

An Innovative Evidence-Based Laboratory Medicine (EBLM) Test to Assist Doctors in the Basic Assessment of Thyroid Function and Hypothalamic-Pituitary Axis

Authors: Adrià Roca¹, Sergio J. Calleja², José D. Santotoribio³

Affiliations: ¹Blueberry Diagnostics, Barcelona, Spain; ²Kience, Wilmington, U.S.; ³Hospital Universitario Puerto Real, Cádiz, Spain

BACKGROUND

Introduction
According to the National Center for Health Statistics (NCHS); there are many diseases whose prevalence is highly concerning. Some of these are related to the thyroid function, and the number of cases of these diseases is high and it is predicted to continue increasing in the next years. For this reason, they are frequently listed among the most common diseases. Such as hypothyroidism (Hypo), subclinical hypothyroidism (sHypo), hyperthyroidism (Hyper), and subclinical hyperthyroidism (sHyper), in the U.S., as well as their main autoimmune causes – Hashimoto thyroiditis (HT) and Graves' disease (GD) –.

Healthcare Reactive Model

Yet, current healthcare system is based on reactive to treat a patient when gets sick¹ (that is, when first symptoms or signs appear). However, many diseases are asymptomatic – clinically silent, subclinical or paucisymptomatic –, or have absence or lack of symptoms and signs until the disease is more advanced.
According to some estimates, healthcare reactive accounts for more than 75% of healthcare spending in the U.S.²
Besides, reactive health increases the lead time³ – the time between the early detection of a disease by screening tests and the time of usual diagnosis after the onset of symptoms or signs and the patient's visit to a doctor –. The lead time in detecting these diseases is crucial, because the sooner they are detected, the better the outcome will be for the patient^{4,5}. Nonetheless, many of these diseases have few or no symptoms in the early stages of the disease, so the patient does not know it until they are in an advanced stage (Figure 1). So, it is mandatory to detect them even in the very early stages, before symptoms appear and when treatment is most likely to be successful.
In response to the healthcare reactive model, the reactive model has been proposed as the solution for a longer and healthier life^{6,7}; but also as the best way to reduce healthcare-related costs^{8,9}. The preventive model is defined as the routine care that the patient receives to maintain its health^{10,11} and for the necessary steps to diagnosing medical conditions before they become a problem.

Healthcare Costs

The high cost of healthcare is a burden on U.S. families¹². About half of U.S. adults say it is difficult to afford healthcare costs^{13,14}, and one in four say they or a family member in their household had problems affording healthcare in the past 12 months¹⁵ – younger adults, those with lower incomes, adults in fair or poor health, and the uninsured are particularly likely to report problems affording healthcare in the last year...
The cost of healthcare can lead some to put off needed care^{16,17} – one in four adults say that in the past 12 months they have skipped or postponed getting healthcare they needed because of the cost^{18,19}...
About six in ten uninsured adults (61%) say they went without needed care because of the cost²⁰...
Healthcare debt is a burden for a large share of Americans – about four in ten adults (41%) report having debt due to medical or dental bills including debts owed to credit cards, collections agencies, family and friends, banks, and other lenders to pay for their healthcare costs²¹, with disproportionate shares of Black and Hispanic adults, women, parents, those with low incomes, and uninsured adults saying they have healthcare debt²²...

Chronic, Morbid and Cancer-Precursor Diseases, and Aging

The World Health Organization (WHO) estimates that chronic illnesses account for half of the global disease burden²³, a figure that will only rise as the world's population ages. Chronic diseases pose a unique challenge as they require proactive, planned and integrated care because they are continuous and often caused by specific and preventable health risks.
On the other hand, long waiting lists coexisting with the pandemic and new cases being treated further and further down the line²⁴. At the same time, we are seeing a consistent rise in chronic illnesses²⁵, such as cardiovascular disease, diabetes and cancer. Moreover, the risk of these chronic diseases actually increases with age as nearly 95% of adults 60 and older have at least one – while nearly 80% have two or more²⁶... With an aging population and these growing numbers, it is clear the model is not sustainable²⁷.

