, and T. J. Collier, "Statistical Controversies in Reporting of Clinical Trials Part 2 of a 4-Part Series on Statistics for Clinical Trials," *Journal of the American College of Cardiology*, vol. 66, no. 23, pp. 2648–2662, 2015.
- [32] J. E. Midgley, R. Larisch, J. W. Dietrich, and R. Hoermann, "Variation in the biochemical response to l-thyroxine therapy and relationship with peripheral thyroid hormone conversion efficiency," *Endocrine Connections*, vol. 4, no. 4, pp. 196–205, 2015.
- [33] J. Berberich, J. W. Dietrich, R. Hoermann, and M. A. Müller, "Mathematical Modeling of the Pituitary–Thyroid Feedback Loop: Role of a TSH-T3-Shunt and Sensitivity Analysis," *Frontiers in Endocrinology*, vol. 9, 2018.
- [34] D. Kent and R. Hayward, "When averages hide individual differences in clinical trials," *American Scientist*, vol. 95, no. 1, pp. 60–68, 2007.
- [35] M. H. Møller, J. P. Ioannidis, and M. Darmon, "Are systematic reviews and meta-analyses still useful research? We are not sure," *Intensive Care Medicine*, vol. 44, no. 4, pp. 518–520, 2018.
- [36] R. Hoermann, J. E. Midgley, R. Larisch, and J. W. Dietrich, "Recent Advances in Thyroid Hormone Regulation: Toward a New Paradigm for Optimal Diagnosis and Treatment," *Frontiers in Endocrinology*, vol. 8, 2017.
- [37] R. Hoermann, J. E. M. Midgley, R. Larisch, and J. W. Dietrich, "Is pituitary TSH an adequate measure of thyroid hormone-controlled homeostasis during thyroxine treatment?" *European Journal of Endocrinology / European Federation of Endocrine Societies*, vol. 168, no. 2, pp. 271–280, 2013.
- [38] M. Ito, A. Miyauchi, M. Hisakado et al., "Biochemical Markers Reflecting Thyroid Function in Athyreotic Patients on Levothyroxine Monotherapy," *Thyroid*, vol. 27, no. 4, pp. 484–490, 2017.
- [39] C. H. Wagner, "Simpson's Paradox in Real Life," *The American Statistician*, vol. 36, no. 1, p. 46, 1982.
- [40] P. J. Curran and D. J. Bauer, "The disaggregation of within-person and between-person effects in longitudinal models of change," *Annual Review of Psychology*, vol. 62, pp. 583–619, 2011.
- [41] S. Andersen, K. M. Pedersen, N. H. Bruun, and P. Laurberg, "Narrow individual variations in serum T<sub>4</sub> and T<sub>3</sub> in normal subjects: a clue to the understanding of subclinical thyroid disease," *The Journal of Clinical Endocrinology & Metabolism*, vol. 87, no. 3, pp. 1068–1072, 2002.
- [42] A. Gelman and J. Hill, *Data Analysis Using Regression and Multi-level/Hierarchical Models*, Cambridge University Press, New York, 2007, ISBN 978-0-521-68689-1.
- [43] A. G. Singal, P. D. Higgins, and A. K. Waljee, "A Primer on Effectiveness and Efficacy Trials," *Clinical and Translational Gastroenterology*, vol. 5, no. 1, pp. e45–e45, 2014.
- [44] L. Chaker, J. Heeringa, A. Dehghan et al., "Normal thyroid function and the risk of atrial fibrillation: The Rotterdam study," *The Journal of Clinical Endocrinology & Metabolism*, vol. 100, no. 10, pp. 3718–3724, 2015.
- [45] R. Hoermann, W. Eckl, C. Hoermann, and R. Larisch, "Complex relationship between free thyroxine and TSH in the regulation of thyroid function," *European Journal of Endocrinology*, vol. 162, no. 6, pp. 1123–1129, 2010.
- [46] G. L. Roef, Y. E. Taes, J.-M. Kaufman et al., "Thyroid hormone levels within reference range are associated with heart rate, cardiac structure, and function in middle-aged men and women," *Thyroid*, vol. 23, no. 8, pp. 947–954, 2013.
