1.0 TITLE PAGE

3014-201-002

A multicenter, randomized, double-blind, dose-conversion study to evaluate the safety and efficacy of hormone replacement therapy with Armour® Thyroid compared to synthetic T4 (levothyroxine) in previously hypothyroid participants, who are euthyroid on T4 replacement therapy

STATISTICAL ANALYSIS PLAN

Version 2: 25 October 2021

2.0 TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

3.0 LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ALP	alkaline phosphatase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
ATH	Armour Thyroid
CFB	change from baseline
CDF	cumulative distribution function
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
eCRF	electronic case report form
ECG	electrocardiogram, electrocardiographic
EOT	end of trial
ITT	intent to treat
MAR	missing at random
MI	multiple imputation
OC	observed cases
PCS	potentially clinically significant
QTc	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using the Bazett formula ($QTcB = QT/(RR)^{1/2}$)
QTcF	QT interval corrected for heart rate using the Fridericia formula ($QTcF = QT/(RR)^{1/3}$)
RM	rescue medication
SAE	serious adverse event
SAP	statistical analysis plan
SE	standard error
SI	*Le Système International d'Unités* (International System of Units)
SOC	system organ class
TBL	total bilirubin
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event

4.0 INTRODUCTION

This statistical analysis plan (SAP) provides technical and detailed elaboration of the statistical analyses of the efficacy and safety data that will be performed for Study 3014-201-002 as outlined and/or specified in a final protocol dated 28 May 2019 and the most recent amendment dated 16 Jan 2020. Specifications of tables, listings and figures are provided in a separate document.

4.1 STUDY DESIGN:

Study 3014-201-002 is a randomized, double-blinded, parallel-group, active-controlled, multicenter, dose-conversion study in previously hypothyroid participants, who are euthyroid on a stable dose of T4 replacement therapy, to evaluate the efficacy and safety of replacement therapy with matching doses of Armour Thyroid in comparison to synthetic T4.

During the coronavirus disease 2019 (COVID-19) pandemic, the mitigation strategy of increasing enrollment has been implemented to overcome the loss of information from the impact of COVID-19. Approximately 284 total participants (an increased enrollment from the original planned sample size of 220 participants specified in the protocol) who meet eligibility criteria will be randomly assigned to study intervention in a 1:1 ratio (approximately 142 participants per arm), stratified by age (< 65 years, ≥ 65 years).

The study will last approximately 33 to 51 weeks (with a planned minimum of 7 visits and a maximum of 15 possible visits if participants require additional dose titration) and will include a Screening Period (up to 3 weeks), double-blinded Titration Period (at least 18 weeks), and a double-blinded Stabilization Period (12 weeks).

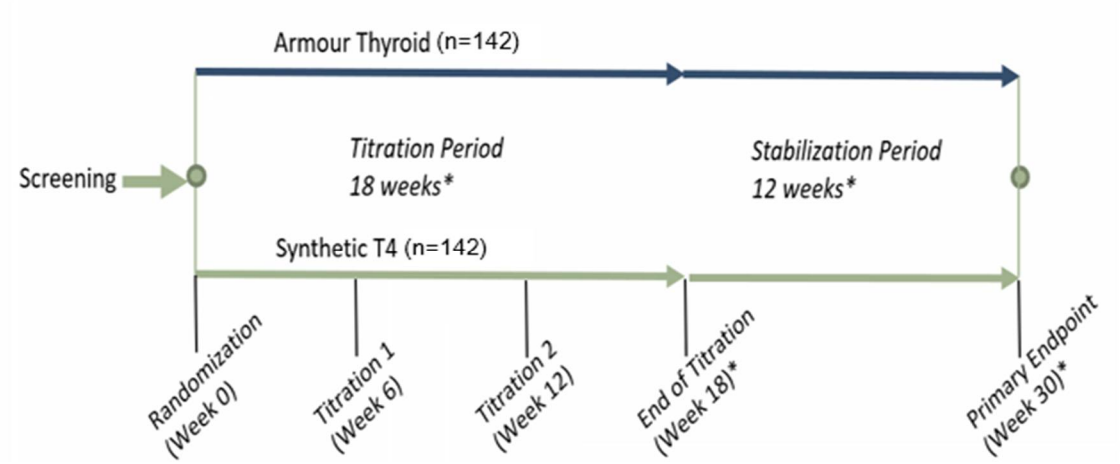
All participants entering the study will be controlled on a stable, FDA-approved, daily dose of synthetic T4 (i.e., euthyroid) for at least 3 months and will be randomized to either the same dose of synthetic T4 or a matching dose of Armour Thyroid during the first visit of the Titration Period (Visit 2). For each available dose of synthetic T4, up to 200 μg, a matching dose of Armour Thyroid will be employed according to a dose-conversion table based on the USP Drug Information 2000. Armour Thyroid and synthetic T4 will be administered as a single oral dose of 2 capsules (one of these capsules may be placebo), once-daily, on an empty stomach at least 30 to 60 minutes before the first meal. Serum TSH levels will be measured prior to study intervention administration during study visits.

During the Titration Period, participants will continue the same dose taken during the last visit or have their dose up- or down-titrated as needed per the Investigators' interpretation of their TSH levels (normal reference range 0.45-4.12 mIU/L, inclusive). For example, Investigators should up-titrate the dose if the TSH value is above 4.12 mIU/L and down-titrate the dose if the TSH value is below 0.45 mIU/L during the Titration Period. Other thyroid-related assessments and evaluation of thyroid clinical status will also be completed.

Investigators may only titrate participants' doses, as needed, until the start of the Stabilization Period. Participants will remain on a stable dose during the Stabilization Period until the end of the study.

Study schematic and schedule of evaluation are presented in Figure 4-1 and Table 4-1 respectively

Figure 4-1 3014-201-002 Study Schematic



-

* If a participant undergoes up to the maximum of 3 additional dose titrations starting at Week 18, the Titration Period may extend up to Week 36; therefore, the Stabilization Period may extend from Week 36 to Week 48 (remaining 12 weeks in duration). The Primary Endpoint will be measured at the end of the Titration Period and the end of the Stabilization Period.

During the Titration Period, study visits are split into parts A and B (i.e., Visit 3a, Visit 3b, Visit 4a, Visit 4b) to allow for any dose titrations:

- Visits 3a and 4a are mandatory for all participants. Study assessments, including bloodwork for total T3, TSH, and study intervention dispensing will be performed.
- Visits 3b and 4b are only required for participants needing a dose titration based on the data obtained from Visits 3a and 4a, respectively. A limited number of assessments and study intervention dispensing will be performed.

Visit 5 is designed as a transition visit between the Titration and Stabilization Periods and may consist of a subset of visits (starting with Visit 5a [Week 18] and extending to Visit 5g [Week 36], as needed). The last visit within Visit 5 serves as both the end of the Titration Period and the beginning of the Stabilization Period.

- Participants with TSH levels within normal reference range at Visit 5a may enter the Stabilization Period (i.e., their next visit will be Visit 6).
- If participants' TSH levels are out of normal reference range at Visit 5a, participants' doses can be up- or down-titrated per the Investigators' interpretation until their TSH levels are within the normal reference range by continuing the Titration Period

beyond Week 18. If the Titration Period is continued, the participants' doses can be up- or down-titrated approximately every 6 weeks until their TSH levels have normalized for up to a maximum of 3 additional titrations (including the titration at Visit 5b). Once their TSH levels are within the normal reference range, participants can enter the Stabilization Period. Participants whose TSH levels have not normalized after the maximum of 3 additional titrations will not enter the Stabilization Period.

- Participants who need dose titration beyond Week 18 will complete a study visit approximately every 6 weeks to determine if their TSH levels are within the normal reference range. At these visits, they will have the same assessments performed as those performed at Visit 5a including study intervention dispensing. These assessment visits will be labeled sequentially 5c (Week 24), 5e (Week 30), and 5g (Week 36), as needed.
- Participants who need dose titration beyond Week 18 will also complete a study visit for additional study intervention dispensing. These study intervention visits will be labeled sequentially as 5b (Week 19), 5d (Week 25), and 5f (Week 31), as needed. Visit 5d and 5f will have all the same assessments and activities performed as Visit 5b.

The Titration Period is at least 18 weeks in duration. Depending on whether a participant requires additional dose titration after Visit 5a, the Titration Period may end at Week 18, 24, 30, or 36 (up to a maximum of 3 additional titrations are allowed starting at Visit 5b).

The Stabilization Period is 12 weeks in duration. Depending on whether a participant requires additional dose titration after Visit 5a, the Stabilization Period may end at Week 30, 36, 42, or 48. If a participant undergoes the maximum of 3 additional dose titrations starting at Week 5b, the Stabilization Period would be from Week 36 to Week 48.

Visit 6 may be at Week 24, 30, 36, or 42 depending on the number of additional titrations needed after Visit 5a. Visit 7 may be at Week 30, 36, 42, or 48 depending on the number of additional titrations needed after Visit 5a.