OBJECTIVES

- To define a minimum blood and/or urine –if needed–, panel capable of confirming –and/or detecting–, the main types of thyroid disease – hypothyroidism (Hypo) and subclinical hypothyroidism (sHypo), hyperthyroidism (Hyper) and subclinical hyperthyroidism (sHyper)–. Besides, if achieved, see if it is also possible to define the subtype of these diseases –primary or secondary–, depending on if the affection is in the thyroid gland or in the pituitary gland. In this process, the price should be a very important variable since this panel should be the cheapest one to enable universal and quality access to healthcare.
- To validate whether this new approach to a Evidence-Based Laboratory Medicine (EBLM) routine blood and/or urine –if needed–, could be used as a non-invasive test to assist doctors in the basic assessment –as well as screening–, of the main types of thyroid disease – hypothyroidism (Hypo) and subclinical hypothyroidism (sHypo), hyperthyroidism (Hyper) and subclinical hyperthyroidism (sHyper)–. The most common autoimmune causes of these diseases are Hashimoto thyroiditis (HT) and Graves' disease (GD) and all of their prevalences in the U.S. population are concerning (Figure 2) –. If achieved, use these results to help the medical community to understand how EBLM and new technologies –mainly machine learning (ML) algorithms (also known as AI-powered diagnostic tools), based in large and quality datasets–, can help healthcare professionals to improve diagnostic accuracy, as well as avoid invasive –and/or unnecessary– procedures.
- To fine-tune the final details of our algorithm as a preliminary step to the upcoming multi-center and international clinical trial of 26,000 patients that will be performed from March 2025 to December 2026 (we are still in the process of recruiting hospitals and medical centers).
- To validate the performance, accuracy and usefulness of several advanced EBLM indices, ratios, scores and/or coefficients –all of them analyzed individually and by different types of groups (and/or parallel)–, to optimize overall specificity (Sp) and sensitivity (Se), respectively –as tools based on machine learning (ML) algorithms for clinical decision support systems (CDSS), to improve healthcare delivery by enhancing medical decisions with targeted clinical knowledge, patient information, and other health information²⁸–.
- Anthropometric indices, ratios and/or products: body mass index (BMI), waist-to-hip ratio (WHR), Deurenberg body fat (%), Palafolls body fat (%), Hodgson Backster body fat (%), body fat mass, ideal body weight, Jackson Rollard, body fat to lose to ideal, body fat density, waist-to-height ratio (WHR), lipid accumulation product (LAP), body adiposity index (BAI), visceral adiposity index (VAI), body shape index (BSI), and conicity index (CI). All these indices, ratios and products were calculated from a few simple clinical variables, such as age, height, weight, neck circumference, waist circumference, and hip circumference.
- Thyroid ratios: free T4 hormone-to-thyroid stimulating hormone (FT4-to-TSH) ratio, and (FT4-to-TSH) ratio as a function of gender and age. All these ratios were calculated from a few simple clinical variables, such as gender, age, height, weight, free T4 hormone (FT4), and thyroid stimulating hormone (TSH).
- To validate the performance and accuracy of the algorithm when used with vendors other than those with which the original algorithms were developed –Synnex (Japan), Synnex (Switzerland) for biochemical and immunoassay–, and to compare the performance of several previous correlation studies alerted about potentially moderate differences in the performance between different reagent vendors –mainly in the normality limits–^{29,32}.
- To collect data for future mid and/or long-term studies related to health economics outcome research (HEOR)³⁰ to analyze the cost effectiveness of machine learning (ML) algorithms as a CDSS.

METHODS

This study was developed as a part of a previous one that has been presented at the European Society for Medical Oncology (ESMO) Congress 2024 for Multi-Cancer Early Detection (MCEd)³¹. From this previous study –with 90 routine laboratory determinations included (Table 5)–, new studies were conducted, such as the one described here.

In this way, on the one hand, to develop the original algorithm for an innovative evidence-based laboratory medicine (EBLM) test to assist doctors in the basic assessment of the thyroid function and hypothalamic-pituitary axis, several approximations were performed until find the best cost-effectiveness ratio (CER) –correlation of the net difference in the costs of two interventions to the net difference in their effectiveness–.