- [47] C. Baumgartner, B. R. Da Costa, T.-H. Collet et al., "Thyroid Function Within the Normal Range, Subclinical Hypothyroidism, and the Risk of Atrial Fibrillation," *Circulation*, vol. 136, no. 22, pp. 2100–2116, 2017.
- [48] S. Wei, W. Wang, N. Liu et al., "U-shaped association between serum free triiodothyronine and recurrence of atrial fibrillation after catheter ablation," *Journal of Interventional Cardiac Electrophysiology*, vol. 51, no. 3, pp. 263–270, 2018.
- [49] J. W. Dietrich, P. Müller, F. Schiedat et al., "Nonthyroidal Illness Syndrome in Cardiac Illness Involves Elevated Concentrations of 3,5-Diiodothyronine and Correlates with Atrial Remodeling," *European Thyroid Journal*, vol. 4, no. 2, pp. 129–137, 2015.
- [50] Y. Zhang, E. I. Dedkov, D. Teplitsky et al., "Both hypothyroidism and hyperthyroidism increase atrial fibrillation inducibility in rats," *Circulation: Arrhythmia and Electrophysiology*, vol. 6, no. 5, pp. 952–959, 2013.
- [51] R. Larisch, S. Schulte, and G. Hildenbrand, "The role of thyroid hormones in anxiety and depression," *Nuklearmedizin / Nuclear Medicine*, vol. 53, 2015, V162 (Abstract).
- [52] R. Baliram, R. Latif, M. Zaidi, and T. F. Davies, "Expanding the Role of Thyroid-Stimulating Hormone in Skeletal Physiology," *Frontiers in Endocrinology*, vol. 8, 2017.
- [53] S. Ameringer, R. C. Serlin, and S. Ward, "Simpson's paradox and experimental research," *Nursing Research*, vol. 58, no. 2, pp. 123–127, 2009.
- [54] U.S. Food and Drug Administration, *Code of federal regulations Title 21*, 2014, <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart314&showFR1&subpart=Node,21:5.0.1.1.4.2>.
- [55] "Supplemental Material for Randomization to Randomization Probability: Estimating Treatment Effects Under Actual Conditions of Use," *Psychological Methods*, vol. 23, pp. 337–350, 2018.
- [56] H. F. Escobar-Morreale, M. J. Obregón, F. Escobar Del Rey, and G. Morreale De Escobar, "Replacement therapy for hypothyroidism with thyroxine alone does not ensure euthyroidism in all tissues, as studied in thyroidectomized rats," *The Journal of Clinical Investigation*, vol. 96, no. 6, pp. 2828–2838, 1995.
- [57] B. Gereben, A. M. Zavacki, S. Ribich et al., "Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling," *Endocrine Reviews*, vol. 29, no. 7, pp. 898–938, 2008.
- [58] P. Beck-Peccoz, A. Lania, A. Beckers, K. Chatterjee, and J. Wemeau, "2013 European Thyroid Association Guidelines for the Diagnosis and Treatment of Thyrotropin-Secreting Pituitary Tumors," *European Thyroid Journal*, vol. 2, no. 2, pp. 76–82, 2013.
- [59] E. A. McAninch, S. Jo, N. Z. Preite et al., "Prevalent polymorphism in thyroid hormone-activating enzyme leaves a genetic fingerprint that underlies associated clinical syndromes," *The Journal of Clinical Endocrinology & Metabolism*, vol. 100, no. 3, pp. 920–933, 2015.



- [60] E. T. Massolt, L. Chaker, T. J. Visser et al., "Serum microRNA profiles in athyroid patients on and off levothyroxine therapy," *PLoS ONE*, vol. 13, no. 4, p. e0194259, 2018.
- [61] S. Piehl, C. S. Hoefig, T. S. Scanlan, and J. Köhrle, "Thyronamines - Past, present, and future," *Endocrine Reviews*, vol. 32, no. 1, pp. 64–80, 2011.
- [62] J. W. Dietrich, G. Landgrafe-Mende, E. Wiora et al., "Calculated Parameters of Thyroid Homeostasis: Emerging Tools for Differential Diagnosis and Clinical Research," *Frontiers in Endocrinology*, vol. 7, 2016.