Table 4-1 Schedule Activities

3014-201-002 Schedule of Activities											
	Screening	**Titration Period (Week 0 to Week 18[a])**						**Stabilization Period (Week 18[a] to Week 30[c])**			**Early Termination**
Visit Number	Visit 1	Visit 2	Visit 3a	Visit 3b[d]	Visit 4a	Visit 4b[d]	Visit 5a[e]	Visit 5b[d,e]	Visit 6[f]	Visit 7[g]	ET
Study Week	-3 to 0	0	6	7	12	13	18[d]	19	24	30	
Visit Window	---	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	
Informed consent	X										
Randomization		X									
Inclusion/exclusion	X	X									
Demographic data	X										
Medical history	X										
AEs	X	X	X	X	X	X	X	X	X	X	X
Clinical laboratory[b,h,n]	X	X	X		X		X		X	X	X
Vital signs[b,i]	X	X	X	X	X	X	X	X	X	X	X
ECG[b]	X	X	X		X		X		X	X	X
Physical examination[j]	X									X	X
BMI	X									X	X
Urine pregnancy test[b] (females of CBP)		X	X		X		X		X	X	X
Serum pregnancy test (females of CBP)[b,n]	X	X	X		X		X		X	X	X
Serum TSH, T4, T3, FT3, FT4 (Thyroid-related assessments)[b,n]	X	X	X		X		X		X	X	X
Thyroglobulin and anti-thyroglobulin[n,o]	X										
PK blood samples[k, n]		X								X	
Thyroid symptom questionnaire (ThyPRO Questionnaire)	X	X	X		X		X		X	X	
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X

	Screening	Titration Period (Week 0 to Week 18[a])						Stabilization Period (Week 18[a] to Week 30[c])			Early Termination
Visit Number	Visit 1	Visit 2	Visit 3a	Visit 3b[d]	Visit 4a	Visit 4b[d]	Visit 5a[e]	Visit 5b[d,e]	Visit 6[f]	Visit 7[g]	ET
Study Week	-3 to 0	0	6	7	12	13	18[d]	19	24	30	
Visit Window	---	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	
Evaluation of thyroid clinical status[l]	X	X	X		X		X		X	X	X
Study intervention dispensing		X	X	X	X	X	X	X	X		
Study intervention compliance		X[m]	X	X	X	X	X	X	X	X	X
Study intervention collection			X	X	X	X	X	X	X	X	X
Potential for dose adjustment				X		X		X			

AE = adverse event; BMI = body mass index; BP = blood pressure; CBP = childbearing potential; ECG = electrocardiogram; ET = early termination; FT3 = free T3; FT4 = free T4; IRB/IEC = institutional review board/independent ethics committee; PK = pharmacokinetic; T3 = liothyronine; T4 = levothyroxine; TSH = thyroid-stimulating hormone

[a] The Titration Period is at least 18 weeks in duration. Depending on whether a participant requires additional dose titration after Visit 5a, the Titration Period may end at Week 18, 24, 30, or 36 (up to a maximum of 3 additional titrations are allowed starting at Visit 5b).

[b] These assessments will be pre-dose.

[c] The Stabilization Period is 12 weeks in duration. Depending on whether a participant requires additional dose titration after Visit 5a, the Stabilization Period may end at Week 30, 36, 42, or 48. If a participant undergoes the maximum of 3 additional dose titrations starting at Week 5b, the Stabilization Period would be from Week 36 to Week 48.

[d] This study intervention dispensing visit is only applicable to participants needing a dose titration.

[e] Visit 5 is designed as a transition visit, which serves as both the end of the Titration Period and the beginning of the Stabilization Period. At Visit 5a (Week 18), if a participant's TSH level is not within normal reference range, the participant may have his/her dose up- or down-titrated per the Investigator's interpretation of his/her TSH level by continuing the Titration Period beyond Week 18 for additional visits, as needed. The additional study intervention dispensing visits will be labeled sequentially (Visit 5b, 5d, and 5f, as needed). Visit 5d and 5f will have all the same assessments and activities performed as Visit 5b. Additional assessment visits will be labeled sequentially (Visit 5c, 5e, 5g, as needed), and will have all the same assessments and activities performed as Visit 5a.

[f] Visit 6 may be at Week 24, 30, 36, or 42 depending on the number of additional titrations needed after Visit 5a.

[g] Visit 7 may be at Week 30, 36, 42, or 48 depending on the number of additional titrations needed after Visit 5a.

[h] Hematology, chemistry, urinalysis.

[i] BP, pulse rate; respiratory rate, and oral temperature

[j] Includes height and weight measurements (height at Screening Visit only)

[k] PK samples will be collected pre-dose and 2 hour (± 30 minutes) post-dose at Visits 2 and 7. Additionally, the date and time of dose on the day of PK sampling will be recorded.

[l] This includes evaluation of hypothyroid symptoms, which will be documented in the eCRF.

[m] Compliance assessment at Visit 2 relates to the participant having taken his/her own synthetic T4 between Visit 1 and Visit 2.

[n] Participants will be required to undergo an overnight (minimum 5 hours) fast for all visits where laboratory assessments will be performed. Participants should also abstain from taking biotin 72 hours prior to clinical laboratory testing.

[o] Thyroglobulin and anti-thyroglobulin testing only applicable for participants who have undergone total thyroidectomy for treatment of thyroid cancer.

5.0 OBJECTIVES

The study objective is to evaluate the efficacy and safety of hormone replacement therapy with matching doses of Armour Thyroid (ATH) in comparison to synthetic T4 in previously hypothyroid participants who are euthyroid on synthetic T4 replacement therapy (25-200 μg T4 daily).

In addition, this study will validate the dose-conversion table by evaluating a safe and effective dose conversion from synthetic T4 to Armour Thyroid in participants who are euthyroid on a stable dose of synthetic T4.

6.0 POPULATIONS FOR ANALYSES

6.1 INTENT-TO-TREAT POPULATION

The Intent-to-Treat (ITT) Population will consist of all randomized participants who took at least 1 dose of study intervention.

6.2 LOW BASELINE TOTAL T3 SYMPTOMATIC SUBSET OF ITT POPULATION (ITT-T3)

The low baseline total T3 symptomatic Population (abbreviated as ITT-T3 Population) will consist of all participants in the ITT Population who had a baseline total T3 < 87 ng/dL and a ThyPRO hypothyroidism symptom scale > 0.

6.3 SAFETY POPULATION

The Safety Population will consist of all participants who took at least 1 dose of study intervention.

6.4 SAFETY-T3 POPULATION

The Safety-T3 Population will consist of all participants who took at least 1 dose of study intervention and who had a baseline total T3 < 87 ng/dL and a ThyPRO hypothyroidism symptom scale > 0.

7.0 PARTICIPANT DISPOSITION

The number and percentage of participants in the four analysis populations (ITT and ITT-T3 Populations, Safety and Safety-T3 Populations) will be summarized by treatment group and study center; the number of participants screened will be summarized overall only by study center.

Screen-failure participants (i.e., participants who consented to participate in the clinical study but were not randomized) and the associated reasons for failure to randomize will be tabulated for all screened participants.

The number and percentage of participants who complete the double blind treatment period and of participants who prematurely discontinue during the same period will be presented for each treatment group and pooled across treatment groups for the ITT and ITT-T3 Populations. The reasons for premature discontinuation from the double blind treatment period as recorded on the termination pages of the electronic case report forms will be summarized (number and percentage) by treatment group for the ITT and ITT-T3 Populations. All participants who prematurely discontinue during the double-blind period will be listed by discontinuation reason for the ITT and ITT-T3 Populations.

8.0 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographic parameters (e.g., age, race, ethnicity, sex) and other baseline characteristics (weight, BMI) will be summarized in total and by treatment group for the four analysis populations (ITT and ITT-T3 Populations, Safety and Safety-T3 Populations).

Abnormalities in participants' medical and surgical histories will be coded using the *Medical Dictionary for Regulatory Activities,* version 21.1 or newer. The number and percentage of participants with abnormalities in medical and surgical histories in each system organ class and preferred term will be summarized by treatment group for the ITT and ITT-T3 Populations.

Prior medication is defined as any medication started before the date of the first dose of double-blind study treatment. *Concomitant medication* is defined as any medication taken on or after the date of the first dose of double-blind study treatment. Any prior medications started more than 30 days before the date of the first dose of double-blind study treatment and any concomitant medications started after the date of the last dose of double-blind study treatment will not be presented in the summary tables, but will be included in the participant data listings.

Both prior and concomitant medication use will be summarized by the number and proportion of participants in each treatment group for the ITT and ITT-T3 Populations. The *WHO Drug Dictionary*, Version B2 March 2017 or newer, will be used to classify prior and concomitant medications by WHO Drug Anatomical/Therapeutic/Chemical category and drug preferred name. Multiple medications used by a participant will only be counted once for the coded drug preferred name or therapeutic class.

9.0 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

9.1 EXTENT OF EXPOSURE

Exposure to the study treatment for the Safety and Safety-T3 Populations during the double-blind treatment period will be summarized for treatment duration, calculated as the number of days from the date of the first dose of double-blind investigational product taken to the date of the last dose taken, inclusive. Descriptive statistics (number of participants, mean, SD, minimum, median, Q1, Q3 and maximum) will be presented by treatment group.

Mean daily dose (by period and overall mean) for the study treatment will be summarized using descriptive statistics (number of patients, mean, SD, median, minimum, Q1, Q3 and maximum). Overall modal daily dose and final daily dose will be summarized by the number and percentage of patients with a pre-specified dose.

Patient-years, defined as exposure to the study treatment in years, will be summarized by treatment for the Safety and Safety-T3 Populations.

9.2 MEASUREMENT OF TREATMENT COMPLIANCE

Dosing compliance for a specified period is defined as the total number of dose-numbers (daily single intervention based on dose level with two capsules, see Table 9-1) actually taken by a patient during that period divided by the number of dose-numbers that were expected to be taken during the same period multiplied by 100 by treatment group. The total number of dose-numbers actually taken during a specific period is calculated as the number of days in that period multiplied by the number of dose-numbers taken each day during that period. The number of dose-numbers expected to be taken for a specific treatment period will be calculated by multiplying the number of days in that period by the number of dose-numbers to be taken per day.

Descriptive statistics for study drug compliance will be presented by treatment group by period and for the whole double-blind treatment period, for the Safety and Safety-T3 Populations.

10.0 EFFICACY ANALYSES

The efficacy analyses will be performed on the ITT Population or ITT-T3 Population, depending on the endpoints.

10.1 PRIMARY EFFICACY ENDPOINT AND ANALYSES

The primary efficacy endpoint is the achievement of TSH values within the normal reference range of 0.45 to 4.12 mIU/L, inclusive, at both the end of the Titration Period and the end of the Stabilization Period (sustained TSH responder). The primary estimand is the difference in the percentage of sustained TSH responders between the Armour Thyroid and synthetic T4 study intervention groups in the ITT Population. The number and percentage of sustained TSH responders will be summarized by treatment group in the ITT Population. Participants with a missing TSH value at either the end of the Titration Period or at the end of the Stabilization Period will be handled based on **N**on-**R**esponder **I**mputation incorporating multiple imputation to handle missing data due to **C**ovid-19 (NRI-C), as outlined later in this section.