- First, the statistical evaluation of the algorithm was performed following the next steps:

- The initial sample –training set of 6,516 patients– was divided in the training and validation sets (80% of the total patients for the training set and the remaining 20% for the validation set), to determine an initial accuracy.
- All the data was pre-processed by converting those numeric variables that are categorical.
- The next step consisted in visualizing the logistic dependent variables and all the different qualitative variables, to verify if the distribution was balanced or not, and if necessary, a corrective method was applied to adjust the unbalance of the classes, by modifying the original size of the whole training data.
- The absent cases were detected, and an imputation treatment was implemented, either with the median or not, –with the most frequent cut-off value that indicates the abnormality of the variable, and the median was used as a replacement value for those observations that were above the cut-off–.
- The initial binary logistic regression (logit) was estimated by the general linear model (GLM) algorithm, with the argument *family = binomial* (using the *logit()* function) because the dependent variable is binary categorical, and the threshold to classify the attribute depended on what we wanted to predict (a priori) the cut-off point was 0.5, because it is the standard cut-off to classify as healthy and sick).
- The logit model achieved was evaluated through the following methods: assessment of the influential values and possibly atypical from the residues of the logit model; multicollinearity analysis –to evaluate the presence and the magnitude of strong linear relationships between predictive variables (independent variables) in the model–; goodness of fit –to determine if the model is valid and adequate for its use in decision making or in making predictions–; calculation of the importance of the predictive variables in the model, considering their weight through the decreasing of the mean average precision and the AUCi decreasing average; as well as the final validation for the model with the validation of the 20%, to determine the sensitivity, specificity, area under the receiver operating characteristic (AUROC) curve, positive predictive value (PPV), and negative predictive value (NPV).
- The cut-off point was optimized, to finally adjust the binary logistic regression model.
- The final evaluation of the binary logistic regression model was performed with the optimal cut-off point.

- Second, several combinations –up to 1 x 10¹⁰–, were performed to find most significant groupings of laboratory determinations –mainly for the thyroid function, but also for all other body functions and systems involved and/or related with this, is both causes and/or consequences–.
- Third, several calculations were performed, mainly those related with thyroid ratios (Table 4). We selected the ones that were Evidence-Based Laboratory Medicine (EBLM). In this way, we performed thyroid ratios were free T4 hormone-to-thyroid stimulating hormone (FT4-to-TSH) ratio³², and (FT4-to-TSH) ratio as a function of gender and age³³.

On the other hand, all patient data was computed with two machine learning (ML) algorithms, such as Evidence-Based Laboratory Medicine Algorithm (EBLM)²⁸ and Artificial Intelligence Recursive Algorithm (AIRA)³⁴ –both developed by Blueberry Diagnostics (Barcelona, Spain) in 2020 to help in COVID-19 diagnosis–, to improve both sensitivity and specificity. EBLM and AIRA include several functions –as algorithms– to assess the following conditions: hypothyroidism (Hypo) and subclinical hypothyroidism (sHypo), hyperthyroidism (Hyper) and subclinical hyperthyroidism (sHyper); and also, the primary and secondary subtypes of these diseases. Besides, AIRA is an algorithm that computes in a recursive way all the values from 0 to 100% for both sensitivity and specificity to find the optimal cut-off values for low, moderate (moderate-low and moderate-high), and high-risk results. Then, parallel approximations to optimize sensitivity (Se) were conducted. Afterwards, serial approximations to enhance specificity (Sp) were also performed. Finally, Cost Effectiveness Ratio Effect Boosting Recursive Optimizer (CEBRO)³⁵ –developed by Kience (Wilmington, DE, U.S.)–, adjusts resulting predictive values –the p-values– with the cost-effectiveness ratio (CER) –correlation of the net difference in the costs of two interventions to the net difference in their effectiveness–.

The training set was very heterogeneous, as it included results analyzed by different suppliers for the same laboratory determinations –Synnex, Horiba, Roche Diagnostics, Siemens Healthineers, and others–, with their own systems and reference values. So, the variability and the potential biases of these results obtained by this way may be too high, thus affecting the quality of the results. In this way, to avoid these potential biases that can alter the results –and consequently, the accuracy of the algorithm–, the present study was designed with one single laboratory –Laboratorio Echevarne (San Cugat del Valles, Barcelona, Spain)–, and consisted in a randomized controlled trial (RCT) with a target sample size (n) of 300.

All patients included in the RCT were recruited through three medical centers in Barcelona (Spain). Patients had to accept and sign the informed consent, the ethics committee approval, and the test requisition form (TRF) –an example of the new TRF designed for the upcoming RCT that will be performed in the U.S. with Empire City Laboratories (Brooklyn, NY, U.S.) is shown (Figure 4)–. This TRF corresponds to the MCEd TRF³¹ –with 90 laboratory determinations included–, from which this other study was developed –by using only 2 routine laboratory determinations of the total analyzed–. In the case of the present study, although target sample size was 300 patients, we decided to enroll 334 patients to allow some backup. However, according to the inclusion and exclusion criteria (Figure 3), 50 patients were initially excluded from the study, because they had ongoing clinically diagnosed pathologies, symptoms or signs, so the sample size dropped to 284 patients. From this new study population, 110 patients were excluded because they didn't show up at the clinical facility for any of the follow-up visits –this is a critical point to improve for the upcoming RCT, since the cost of each patient is very high and for this RCT almost a third of the patients already tested were lost–, so the new n consisted of 154 patients. From this new n, 2 patients were excluded because some of their laboratory parameters and/or clinical information were wrong or incorrect. Thus, a final n of 152 patients was achieved. In this final n, both genders (male and female) were equally represented and the mean age of the participants was 53.34 years, being 54.64 (40 – 82) in the male population and 52.03 (41 – 77) in the female population (Tables 1–5).

Patients' blood samples were obtained from October 2021 to June 2023. Blood samples were obtained by peripheral venipuncture in all participants. All analyses were performed in Laboratorio Echevarne. After centrifugation of the blood samples, all the analytes were quantified. The serum hormones were measured by Atellica analyzers (Siemens, Munich, Germany). The majority of the tests were performed within two days after samples were obtained. This is important to note that the laboratory where the RCT was performed did not have the necessary equipment for the analysis of the thyroid and pituitary hormones of the RCT –so, false positives (FP) and false negatives (FN) could occur because most major studies were conducted in hospitals using these equipments. Thus, the present study could have an additional bias due to the difference between the analyzers used in the modeling process of the algorithm and the RCT. In this way, although several previous correlation studies alerted about potentially moderate differences in the performance between different reagent vendors –as mentioned above–, the reason why the validation of the algorithm was done in a lab with no Roche Diagnostics analyzers was to validate also its robustness and overall performance against different vendors. Therefore, in the upcoming clinical trial of 26,000 patients, several parallel studies of 120 patients will be performed by analysing blood and/or urine samples with different vendors to achieve correlation coefficients and curves, in order to further evaluate their overall robustness.

Once the results were obtained, all those that exhibited abnormal values –outside their reference ranges– were re-processed, to make sure that they were not obtained due to technical errors and were real and potentially pathogenic. When all laboratory determinations were finished per each patient, they were processed by the latest version of the above-mentioned algorithm giving their final result as well as determining the final accuracy with a validation set of 152 patients. In turn, all patients with suspicious findings were referred to the corresponding medical centers for follow-up and subsequent classification in each of the groups –case and control–, designed for the biostatistics upcoming phase. Patients attended follow-up visits in one year.

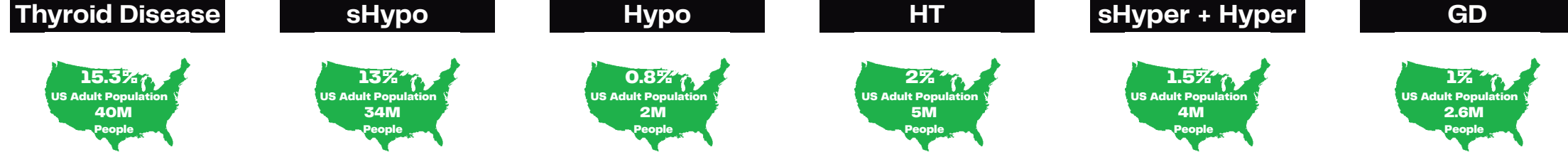


Figure 2. Prevalences of main thyroid disease and its types –hypothyroidism (Hypo), subclinical hypothyroidism (sHypo), hyperthyroidism (Hyper), and subclinical hyperthyroidism (sHyper)–, in the U.S., as well as their main autoimmune causes –Hashimoto thyroiditis (HT) and Graves' disease (GD)–.