The non-inferiority (NI) hypothesis test for the primary efficacy endpoint, based on the ITT Population, will be performed at a 1-sided 2.5% level of significance (equivalent to a two-sided 5% level of significance). The null hypothesis is that Armour Thyroid is inferior to synthetic T4, which will be tested against the alternative hypothesis that Armour Thyroid is non-inferior to synthetic T4, as follows:

$$H_0: p_A - p_T \leq -\Delta \quad verus \; H_a: p_A - p_T > -\Delta$$

Where: p_A and p_T are the proportions of TSH responders in the Armour Thyroid and synthetic T4 study intervention groups, respectively; [redacted]

The analysis of the primary efficacy endpoint will be performed on the ITT Population using stratum-adjusted Mantel-Haenszel by age group (age < 65 years or ≥ 65 years) as described in (Agresti 2013, p. 231).

The stratified proportion difference is

$$\hat{d}_{MH} = \left(\sum_h \hat{d}_h w_h\right) / \left(\sum_h w_h\right)$$

where $\hat{d}_h$ is the proportion difference in stratum h and $w_h = n_{h1\cdot} n_{h2\cdot} / n_h$.

The proportion difference in stratum h is computed as

$\hat{d}_h = \hat{p}_{h1} - \hat{p}_{h2} = (n_{h11}/n_{h1\cdot}) - (n_{h21}/n_{h2\cdot})$

where $\hat{p}_{h1}$ is the proportion of TSH responder for ATH and $\hat{p}_{h2}$ is the proportion of TSH responder for synthetic T4 at stratum h.

The variance of $\hat{d}_{MH}$ as

$$\hat{\sigma}^2(\hat{d}_{MH}) = \left(\hat{d}_{MH} \sum_h P_h + \sum_h Q_h\right) / \left(\sum_h w_h\right)^2$$

where

$$P_h = \left(n_{h1\cdot} n_{h21} - n_{h2\cdot} n_{h11} + n_{h1\cdot} n_{h2\cdot} (n_{h2\cdot} - n_{h1\cdot})/2\right) / n_h^2$$

$$Q_h = (n_{h11}(n_{h2\cdot} - n_{h21}) + n_{h21}(n_{h1\cdot} - n_{h11}))/2n_h$$

The $100(1-\alpha)\%$ confidence limits for the stratified proportion difference are

$$\hat{d}_{MH} \pm \left(z_{\alpha/2} \times \hat{\sigma}(\hat{d}_{MH})\right).$$

[redacted]

[redacted]

The **N**on-**R**esponder **I**mputation incorporating multiple imputation to handle missing data due to **C**ovid-19 (NRI-C) will be performed for participants who have missing TSH values either at the end of the Titration Period or at the end of the Stabilization Period. The NRI-C will categorize any participant who does not have an evaluation at the end of Titration Period or at the end of Stabilization Period as a non-responder. The only exception is that missing data solely due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic will be handled by Multiple Imputation (MI) under the Missing at Random (MAR) mechanism.

The analysis using NRI-C will be carried out in the following steps.

Step 1. Imputation

Step 1.1 Multiple imputation for TSH values: 100 imputed datasets of TSH values at the end of Titration Period and the end of Stabilization Period will be generated using SAS PROC MI FCS REG. The random seed 12345 will be used. The following covariates will be included in the imputation model:

1) Randomized treatment group
2) Stratification group (age < 65 years or ≥ 65 years)
3) Baseline TSH
4) TSH at the end of Titration Period for the imputation of missing TSH at the end of Stabilization Period

Step 1.2. Imputing Responder Status: In the imputed datasets from Step 1.1, the responder status (yes/no) will be derived from the TSH values at the end of Titration

Period and the end of Stabilization Period according to the normal reference range (0.45-4.12 mIU/L, inclusive) based on the imputed values with the following exception:

The participants who have missing TSH values either at the end of the Titration Period or at the end of the Stabilization Period for reasons other than solely due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic, the non-responder status will be assigned to these participants in the imputed datasets regardless of the imputed values.

Step 2. Analysis: Each of the 100 imputed datasets from Step 1.2 will be analyzed separately using the above mentioned stratified Mantel-Haenszel test.

Step 3. Pooling: The point estimate and 95% confidence interval of the stratified proportion difference from the 100 imputed datasets are combined into one overall result using **PROC MIANALYZE.**

Sensitivity Analyses for the Primary Efficacy Endpoint

Participants who are missing TSH values either at the end of the Titration Period or at the end of the Stabilization Period will be assessed by the following sensitivity analyses. Additional sensitivity analyses will be performed if useful and appropriate.

Non-Responder Imputation

The non-responder imputation sensitivity analysis will be performed to investigate the impact of missing TSH values**.** If a participant has missing TSH values at either the end of the Titration Period or the end of the Stabilization Period, then this participant will be imputed as a TSH non-responder**.**

Responder Imputation

The responder imputation sensitivity analysis will be performed to investigate the impact of missing TSH values**.** If a participant has missing TSH values at either the end of the Titration Period or the end of the Stabilization Period, then this participant will be imputed as a Sustained TSH responder**.** If a participant has the TSH value out of the normal reference range at the end of titration or stabilization, the participant will be considered as non-responder by the definition of sustained TSH response regardless of missing data.

Complete Case

Participants who have not completed the study or who are missing TSH values at either the end of the Titration Period or the end of the Stabilization Period will be excluded from the analysis. That is, the non-inferiority between Armour Thyroid and synthetic T4 will be established in the complete-case.

NRI incorporating Multiple Imputation (NRI-MI)

The **N**on-**R**esponder **I**mputation incorporating **M**ultiple **I**mputation (NRI-MI) will be carried out for all the missing TSH values in the steps similar to the ones described above for NRI-C, with the exception of Step 1.2.

> **Step 1.2. Imputing Responder Status:** In the imputed datasets from Step 1.1, the responder status (yes/no) will be derived from the TSH values at the end of Titration Period and the end of Stabilization Period according to the normal reference range (0.45-4.12 mIU/L, inclusive) based on the imputed values with the following exception: For participants who prematurely discontinue from the study due to lack of efficacy, adverse events related to study drug, or death, the non-responder status will be assigned to these participants in the imputed datasets regardless of the imputed values.

Multiple Imputation

The **M**ultiple **I**mputation (MI) analysis will be carried out for all the missing TSH values regardless of reasons for missingness in the steps similar to the ones described above for NRI-C, with the exception of Step 1.2.

> **Step 1.2. Imputing Responder Status:** In the imputed datasets from Step 1.1, the responder status (yes/no) will be derived from the TSH values at the end of Titration Period and the end of Stabilization Period according to the normal reference range based on the imputed values.

Tipping Point

Tipping-point analysis will be performed to investigate the impact of missing TSH values on the study results, that is to identify difference of minimum numbers of Sustained TSH responders between treatment groups in the missing cohort which results in conclusion change (Yan, Lee and Li, 2007).

Stratified Newcombe Confidence Interval

For the sensitivity analyses of Non-Responder Imputation, Responder Imputation, and Complete Case, the 2-sided 95% confidence interval for stratified proportion difference will also be performed in the ITT Population using stratified Newcombe confidence interval, which is based on a stratified method with Cochran-Mantel-Haenszel weights (Yan and Su 2012).

10.2 SECONDARY EFFICACY ENDPOINTS

The Achievement of Titration TSH Response

The achievement of titration TSH Response is defined as TSH values that are within the normal reference range of 0.45 to 4.12 mIU/L, inclusive, at the end of the Titration Period (Titration TSH Responder). The secondary estimand is the difference in the percentage of Titration TSH Responders between the Armour Thyroid and synthetic T4 study intervention groups in the ITT population. The number and percentage of titration TSH responders will be summarized by treatment group in the ITT Population along with the 2-sided 95% CI for the stratified proportion difference. The non-inferiority (NI) hypothesis test for the secondary efficacy endpoint, based on the ITT Population, will be performed at a 1-sided 2.5% level of significance (equivalent to a two-sided 5% level of significance). The analysis of the secondary efficacy endpoint will be performed on the ITT Population using stratum-adjusted Mantel-Haenszel by age group (age < 65 years or ≥ 65 years) as described in the Section 10.1 (Agresti 2013, p. 231).

The **N**on-**R**esponder **I**mputation incorporating multiple imputation to handle missing data due to **C**ovid-19 (NRI-C) will be performed for participants who have missing TSH values at the end of the Titration Period similar to the primary efficacy analysis outlined in Section 10.1.

Sensitivity Analyses for the Secondary Efficacy Endpoint

Participants who are missing TSH values at the end of the Titration Period will be assessed by the following sensitivity analyses. Additional sensitivity analyses will be performed if useful and appropriate.

Non-Responder Imputation

The non-responder imputation sensitivity analysis will be performed to investigate the impact of missing TSH values**.** If a participant has missing TSH value at the end of the Titration Period, then this participant will be imputed as a titration TSH non-responder**.**

Responder Imputation

The responder imputation sensitivity analysis will be performed to investigate the impact of missing TSH values**.** If a participant has missing TSH value at the end of the Titration Period, then this participant will be imputed as a titration TSH responder**.**

Complete Case

Participants who have not completed the Titration Period or who have missing TSH values at the end of the Titration Period will be excluded from the analysis.

NRI incorporating Multiple Imputation (NRI-MI)

The **N**on-**R**esponder **I**mputation incorporating **M**ultiple **I**mputation (NRI-MI) will be carried out for all the missing TSH values in the steps similar to the ones described above for NRI-C, with the exception of Step 1.2.

Step 1.2. Imputing Responder Status: In the imputed datasets from Step 1.1, the responder status (yes/no) will be derived from the TSH values at the end of Titration Period according to the normal reference range (0.45-4.12 mIU/L, inclusive).

For participants who prematurely discontinues from the study drug during the Titration Period due to lack of efficacy, adverse events related to study drug, or death, the non-responder status will be assigned to these participants in the imputed datasets regardless of the imputed values.

Multiple Imputation

The **M**ultiple **I**mputation (MI) analysis will be carried out for all the missing TSH values regardless of reasons for missingness in the steps similar to the ones described above for NRI-C, with the exception of Step 1.2.

Step 1.2. Imputing Responder Status: In the imputed datasets from Step 1.1, the responder status (yes/no) will be derived from the TSH values at the end of Titration Period according to the normal reference range.