Inclusion Criteria

- Men and women with not known currently clinically diagnosed pathologies, without symptoms or signs, white/Caucasian ethnicity, aged 40, onwards; and
- Patients whose high anthropometric indicators and/or lifestyle habits may predispose them to non-malignant, highly prevalent, morbid, cancer-precursor, and deadly diseases (Figure 2).
 - Body mass index (BMI), waist-to-hip ratio (WHR), waist-to-height ratio (WtHR), hemoglobin/glycemic waist (HW), visceral adiposity index (VAI) –which are indicative of overweight and obesity–; and/or
 - High reprocessed meat consumption, low fruits and vegetables intakes, lack of physical activity, smoking habits, and/or alcohol intake.

Exclusion Criteria

- Patients who didn't meet the inclusion criteria or meet any of the following criteria:
 - Patients with any confirmed or found cancer during the whole study.
 - Patients that didn't show up at the clinical facility in any of the follow-up visits.
 - Patients lacking any of the laboratory parameters and/or clinical information.

Figure 3. Inclusion and exclusion criteria for the selection of the study population, as well as the graphical flow of the patients that were selected in the study population.

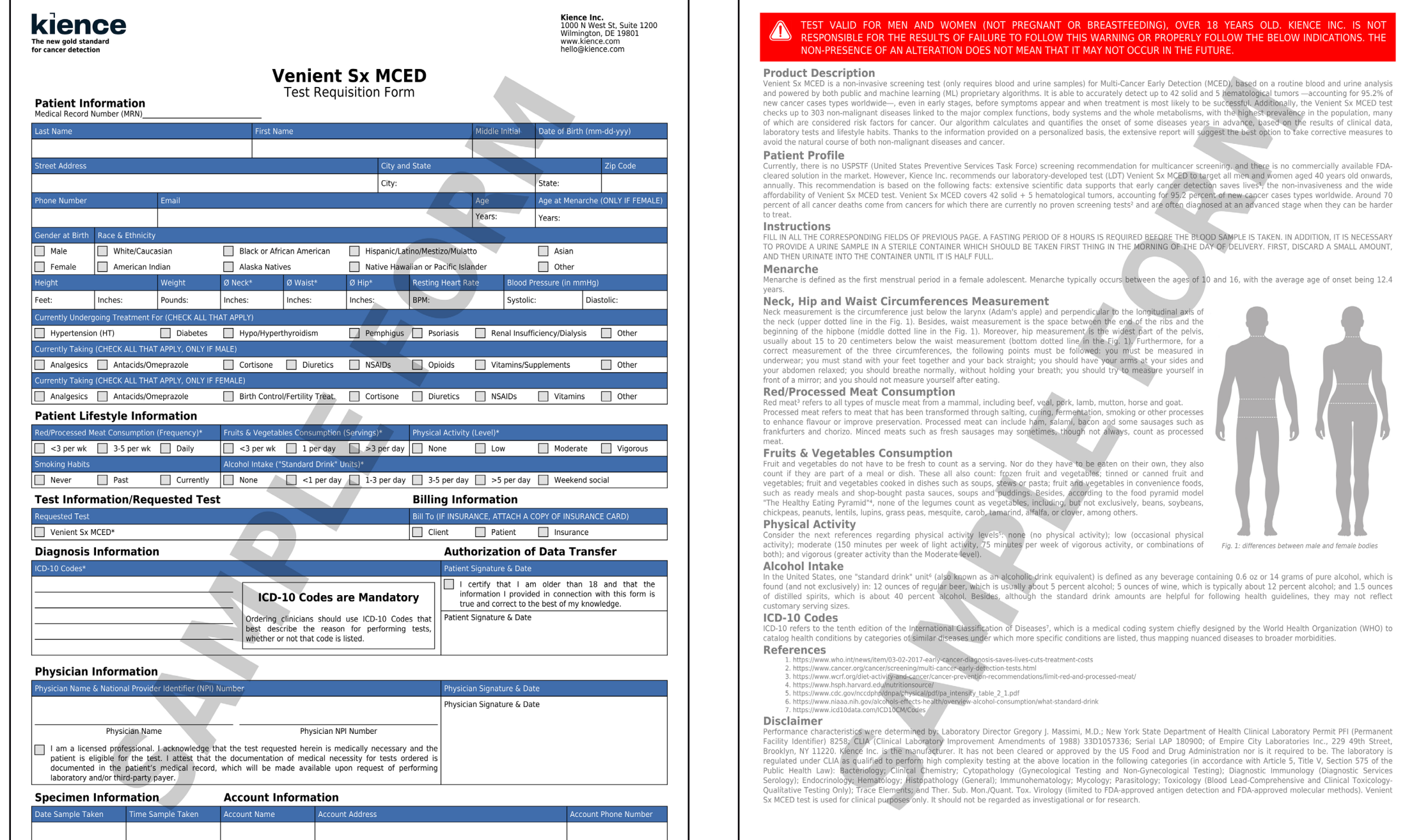


Figure 4. Test requisition form (TRF) for the RCT to be performed in the U.S. (Empire City Laboratories Inc., Brooklyn, NY, U.S.). This new RCT will be based in a sample size (n) of 1,000 patients –paying special attention to the proportion of different ethnicities of the population of the U.S.–.

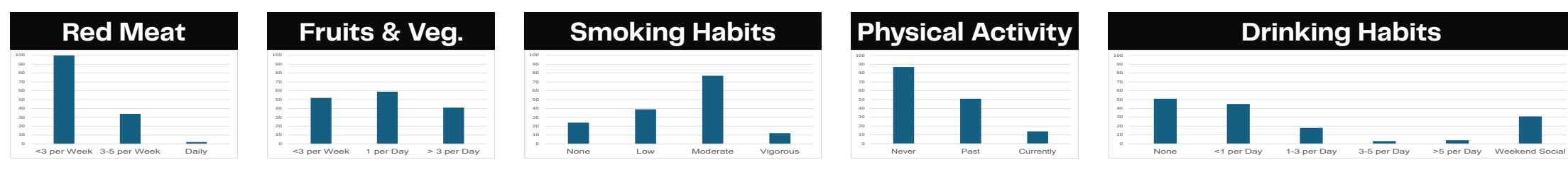


Figure 5. Lifestyle –red and processed meat consumption (displayed as frequency), fruits and vegetables consumption (displayed as serving), physical activity (displayed as level), smoking habit, and drinking habits (displayed as standard drink units)–, for all patients.

Parameter	Gender	Min	Max	Female
Age (years)	Female	43.01 (31.86)	82.00 (82.00)	52.03 (37.22)
Height (cm)	Female	150.71 (150.14)	176.77 (150.14)	163.51 (149.17)
Weight (kg)	Female	54.64 (38.28)	100.00 (47.00)	68.33 (35.13)
Neck Circumference (cm)	Female	36.08 (36.08)	40.00 (39.40)	39.79 (36.59)
Waist Circumference (cm)	Female	66.97 (52.13)	96.71 (66.97)	76.62 (52.13)
Body Mass Index (BMI)	Female	23.79 (19.12)	39.46 (23.19)	26.41 (22.77)
Waist-Hip Ratio (WHR)	Female	0.91 (0.79)	1.17 (0.79)	1.00 (0.79)
Waist-Height Ratio (WtHR)	Female	0.38 (0.38)	0.50 (0.38)	0.46 (0.38)
Hemoglobin (g/dL)	Female	11.6 (10.7)	17.0	13.27 (10.7)
Glucose (mg/dL)	Female	73.66 (143.130)	77.03 (143.130)	70.09 (143.130)
Diastolic Pressure (mmHg)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)
Triglyceride (mg/dL)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)

Table 1. Clinical data, displayed as the mean value of each parameter with the corresponding minimal (Min) and maximal (Max) value, for all patients and by gender: male and female.