Tipping Point

Tipping-point analysis will be performed to investigate the impact of missing TSH values on the study results, that is to identify difference of minimum numbers of titration TSH responders between treatment groups in missing cohort which results in conclusion change (Yan, Lee and Li, 2007).

10.3 MULTIPLE COMPARISONS PROCEDURE

The overall type I error rate will be controlled at a two-sided 0.05 level for the primary and secondary efficacy endpoints using the fixed sequential procedure in the following order:

- Step 1: Testing for non-inferiority of proportions of sustained TSH responders in the ITT Population.
- Step 2: Testing for non-inferiority of proportions of titration TSH responders in the ITT Population.

10.4 ADDITIONAL AND EXPLORATORY EFFICACY ENDPOINTS

[redacted]

11.0 SAFETY ANALYSES

The safety analysis will be performed on the Safety and Safety-T3 Populations. The safety parameters will include AEs, clinical laboratory measurements including fasting T3 and T4 (total T3, total T4, free T3 and free T4), vital sign values, ECG parameters, and physical examination findings. For each safety parameter of the clinical laboratory, vital sign, and ECG parameters, the last non-missing safety assessment prior to or on the first dose of study intervention will be used as the baseline for all analyses of that safety parameter. Continuous variables will be summarized by number of patients and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of patients.

11.1 ADVERSE EVENTS

AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 19.0 or newer.

An AE will be considered a TEAE if:

- The AE began on or after the date of the first dose of study intervention; or
- The AE was present before the date of the first dose of study intervention, but increased in severity or became serious on or after the date of the first dose of study intervention

An AE that occurs more than 35 days after the last dose of study intervention will not be counted as a TEAE.

An AE will be considered a TESAE if it is a TEAE that additionally meets any SAE criteria.

The number and percentage of participants reporting TEAEs in each study intervention group will be tabulated by system organ class and preferred term and by system organ class, preferred term, and severity.

The number and percentage of participants reporting treatment related TEAEs in each study intervention group will be tabulated by system organ class and preferred term. If more than 1 AE is coded to the same preferred term for the same participant, the participant will be counted only once for that preferred term using the most severe and most related occurrence for the summarizations by severity and by relationship to study intervention.

Summary tables will be provided for participants with SAEs and participants with AEs leading to discontinuation if 5 or more participants reported such events. Listings of all AEs, SAEs, and AEs leading to discontinuation by participant will be presented.

11.2 CLINICAL LABORATORY PARAMETERS

Descriptive statistics for clinical laboratory values (in SI units) at baseline (screening) and changes from baseline at each assessment will be presented by study intervention for each clinical laboratory assessment.

• **Laboratory Assessments**	• **Parameters**			
Thyroid-related	• Serum TSH, T4, T3, FT4, and FT3			
Thyroglobulin and anti-thyroglobulin	• Thyroglobulin and anti-thyroglobulin testing only applicable for participants who have undergone total thyroidectomy for treatment of thyroid cancer			
Hematology	Platelet count RBC count Hemoglobin Hematocrit	RBC indices: MCV MCH MCHC %Reticulocytes	WBC count with differential (absolute): Neutrophils Lymphocytes Monocytes Eosinophils Basophils	
Clinical Chemistry	BUN Creatinine Glucose HbA1c	Potassium Sodium Calcium	AST ALT Alkaline phosphatase	Total, direct and indirect bilirubin Total protein Cholesterol, chloride, albumin
Routine Urinalysis[a]	• Specific gravity • pH, glucose, protein, blood, ketones • Microscopic examination (if blood or protein is abnormal)			
Pregnancy Test[a]	• Urine and serum test (females of CBP)			

ALT = alanine aminotransferase; AST = aspartate aminotransferase; BUN = blood urea nitrogen; CBP = childbearing potential; FT3 = free T3; FT4 = free T4; MCH = mean corpuscular hemoglobin; MCHC = mean corpuscular hemoglobin concentration; MCV = mean corpuscular volume; RBC = red blood cell; T3 = liothyronine; T4 = levothyroxine; TSH = thyroid-stimulating hormone; WBC = white blood cell

a Serum pregnancy tests will be standard at the Screening Visit (Visit 1) and Visit 7; local urine testing will be standard for all other visits, as specified in the schedule of activities, unless serum testing is required by local regulation or IRB/IEC.

Clinical laboratory test values will be considered potentially clinically significant (PCS) if they meet either the lower-limit or higher-limit PCS criteria listed in Table 11-1. The number and percentage of participants who have PCS post-baseline clinical laboratory values will be tabulated by study intervention at each assessment. The percentages will be calculated relative to the number of participants who have available non-PCS baseline values and at least 1 post-baseline assessment. The numerator will be the total number of participants with at least 1 PCS post-baseline value. A supportive listing of participants with PCS post-baseline values will be provided for the Safety and Safety-T3 Populations.

Table 11-1 Criteria for Potentially Clinically Significant Laboratory Tests

Parameter	***SI Unit***	***Lower Limit***	***Higher Limit***
CHEMISTRY			
Albumin	g/L	< 0.9 * LLN	> 1.1 * ULN
Alanine Aminotransferase (ALT)	U/L	—	≥ 3 * ULN
Alkaline Phosphatase	U/L	—	≥ 3 * ULN
Aspartate Aminotransferase (AST)	U/L	—	≥ 3 * ULN
Calcium	mmol/L	< 0.9 * LLN	> 1.1 * ULN
Chloride	mmol/L	< 0.9 * LLN	> 1.1 * ULN
Cholesterol	mmol/L	—	> 1.6 * ULN
Creatinine	µmol/L	—	> 1.3 * ULN
Potassium	mmol/L	< 0.9 * LLN	> 2.0 * ULN
Glucose, Fasting	mmol/L	< 0.9 * LLN	> 1.4 * ULN
Sodium	mmol/L	< 0.9 * LLN	> 1.1 * ULN
Total Bilirubin	µmol/L	—	> 1.5 * ULN
Total Protein	g/L	< 0.9 * LLN	> 1.1 * ULN
Triglycerides, Fasting	mmol/L	—	≥ 3 * ULN
Urea (BUN)	mmol/L	—	> 1.2 * ULN
Magnesium	mmol/L	< 0.9 * LLN	> 1.1 * ULN
Bicarbonate	mmol/L	< 0.9 * LLN	> 1.1 * ULN
Phosphate	mmol/L	< 0.9 * LLN	> 1.1 * ULN
Uric Acid	µmol/L	< 0.9 * LLN	> 1.1 * ULN
HEMATOLOGY			
Basophils Absolute Cell Count	10^9/L	—	> 3 * ULN
Eosinophils Absolute Cell Count	10^9/L	—	> 3 * ULN
Hematocrit	Ratio	< 0.9 * LLN	> 1.1 * ULN
Hemoglobin	g/L	< 0.9 * LLN	> 1.1 * ULN
Lymphocytes Absolute Cell Count	10^9/L	< 0.8 * LLN	> 1.5 * ULN
MCH	PG	—	> 3 * ULN
MCHC	G/L	—	> 3 * ULN
MCV	fL	< 0.9 * LLN	> 1.1 * ULN
Monocytes Absolute Cell Count	10^9/L	—	> 3 * ULN
Neutrophils Absolute Cell Count	10^9/L	< 0.8 * LLN	> 1.5 * ULN
Platelet Count	10^9/L	< 0.5 * LLN	> 1.5 * ULN
Red Blood Cell Count (Erythrocyte Count)	10^{12}/L	< 0.9 * LLN	> 1.1 * ULN
White Blood Cell Count	10^9/L	< 0.7 * LLN	> 1.5 * ULN
URINALYSIS			
pH		< 0.9 * LLN	> 1.1 * ULN
Specific Gravity		—	> 1.1 * ULN

LLN: Lower limit of normal value provided by the laboratory.
ULN: Upper limit of normal value provided by the laboratory.

11.3 VITAL SIGNS

Descriptive statistics for vital signs (systolic and diastolic BP, pulse rate, weight, respiration rate, and temperature) at baseline (screening) and changes from baseline at each assessment will be presented by study intervention.

Vital sign values will be considered PCS if they meet both the observed-value criteria and the change from baseline criteria listed in Table 11-2. The number and percentage of participants who have PCS post-baseline clinical laboratory values will be tabulated by study intervention at each assessment. The percentages will be calculated relative to the number of participants who have available non-PCS baseline values and at least 1 post-baseline assessment. The numerator will be the total number of participants with at least 1 PCS post-baseline value. A supportive listing of participants with PCS post-baseline values will be provided for the Safety and Safety-T3 Populations.

Table 11-2 Criteria for Potentially Clinically Significant Vital Signs

Parameter	***Flag***	***Criteria[a]***	
		Observed Value	*Change From Baseline*
Sitting systolic blood pressure, mm Hg	High	≥ 140	Increase of ≥ 20
	Low	≤ 90	Decrease of ≥ 20
Sitting diastolic blood pressure, mm Hg	High	≥ 90	Increase of ≥ 15
	Low	≤ 50	Decrease of ≥ 15
Sitting pulse rate, bpm	High	≥ 100	Increase of ≥ 15
	Low	≤ 50	Decrease of ≥ 15
Weight, kg	High	—	Increase of ≥ 7%
	Low	—	Decrease of ≥ 7%

a A post-baseline value is considered potentially clinically significant if it meets both the observed-value and the change from baseline criteria.
bpm = beats per minute.

11.4 ELECTROCARDIOGRAM

Descriptive statistics for ECG parameters (heart rate, PR interval, QRS duration, QT interval, and QTc) at baseline, EOS, changes from baseline at EOS, interim ECG parameters at baseline (pre-dose at each period), and changes from baseline at post-dose will be presented by study intervention.

Electrocardiographic parameter values are considered PCS if they meet or exceed the higher-limit PCS criteria listed in Table 11-3. The number and percentage of participants who have PCS post-baseline ECG values will be tabulated by study intervention for EOS assessments and interim assessments. The percentages will be calculated relative to the number of participants who have available non-PCS baseline values and at least 1 post-baseline assessment. The numerator will be the total number of participants with at least 1 PCS post-baseline value. A supportive listing of participants with PCS post-baseline values will be provided, including the participant number and baseline and post-baseline values, for the Safety and Safety-T3 Populations.