Parameter	Gender	Min	Max	Female
Age (years)	Female	43.01 (31.86)	82.00 (82.00)	52.03 (37.22)
Height (cm)	Female	150.71 (150.14)	176.77 (150.14)	163.51 (149.17)
Weight (kg)	Female	54.64 (38.28)	100.00 (47.00)	68.33 (35.13)
Neck Circumference (cm)	Female	36.08 (36.08)	40.00 (39.40)	39.79 (36.59)
Waist Circumference (cm)	Female	66.97 (52.13)	96.71 (66.97)	76.62 (52.13)
Body Mass Index (BMI)	Female	23.79 (19.12)	39.46 (23.19)	26.41 (22.77)
Waist-Hip Ratio (WHR)	Female	0.91 (0.79)	1.17 (0.79)	1.00 (0.79)
Waist-Height Ratio (WtHR)	Female	0.38 (0.38)	0.50 (0.38)	0.46 (0.38)
Hemoglobin (g/dL)	Female	11.6 (10.7)	17.0	13.27 (10.7)
Glucose (mg/dL)	Female	73.66 (143.130)	77.03 (143.130)	70.09 (143.130)
Diastolic Pressure (mmHg)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)
Triglyceride (mg/dL)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)

Table 2. Anthropometric indices and ratios, displayed as the mean value of each parameter with the corresponding minimal (Min) and maximal (Max) value, for all patients and by gender: male and female.

Parameter	Gender	Min	Max	Female
Age (years)	Female	43.01 (31.86)	82.00 (82.00)	52.03 (37.22)
Height (cm)	Female	150.71 (150.14)	176.77 (150.14)	163.51 (149.17)
Weight (kg)	Female	54.64 (38.28)	100.00 (47.00)	68.33 (35.13)
Neck Circumference (cm)	Female	36.08 (36.08)	40.00 (39.40)	39.79 (36.59)
Waist Circumference (cm)	Female	66.97 (52.13)	96.71 (66.97)	76.62 (52.13)
Body Mass Index (BMI)	Female	23.79 (19.12)	39.46 (23.19)	26.41 (22.77)
Waist-Hip Ratio (WHR)	Female	0.91 (0.79)	1.17 (0.79)	1.00 (0.79)
Waist-Height Ratio (WtHR)	Female	0.38 (0.38)	0.50 (0.38)	0.46 (0.38)
Hemoglobin (g/dL)	Female	11.6 (10.7)	17.0	13.27 (10.7)
Glucose (mg/dL)	Female	73.66 (143.130)	77.03 (143.130)	70.09 (143.130)
Diastolic Pressure (mmHg)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)
Triglyceride (mg/dL)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)

Table 3. Selected laboratory data, displayed as the mean value of each parameter with the corresponding minimal (Min) and maximal (Max) value, for all patients and by gender: male and female.

Parameter	Gender	Min	Max	Female
Age (years)	Female	43.01 (31.86)	82.00 (82.00)	52.03 (37.22)
Height (cm)	Female	150.71 (150.14)	176.77 (150.14)	163.51 (149.17)
Weight (kg)	Female	54.64 (38.28)	100.00 (47.00)	68.33 (35.13)
Neck Circumference (cm)	Female	36.08 (36.08)	40.00 (39.40)	39.79 (36.59)
Waist Circumference (cm)	Female	66.97 (52.13)	96.71 (66.97)	76.62 (52.13)
Body Mass Index (BMI)	Female	23.79 (19.12)	39.46 (23.19)	26.41 (22.77)
Waist-Hip Ratio (WHR)	Female	0.91 (0.79)	1.17 (0.79)	1.00 (0.79)
Waist-Height Ratio (WtHR)	Female	0.38 (0.38)	0.50 (0.38)	0.46 (0.38)
Hemoglobin (g/dL)	Female	11.6 (10.7)	17.0	13.27 (10.7)
Glucose (mg/dL)	Female	73.66 (143.130)	77.03 (143.130)	70.09 (143.130)
Diastolic Pressure (mmHg)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)
Triglyceride (mg/dL)	Female	69.11 (62.110)	69.23 (62.110)	69.09 (62.110)

Figure 6. Sample graphic for the thyroid function basic assessment shared with healthcare professionals to validate the usefulness of this new approach to a routine blood test –diagnostic accuracy, non-invasiveness, time savings, cost savings, turnaround time for results, ease of interpretation of results, or savings on additional complementary tests, among others–. The graphic for the thyroid function shows a white dot corresponding the thyroid stimulating hormone (TSH) –plotted on the X-axis– and the free T4 (FT4) hormone –plotted on the Y-axis–, over a colored background (if any value is greater or smaller than X-axis or Y-axis ranges of the graphic, the dot is colored in red and placed over the corresponding border). The background is colored according what these levels mean once putted together in order healthcare professionals could figure in an easy way how is the patient's thyroid function according current worldwide guidelines that is, by using a colored scale based on a risk stratification from green (Risk 0) to dark red (Risk 4).