Table 11-3 Criteria for Potentially Clinically Significant Electrocardiographic Values

Parameter	***Unit***	***Criterion Value*** [a]	***Change from Baseline*** [a]
QRS duration	msec	≥ 150	—
PR interval	msec	≥ 250	—
QTcF interval for male	msec	> 450	Increase of > 60
QTcF interval for female	msec	> 470	Increase of > 60

a post-baseline (end-of-study or post-dose) value will be considered potentially clinically significant if it meets the criterion value or the change from baseline or pre-dose value.

PCS = potentially clinically significant;
QTcF = QT interval/$(60/RR)^{1/3}$.

11.5 OTHER SAFETY PARAMETERS

11.5.1 Physical Examination

Any new physical examination abnormality identified on the post-baseline physical examination or any physical examination abnormality noted as worsened from Screening (Visit 1) to the post-baseline physical examination will be reported as an AE if considered by the investigator to be clinically significant. No separate data analysis for physical exams is planned.

11.5.2 Potential Hy's Law

Potential Hy's Law criteria within a 24-hour window is defined by a post baseline elevation of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) ≥ 3x ULN, along with total bilirubin (TBL) ≥ 2x ULN and a non-elevated alkaline phosphatase (ALP) < 2x ULN, all based on blood draws collected within a 24-hour period.

Potential Hy's Law criteria without time window (e-DISH) is defined by maximum of post baseline elevation of ALT or AST ≥ 3x ULN, along with maximum of post baseline elevation of TBL ≥ 2x ULN and a non-elevated alkaline phosphatase (ALP) < 2x ULN.

Participants who meet the potential Hy's Law criteria from the first dose of study drug to within 35 days after the last dose of study treatment will be summarized. Supportive tabular displays will also be provided.

12.0 HEALTH OUTCOMES ANALYSES

Analyses for health economics and outcomes parameters will be described in a separate document.

13.0 INTERIM ANALYSIS

No interim analysis is planned for this study.

14.0 DETERMINATION OF SAMPLE SIZE

[redacted]

15.0 STATISTICAL SOFTWARE

Statistical analyses will be performed using version 9.4 (or newer) of SAS on a Linux operating system.

16.0 DATA HANDLING CONVENTIONS

16.1 BASELINE DEFINITION

Unless otherwise specified, the baseline value will be the last non-missing value obtained prior to or on the first treatment dose.

16.2 VISIT TIME WINDOWS

The following tables present the visits assigned for analyses and the corresponding range of treatment days (window) during which an actual visit may occur.

1) Participants for 18-week Titration Period Visit Time Windows for Analysis

Derived Visit		***Window***
Visit 2 Baseline	Day 1	Days ≤ 1
Visit 3a Week 6	Day 42	Days [2, 63]
Visit 4a Week 12	Day 84	Days [64, 105]
Visit 5a Week 18	Day 126	Days [106, last day of TP]
Visit 6 Week 24	Day 168	Days [first day of SP, 189]
Visit 7 Week 30	Day 210	Days > 189

2) Participants for 24-week Titration Period Visit Time Windows for Analysis

Derived Visit		***Window***
Visit 2 Baseline	Day 1	Days ≤ 1
Visit 3a Week 6	Day 42	Days [2, 63]
Visit 4a Week 12	Day 84	Days [64, 105]
Visit 5a Week 18	Day 126	Days [106, 147]
Visit 5c Week 24	Day 168	Days [148, last day of TP]
Visit 6 Week 30	Day 210	Days [first day of SP, 231]
Visit 7 Week 36	Day 252	Days > 231

3) Participants for 30-week Titration Period Visit Time Windows for Analysis

Derived Visit		***Window***
Visit 2 Baseline	Day 1	Days ≤ 1
Visit 3a Week 6	Day 42	Days [2, 63]
Visit 4a Week 12	Day 84	Days [64, 105]
Visit 5a Week 18	Day 126	Days [106, 147]
Visit 5c Week 24	Day 168	Days [148, 189]
Visit 5e Week 30	Day 210	Days [190, last day of TP]
Visit 6 Week 36	Day 252	Days [first day of SP, 273]
Visit 7 Week 42	Day 294	Days > 273

4) Participants for 36-week Titration Period Visit Time Windows for Analysis

Derived Visit		***Window***
Visit 2 Baseline	Day 1	Days ≤ 1
Visit 3a Week 6	Day 42	Days [2, 63]
Visit 4a Week 12	Day 84	Days [64, 105]
Visit 5a Week 18	Day 126	Days [106, 147]
Visit 5c Week 24	Day 168	Days [148, 189]
Visit 5e Week 30	Day 210	Days [190, 231]
Visit 5g Week 36	Day 252	Days [232, last day of TP]
Visit 6 Week 42	Day 294	Days [first day of SP, 315]
Visit 7 Week 48	Day 336	Days > 315

TP = Titration Period; SP = Stabilization Period.

Last day of TP is the day of last TSH visit prior to visit 6.

First day of SP is defined as the last day of TP + 1.

If the visit date is on or after the date of the first dose of double-blind study treatment, the study day is calculated by visit date – date of the first dose of double-blind investigational product + 1. If the visit date is before the date of the first dose of double-blind investigational product, the study day is calculated by visit date – date of the first dose of double-blind investigational product. Therefore, a negative day indicates a day before the start of the double blind investigational product.

If a participant has 2 or more visits within the same window, the last visit with a non-missing value will be used for analysis.

The titration visits 5b, 5d, and 5f will be included in the analysis visits 5a, 5c, 5e, or 5g based on the definition of analysis visit window.

16.3 REPEATED OR UNSCHEDULED ASSESSMENTS OF SAFETY PARAMETERS

If a participant has repeated assessments before the start of double-blind treatment with investigational product, the results from the final assessment made before the date of the first dose will be used as baseline. If EOT assessments are repeated or unscheduled, the last post-baseline assessment will be used as the EOT assessment for generating summary statistics.

However, all post-baseline assessments will be used for PCS value determination, and all assessments will be presented in the data listings.

16.4 MISSING DATE OF THE LAST DOSE OF STUDY TREATMENT

When the date of the last dose of the double-blind investigational product is missing, all efforts should be made to obtain the date from the Investigator. If it is still missing after all efforts have been made, the last diary date will be used as the last dose date.

16.5 MISSING SEVERITY ASSESSMENT FOR ADVERSE EVENTS

If severity is missing for an AE that started before the date of the first dose of double-blind investigational product, an intensity of mild will be assigned. If severity is missing for an AE that started on or after the date of the first dose of double-blind investigational product, an intensity of severe will be assigned. The imputed values for severity assessment will be used for the incidence summary; the values will be shown as missing in the data listings.

16.6 MISSING CAUSAL RELATIONSHIP TO STUDY TREATMENT FOR ADVERSE EVENTS

If the causal relationship to the investigational product is missing for an AE that started on or after the date of the first dose of double-blind investigational product, a causality of yes will be assigned. The imputed values for causal relationship to double-blind treatment will be used for the incidence summary; the values will be shown as missing in the data listings.

16.7 MISSING DATE INFORMATION FOR ADVERSE EVENTS

The following imputation rules apply to cases in which the start date is incomplete (i.e., partial missing) for AEs.

Missing day and month

- If the year is the same as the year of the date of the first dose of double-blind investigational product, then the day and month of the date of the first dose of double-blind investigational product will be assigned to the missing fields.

- If the year is prior to the year of the date of the first dose of double-blind investigational product, then December 31 will be assigned to the missing fields.

- If the year is after the year of the date of the first dose of double-blind investigational product, then January 1 will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year are the same as the month and year of the date of the first dose of double-blind investigational product, then the date of the first dose of double-blind investigational product will be assigned to the missing day.

- If either the year is before the year of the date of the first dose of double-blind investigational product or if both years are the same but the month is before the month of the date of the first dose of double-blind investigational product, then the last day of the month will be assigned to the missing day.

- If either the year is after the year of the date of the first dose of double-blind investigational product or if both years are the same but the month is after the month of the date of the first dose of double-blind investigational product, then the first day of the month will be assigned to the missing day.

If the stop date is complete and the imputed start date as above is after the stop date, the start date will be imputed by the stop date.

If the start date is completely missing and the stop date is complete, then the following algorithm is used to impute the start date:

- If the stop date is on or after the date of the first dose of double-blind investigational product, the date of the first dose of double-blind investigational product will be assigned to the missing start date.

- If the stop date is before the date of the first dose of double-blind investigational product, the stop date will be assigned to the missing start date.

16.8 MISSING DATE INFORMATION FOR PRIOR OR CONCOMITANT MEDICATIONS

For prior or concomitant medications, excluding rescue medications, incomplete (i.e., partly missing) start dates and/or stop dates will be imputed. When the start date and the stop date are both incomplete for a participant, the start date will be imputed first.

16.8.1 Incomplete Start Date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication start date. If the stop date is complete (or imputed) and the imputed start date is after the stop date, the start date will be imputed using the stop date.

Missing month and day

- If the year of the incomplete start date is the same as the year of the first dose of study treatment, the month and day of the first dose of study treatment will be assigned to the missing fields

- If the year of the incomplete start date is before the year of the first dose of study treatment, *December 31* will be assigned to the missing fields

- If the year of the incomplete start date is after the year of the first dose of study treatment, *January 1* will be assigned to the missing fields

Missing month only

- If only the month is missing, the day will be treated as missing and both the month and the day will be replaced according to the above procedure

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the first dose of study treatment, the day of the first dose of study treatment will be assigned to the missing day

- If either the year of the incomplete start date is before the year of the date of the first dose of study treatment or if both years are the same but the month of the incomplete start date is before the month of the date of the first dose of study treatment, the last day of the month will be assigned to the missing day.

- If either the year of the incomplete start date is after the year of the date of the first dose of study treatment or if both years are the same but the month of the incomplete start date is after the month of the date of the first dose of study treatment, the first day of the month will be assigned to the missing day

16.8.2 Incomplete Stop Date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication stop date. If the date of the last dose of study treatment is missing, impute it as descripted in Section 16.4. If the imputed stop date is before the start date (imputed or non-imputed start date), the imputed stop date will be equal to the start date.

Missing month and day

- If the year of the incomplete stop date is the same as the year of the last dose of study treatment, the month and day of the last dose of study treatment will be assigned to the missing fields

- If the year of the incomplete stop date is before the year of the last dose of study treatment, *December 31* will be assigned to the missing fields

- If the year of the incomplete stop date is after the year of the last dose of study treatment, *January 1* will be assigned to the missing fields

Missing month only

- If only the month is missing, the day will be treated as missing and both the month and the day will be replaced according to the above procedure

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the last dose of study treatment, the day of the last dose of study treatment will be assigned to the missing day

- If either the year of the incomplete stop date is before the year of the date of the last dose of study treatment or if both years are the same but the month of the incomplete stop date is before the month of the date of the last dose of study treatment, the last day of the month will be assigned to the missing day

- If either the year of the incomplete stop date is after the year of the date of the last dose of study treatment or if both years are the same but the month of the incomplete stop date is after the month of the date of the last dose of study treatment, the first day of the month will be assigned to the missing day.

16.9 CHARACTER VALUES OF CLINICAL LABORATORY PARAMETERS

If the reported value of a clinical laboratory parameter cannot be used in a statistical summary table because, for example, the value is a character string rather than a numerical type, a coded value needs to be appropriately determined and used in the statistical analyses. However, the actual values as reported in the database will be presented in data listings.

Table 16-1 Examples for Coding of Special Character Values for Clinical Laboratory Parameters

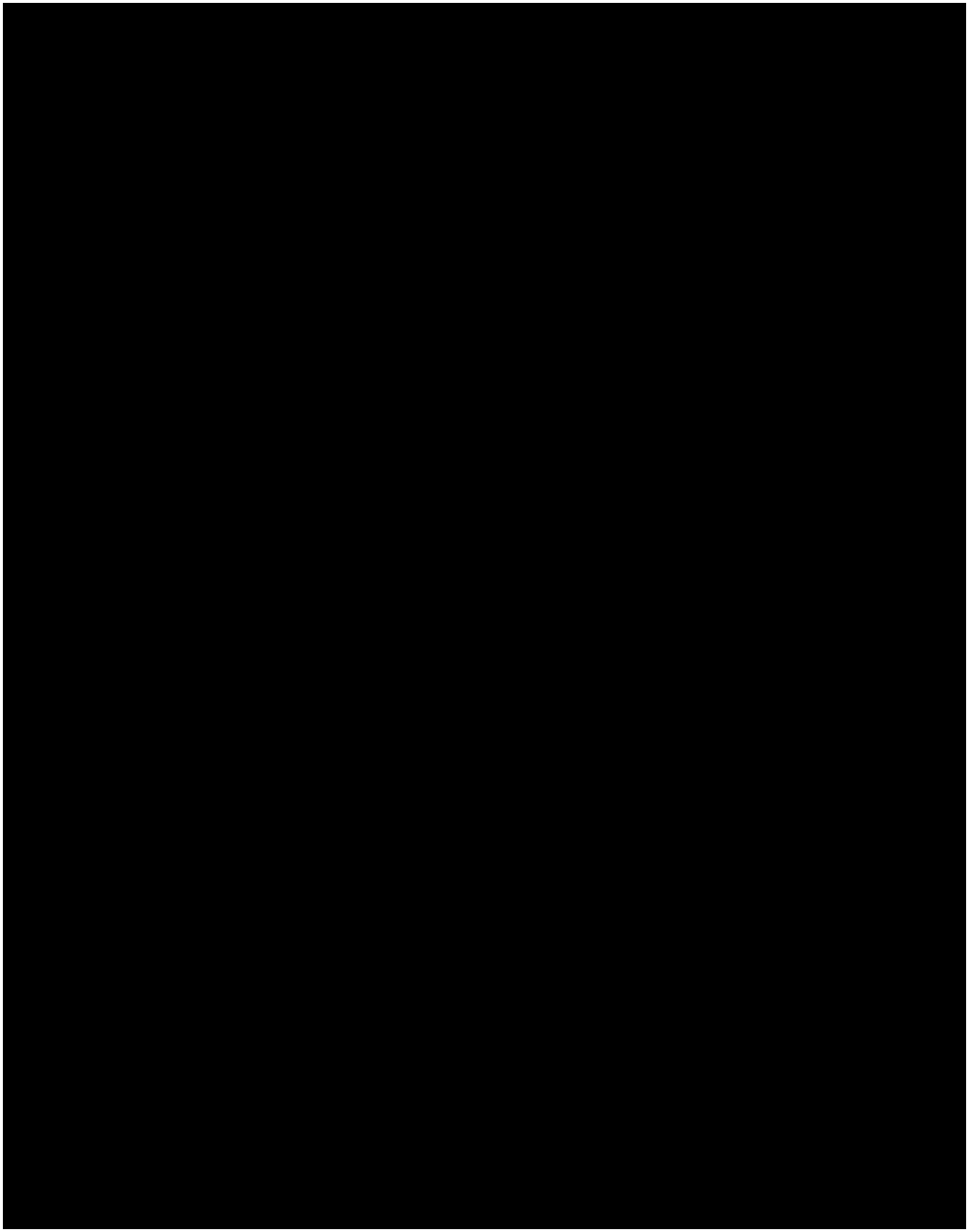
Laboratory Test (Unit)	***Possible Lab Results (in SI units)***	***Coded Value for Analysis***
Chemistry: ALT	< 5	0
Chemistry: AST	< 5	0
Chemistry: Bilirubin, Total	< 2	0
Urinalysis: Glucose	= OR > 55, >= 55, > 0	Positive
	<= 0, Negative	Negative
Urinalysis: Ketones	= OR > 8.0, >=8.0, > 0	Positive
	<= 0, Negative	Negative
Urinalysis: pH	> 8.0, >= 8.0	8.0
	>= 8.5,	8.5
Urinalysis: Protein	= OR > 3.0, >=3.0, > 0	Positive
	<= 0	Negative

17.0 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

[redacted]

18.0 SUMMARY OF CHANGES FROM THE ORIGINAL SAP

19.0 APPENDIX A: JUSTIFICATION FOR NON-INFERIORITY MARGIN



20.0 APPENDIX B: ATTRIBUTES OF THE ESTIMAND FOR PRIMARY AND SECONDARY ENDPOINTS

Estimand	Attributes of the Estimands				
	Treatment	**Endpoint**	**Population**	**Intercurrent Events**	**Statistical Summary**
Primary	Armour Thyroid vs. Synthetic T4	Achievement of Sustained TSH Response	ITT	IE1: Premature discontinuation of study drug solely due to COVID-19 related reasons IE2: Premature discontinuation of study drug due to other reasons not attributable to COVID-19 All data after IE1 or IE2 will be used. Participants who encounter IE1 will be imputed via multiple imputation under the Missing at Random (MAR) mechanism. Participants who encounter IE2 will be imputed as non-responders.	Stratum-adjusted difference in the percentage of sustained TSH response at end of Titration Period and Stabilization Period, and associated 2-sided 95% confidence interval.
Secondary	Armour Thyroid vs. Synthetic T4	Achievement of titration TSH Response	ITT	IE1: Premature discontinuation of study drug during the Titration Period solely due to COVID-19 related reasons IE2: Premature discontinuation of study drug during the Titration Period due to other reasons not attributable to COVID-19 All data after IE1 or IE2 will be used. Participants who encounter IE1 will be imputed via multiple imputation under the Missing at Random (MAR) mechanism. Participants who encounter IE2 will be imputed as non-responders.	Stratum-adjusted difference in the percentage of titration TSH response at end of Titration Period, and associated 2-sided 95% confidence interval.

21.0 APPENDIX C: THYPRO QUESTIONNAIRE

21.1 SCORING OF THE THYPRO QUESTIONNAIRE

Each of the 13 ThyPRO scales is scored as a summary score and linearly transformed to range 0-100. This scoring procedure is described step by step below.

Naming of items

Each item is named TQ{relevant item number}. Item number is indicated left of each item in the questionnaire. Thus, the first item, 'sensation of fullness in the neck' is named TQ1A in the database.

Scoring of responses

Item responses are scored

0 for 'Not at all'

1 for 'A little'

2 for 'Some'

3 for 'Quite a bit'

4 for 'Very much'

Scale content

Each scale consists of the following items:

Goitre symptoms scale: TQ1A TQ1B TQ1C TQ1D TQ1E TQ1F TQ1G TQ1H TQ1I TQ1J TQ1K;

Hyperthyroid symptoms scale: TQ1L TQ1M TQ1N TQ1O TQ1P TQ1R TQ1S TQ1T;

Hypothyroid symptoms scale: TQ1Q TQ1CC TQ1DD TQ1EE;

Eye symptoms scale: TQ1U TQ1V TQ1W TQ1X TQ1Y TQ1Z TQ1AA TQ1BB;

Tiredness scale: TQ2A TQ2B TQ2C TQ2D TQ3A TQ3B TQ3C;

Cognitive problems scale: TQ4A TQ4B TQ4C TQ4D TQ4E TQ4F;

Anxiety scale: TQ5A TQ5B TQ5C TQ5D TQ5E TQ5F;

Depressivity scale: TQ6A TQ6B TQ6C TQ6D TQ6E TQ6F TQ6G;

Emotional Susceptibility scale: TQ7A TQ7B TQ7C TQ7D TQ7E TQ7F TQ7G TQ7H TQ7I;

Impaired Social life scale: TQ8A TQ8B TQ8C TQ8D;

Impaired Daily life scale: TQ9A TQ9B TQ9C TQ9D TQ9E TQ9F;

Impaired Sexlife scale: TQ10A TQ10B;

Cosmetic Complaints scale: TQ11A TQ11B TQ11C TQ11D TQ11E TQ11F;

Scale scoring

In principle, scales are scored by simply adding the response values (0-4) for all the items in a scale. Thus, if a patient answered 'Not at all' to all items in the Social impairment scale, the score would be 0; whereas if she answered 'Not at all' to two items, 'A little' to one item and 'Some' to the last item, the score would be 3 (and max score would be 16). However, three additional features expand this procedure: reversal of positively worded items, imputation for individual missing item responses and linear transformation.

Reversal of positively worded items

As part of the scoring procedure, items 3A, 3B, 3C, 6F, 6G, 7H and 7I have to be reversed, i.e. 'Not at all' scored as 4, 'A little' as 3, 'Some' as 2, 'Quite a bit' as 1 and 'Very much' scored as 0.

Scales are transformed linearly to range 0-100 according to the formula

Transformed score=(raw sum score/highest possible raw sum score)*100.

Individual missing items will be imputed as follows. If half or more of the items in a scale is completed, missing items are substituted by the mean of the completed items. Otherwise, total score (or scale) will be set to missing.

For example, the abovementioned patient answering 'Not at all' to two items, 'A little' to one item and 'Some' to the last item, would have a raw score of 3 (0+0+1+2). Highest possible raw score would be 16 (4+4+4+4). The transformed 0-100 score would then be 3/16*100=19.

21.2 SCORING OF THE THYPRO-39 QUESTIONNAIRE

Each of the 13 ThyPRO-39 scales is scored as a summary score and transformed to range 0-100. This scoring procedure is described step by step below.

Naming of items

Each item is named TQ{relevant item number}. Item number is indicated left of each item in the questionnaire. Thus, the first item, 'sensation of fullness in the neck' is named TQ1A in the database.

Scoring of responses

Item responses are scored

0 for 'Not at all'

1 for 'A little'

2 for 'Some'

3 for 'Quite a bit'

4 for 'Very much'/'Completely'

Scale content

Each scale consists of the following items:

Goitre symptoms scale: TQ1A TQ1C TQ1H

Hyperthyroid symptoms scale: TQ1l TQ1M TQ1N TQ1T

Hypothyroid symptoms scale: TQ1Q TQ1CC TQ1DD TQ1EE

Eye symptoms scale: TQ1W TQ1X TQ1BB

Tiredness scale: TQ2A TQ2C TQ3B

Cognitive problems scale: TQ4A TQ4B TQ4F;

Anxiety scale: TQ5B TQ5C TQ5E

Depressivity scale: TQ6A TQ6E TQ6G

Emotional Susceptibility scale: TQ7C TQ7D TQ7H

Impaired Social life scale: TQ8A TQ8B TQ8C

Impaired Daily life scale: TQ9A TQ9C TQ9E

Cosmetic Complaints scale: TQ11A TQ11D TQ11E

Overall QoL: TQ12

Composite scale: TQ2A TQ2C TQ3B TQ4A TQ4B TQ4F TQ5B TQ5C TQ5E TQ6A TQ6E TQ6G TQ7C TQ7D TQ7H TQ8A TQ8B TQ8C TQ9A TQ9C TQ9E TQ12

Scoring the individual scales

The raw scale scores are derived by simply adding the response values (0-4) for all the items in a scale. However, three additional features expand this procedure:

1. reversal of positively worded items,

2. imputation for individual missing item responses and

3. 0-100 transformation.

Reversal of positively worded items

As part of the scoring procedure, items 3B, 6G and 7H have to be reversed, i.e. 'Not at all' scored as 4, 'A little' as 3, 'Some' as 2, 'Quite a bit' as 1 and 'Very much'/'Completely scored as 0.

Imputation for individual missing items

If half or more of the items in a scale is completed, missing items are substituted by the mean of the completed items.

Linear transformation

All scales (except for the Hypothyroid Symptoms, the Overall QoL scale and the Composite scale, see below) are transformed to range 0-100 according to Table 21-1 below:

Table 21-1 **To the left is the raw score. The corresponding 0-100 score is tabulated for each scale separately. For example, a patient with a raw score on the Goiter Symptoms scale of 6 (e.g. because she answered "Some" to all three Goiter items), will have a 0-100 Goiter Symptoms score of 37. A raw score of 6 on the Tiredness scale, would yield a 0-100 score of 50.**

Raw sum score	Final rescaled short-form score										
	Goiter	Hyper	Eye	Tired	Cogni-tion	Anxiety	Depres-sion	Suscep-tibility	Social Life	Daily Life	Appear-ance
0	2	2	1	0	1	1	0	1	0	0	1
1	10	8	8	8	7	10	7	7	8	7	12
2	15	13	14	17	14	18	14	13	17	15	21
3	20	18	20	25	21	26	22	21	25	22	28
4	26	23	25	33	29	34	29	28	33	30	36
5	31	28	32	42	37	41	37	36	42	38	43
6	37	33	38	50	44	49	45	44	50	46	51
7	43	38	45	58	52	56	54	52	58	54	59
8	49	44	52	67	60	63	63	60	67	62	66
9	57	49	60	75	68	71	71	68	75	71	73
10	64	55	68	83	76	79	80	77	83	80	80
11	73	60	78	92	85	87	89	86	92	89	87
12	84	66	89	100	95	96	97	95	100	98	96
13		71									
14		77									
15		84									
16		90									

Transformation of the Hypothyroid Symptoms scale:

The ThyPRO-39 Hypothyroid Symptoms is identical to the original ThyPRO Hypothyroid Symptoms scale and is thus transformed to 0-100 according to the formula

Transformed score=(raw sum score/16)*100

For example, if a patient answered 'Not at all' to two items, 'A little' to one item and 'Some' to the last item, she would have a raw score of 3 (0+0+1+2). The transformed 0-100 score would then be 3/16*100=19.

Transformation of the Overall QoL-impact scale/item:

The Overall QoL item (TQ12) is rescaled to 0-100 simply by taking the mean raw score and multiply by 25.

Scoring the Composite scale

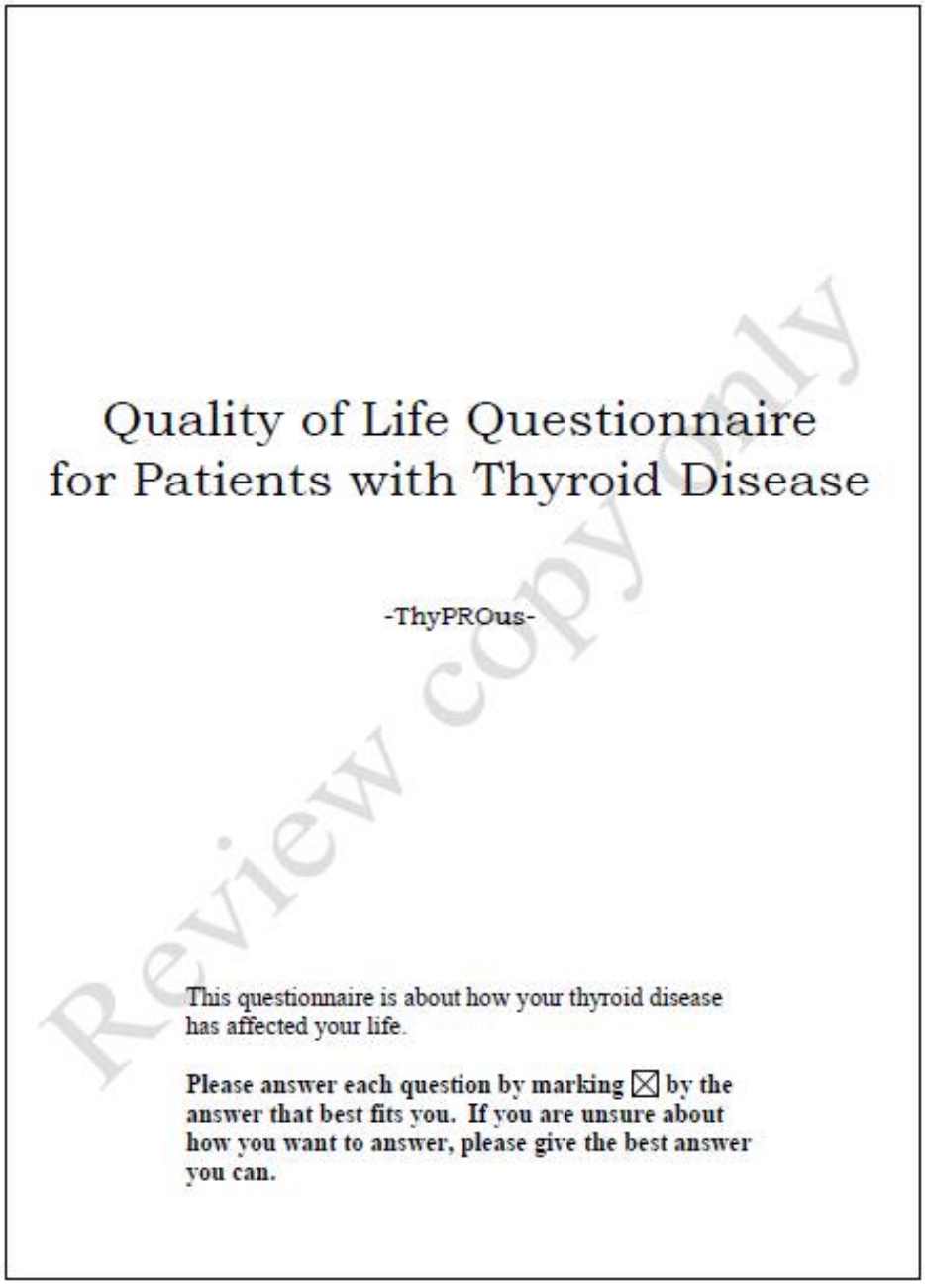
The Composite scale is based on the 22 items from the Tiredness, Cognition, Anxiety, Depressivity, Emotional Susceptibility, Impaired Social life Impaired Daily Life and Overall QoL scales:

TQ2A TQ2C TQ3B TQ4A TQ4B TQ4F TQ5B TQ5C TQ5E TQ6A TQ6E TQ6G TQ7C TQ7D TQ7H TQ8A TQ8B TQ8C TQ9A TQ9C TQ9E TQ12

The raw score is derived as described above, by summation (with imputation for missing), to range 0-88. The raw score is transformed to 0-100 according to the formula

Transformed score=(raw sum score/88)*100

21.3 PATIENT-REPORTED OUTCOMES QUESTIONNAIRES, DESCRIPTIONS, AND INSTRUCTIONS – THYPRO QUESTIONNAIRE



Quality of Life Questionnaire
for Patients with Thyroid Disease

-ThyPROus-

This questionnaire is about how your thyroid disease has affected your life.

Please answer each question by marking ☒ by the answer that best fits you. If you are unsure about how you want to answer, please give the best answer you can.

1

The first section of the questionnaire is about symptoms, tiredness, memory, mood, and health.

Please base your **answers on how you have been feeling in general** during the past 4 weeks.

1. ***The first questions are about symptoms***

During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
1a - had the sensation of fullness in the neck?	☐	☐	☐	☐	☐
1b - had a visible swelling in the front of your neck?	☐	☐	☐	☐	☐
1c - felt pressure in your throat?	☐	☐	☐	☐	☐
1d - had pain in the front of your throat?	☐	☐	☐	☐	☐
1e - had pain in your neck that could be felt in your ears?	☐	☐	☐	☐	☐
1f - had the sensation of a lump in your throat?	☐	☐	☐	☐	☐
1g - had the need to clear your throat frequently?	☐	☐	☐	☐	☐
1h - felt discomfort swallowing?	☐	☐	☐	☐	☐
1i - had difficulty swallowing?	☐	☐	☐	☐	☐
1j - had the sensation of suffocating?	☐	☐	☐	☐	☐
1k - been hoarse?	☐	☐	☐	☐	☐
1l - had trembling hands?	☐	☐	☐	☐	☐
1m - had a tendency to sweat a lot?	☐	☐	☐	☐	☐
1n - experienced palpitations (rapid heart beat)?	☐	☐	☐	☐	☐
1o - experienced shortness of breath?	☐	☐	☐	☐	☐
1p - been sensitive to heat?	☐	☐	☐	☐	☐
1q - been sensitive to cold?	☐	☐	☐	☐	☐
1r - had an increased appetite?	☐	☐	☐	☐	☐

	During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
1s	- had loose stools?	☐	☐	☐	☐	☐
1t	- had an upset stomach?	☐	☐	☐	☐	☐
1u	- had moist or watery eyes?	☐	☐	☐	☐	☐
1v	- had bags under the eyes or swollen eyelids?	☐	☐	☐	☐	☐
1w	- had the sensation of dryness or "grittiness" in the eyes?	☐	☐	☐	☐	☐
1x	- had impaired vision?	☐	☐	☐	☐	☐
1y	- felt pressure in (or behind) the eyes?	☐	☐	☐	☐	☐
1z	- had double vision?	☐	☐	☐	☐	☐
1aa	- had eye pain?	☐	☐	☐	☐	☐
1bb	- been very sensitive to light?	☐	☐	☐	☐	☐
1cc	- had swollen hands or feet?	☐	☐	☐	☐	☐
1dd	- had dry skin?	☐	☐	☐	☐	☐
1ee	- had itchy skin?	☐	☐	☐	☐	☐

2. ***The following questions are about tiredness***

	During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
2a	- been tired?	☐	☐	☐	☐	☐
2b	- been exhausted?	☐	☐	☐	☐	☐
2c	- had difficulty getting motivated to do anything at all?	☐	☐	☐	☐	☐
2d	- felt worn out?	☐	☐	☐	☐	☐

3

3. *The following questions are about your vitality*

During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
3a - felt full of life?	☐	☐	☐	☐	☐
3b - felt energetic?	☐	☐	☐	☐	☐
3c - been able to cope with the demands of your life?	☐	☐	☐	☐	☐

4. *The following questions are about memory and concentration*

During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
4a - had difficulty remembering?	☐	☐	☐	☐	☐
4b - had slow or unclear thinking?	☐	☐	☐	☐	☐
4c - had difficulty finding the right words?	☐	☐	☐	☐	☐
4d - been confused?	☐	☐	☐	☐	☐
4e - had difficulty learning something new?	☐	☐	☐	☐	☐
4f - had difficulty concentrating?	☐	☐	☐	☐	☐

4

5. ***The following questions are about nervousness and tension***

	During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
5a	- felt nervous?	☐	☐	☐	☐	☐
5b	- felt afraid or anxious?	☐	☐	☐	☐	☐
5c	- felt tense?	☐	☐	☐	☐	☐
5d	- been concerned about being seriously ill?	☐	☐	☐	☐	☐
5e	- felt uneasy?	☐	☐	☐	☐	☐
5f	- felt restless?	☐	☐	☐	☐	☐

6. ***The following questions are about psychological well-being***

	During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
6a	- felt sad?	☐	☐	☐	☐	☐
6b	- felt depressed?	☐	☐	☐	☐	☐
6c	- felt discouraged?	☐	☐	☐	☐	☐
6d	- cried easily?	☐	☐	☐	☐	☐
6e	- felt unhappy?	☐	☐	☐	☐	☐

	During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
6f	- felt happy?	☐	☐	☐	☐	☐
6g	- had self-confidence?	☐	☐	☐	☐	☐

5

7. *The following questions are about having difficulty coping or having mood swings*

During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
7a - had difficulty coping?	☐	☐	☐	☐	☐
7b - felt "not like yourself"?	☐	☐	☐	☐	☐
7c - noticed you easily felt stressed?	☐	☐	☐	☐	☐
7d - had mood swings?	☐	☐	☐	☐	☐
7e - felt irritable?	☐	☐	☐	☐	☐
7f - felt frustrated?	☐	☐	☐	☐	☐
7g - felt angry?	☐	☐	☐	☐	☐

	Not at all	A little	Some	Quite a bit	Com-pletely
7h - felt in control of your life?	☐	☐	☐	☐	☐
7i - felt in balance?	☐	☐	☐	☐	☐

The remainder of the questionnaire is about **how your thyroid disease may have affected various aspects of your life**

8. ***The following questions are about your relationships with other people***

During the past 4 weeks, <u>has your thyroid disease caused</u> you to	Not at all	A little	Some	Quite a bit	Very much
- have difficulty being together with other people (for example, spouse, children, boy/girlfriend, friends, or others)?	☐	☐	☐	☐	☐
- feel you were a burden to other people?	☐	☐	☐	☐	☐
- have conflicts with other people?	☐	☐	☐	☐	☐

During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
- felt that people in your surroundings have lacked understanding of your thyroid disease?	☐	☐	☐	☐	☐

9. *The following questions are about your daily activities*

During the past 4 weeks, has your thyroid disease caused you to	Not at all	A little	Some	Quite a bit	Very much
- have difficulty managing your daily life?	☐	☐	☐	☐	☐
- limit your leisure activities or hobbies?	☐	☐	☐	☐	☐
- not be able to participate in life around you?	☐	☐	☐	☐	☐
- have difficulty getting around (for example, walking, running, bicycling, or driving a car)?	☐	☐	☐	☐	☐
- feel as if everything takes longer to do?	☐	☐	☐	☐	☐

During the past 4 weeks, has your thyroid disease caused you to	I do not work	Not at all	A little	Some	Quite a bit	Very much
- have difficulty managing your job (for example, finding it hard to cope or calling in sick)?	☐	☐	☐	☐	☐	☐

8

10. ***The following questions are about your sex life***

	During the past 4 weeks have you	Not at all	A little	Some	Quite a bit	Very much
10a	- felt your thyroid disease had a negative influence on your sex life?	☐	☐	☐	☐	☐
10b	- had a decreased sexual desire?	☐	☐	☐	☐	☐

11. ***Thyroid diseases (or their treatment) may affect your appearance. (For example, by causing swelling of the neck, swollen face, hands, or feet, or changes in weight or to the eyes.)***

	During the past 4 weeks,	Not at all	A little	Some	Quite a bit	Very much
11a	- has your thyroid disease affected your appearance (for example, swelling of the neck, eye changes, weight changes)?	☐	☐	☐	☐	☐
11b	- have you been unsatisfied with your appearance because of your thyroid disease?	☐	☐	☐	☐	☐
11c	- have you tried to camouflage or mask visible signs of your thyroid disease (for example, by wearing a scarf or sunglasses)?	☐	☐	☐	☐	☐
11d	- have you been bothered by other people looking at you?	☐	☐	☐	☐	☐
11e	- has your thyroid disease influenced which clothes you wear?	☐	☐	☐	☐	☐
11f	- has your thyroid disease made you feel too fat?	☐	☐	☐	☐	☐

12. ***The final question is about to what extent your thyroid disease has affected you overall during the past 4 weeks***

During the past 4 weeks,	Not at all	A little	Some	Quite a bit	Very much
12 - has your thyroid disease had a negative effect on your quality of life?	☐	☐	☐	☐	☐

Please go back and check that you have answered all the questions.

Thank you very much for your help answering this questionnaire!

22.0 REFERENCES

1. Agresti A (2013). Categorical Data Analysis. 3rd ed. Hoboken, NJ: John Wiley & Sons.

2. Caty SA, Marta OB, Jordi R, Leonardo G, Inaki G and Elisabeth M (2017). Factors Associated with the Stability of Thyroid-Stimulating Hormone Values in Hypothyroidism. Arch Med. 2017 9:2.

3. Comtois R, Faucher L, Lafleche L (1995). Outcome of Hypothyroidism Caused by Hashimoto's Thyroiditis. Arch Intern Med. 1995 Jul;155: 1404-1408.

4. Hoang TD, Olsen CH, Mai VQ, Clyde PW, Shakir MK. (2013). Desiccated thyroid extract compared with levothyroxine in the treatment of hypothyroidism: a randomized, double-blind, crossover study. J Clin Endocrinol Metab. 2013 May; 98(5):1982-90.

5. Hofling D, Chavantes MC, Juliano AG, Cerri G, Knobel M, Yoshimura E, Chammas, MC (2013). Low-level laser in the treatment of patients with hypothyroidism induced by chronic autoimmune thyroiditis: a randomized, placebo-controlled clinical trial. Lasers Med. Sci. 2013 Jun; 28: 743-753.

6. Kuye R, Riggs C, King J, Heilmann R, Kurz D and Milchak J (2020) Thyroid Stimulating Hormone Stability in Patients Prescribed Synthetic or Desiccated Thyroid Products: A Retrospective Study. The Annals of Family Medicine. 2020 September;18 (5) 452-454.

7. Yan X, Su XG.(2012) Stratified Wilson and Newcombe confidence intervals for multiple binomial proportions. Stat Biopharm Res. 2012 2(3):329-335.

8. Yan Y, Lee S and Li N. (2009) Missing Data Handling Methods in Medical Device Clinical Trials. Journal of Biopharmaceutical Statistics, 2009 19:6, 1085-1098.