Table 5. Raw laboratory data, displayed as the mean value of each parameter with the corresponding minimal (Min) and maximal (Max) value, for all patients and by gender: male and female. The data for this study comes from the results obtained from a few of the thyroid laboratory determinations (see selected laboratory determinations in Table 3), since the main approach of this study is to generate the cheapest and universally useful panel possible for as many of these diseases as possible.

Urine leukocytes, urine nitrites, urine protein, urine glucose, urine ketones, urine urobilinogen, urine urobilin, urine red blood cells and urine hemoglobin are expressed as a for "no/none", 1 for "positive +", 2 for "positive ++" and 3 for "positive +++".

Acknowledgments: This study was funded by Bioprognos and Blueberry Diagnostics. Medical writing support was provided by i2e3 Proccoms team. We would also like to thank several doctors for their insights in different fields of medicine that have helped, directly or indirectly, to achieve the results of this publication (for some of them it was not possible to include more references to their work because space reasons): Rafael Molina¹, Jaime Trapé, as well as the members of the Spanish Society of Clinical Chemistry and Laboratory Medicine (SECCM) –Barcelona Office–, Juan Yáñez –endocrinologic and metabolic diseases–, Sabine Kahn –hepatic diseases–, Matteo Ferro, Gabriele Avcioglu, Noha Mohamed Said –inflammatory indines–, Markus Kettler and Mark B. Leonard –kidney disease improving global outcomes (KDIGO) guidelines–, Andrew S. Levey –Chronic kidney disease-epidemiology collaboration (CKD-EPI) equation–, Mario Plebani –evidence-based laboratory medicine (EBLM), value-based medicine (VBM) and value-based laboratory medicine (VBLM), laboratory medicine and integrated diagnostic–.

Disclosures: José D. Santotoribio has no conflicts of interest to declare. Sergio J. Calleja is a shareholder of Kience and Blueberry Diagnostics. Sergio J. Calleja has financial interests in Bioprognos.

Correspondence: sergio.calleja@kience.com; adria.roca@blueberrydx.com; jose.d.santotoribio.sspa@juntadeandalucia.es.

RESULTS

This new approach to a routine blood test has detected all the cases of thyroid disease –hypothyroidism and hypothyroidism– of primary and secondary origin, in the randomized controlled trial (RCT) validation set, as can be seen in Table 5. It could be

Given that the prevalence of *Helicobacter pylori* infection in the raw laboratory data is very low (Table 5) and it could be an indicative that the study population was biased. This is because the overall prevalence in Spain is much higher since the one achieved in this study³⁶. Since *H. pylori* infection is more frequent in people with lower income³⁷ this low prevalence in the validation group suggests that the study population is principally composed by middle-to-high class patients, with moderate-to-high income.

In Table 6, the results are displayed for the following groups: the training set was tested for all 2 laboratory determinations, but with different providers with their own reference limits and the validation set was tested for the same 2 laboratory determinations, but all of them were performed by the same providers. The sample size (n) of the training set was 6,516 patients, but this uniformity of the providers resulted in a reduction of the sample size, because only those patients tested in Laboratorio Echevarne were used in the validation set, and for this reason, the sample size dropped to 152 patients.

In training set, the sensitivity (Se) and specificity (Sp) obtained were 94.82% and 95.76%, respectively. The estimated area under the receiver operating characteristic (AUROC) curve was 95.30%, and the positive predictive value (PPV) and the negative predictive value (NPV) were 97.65% and 90.81%, respectively (Table 6).

Finally, in validation set, it can be seen that false positives (FP) and false negatives (FN) were totally eliminated, because the Se and Sp, as well as the PPV and the NPV, increased to 100.00%. But it must be considered that the sample size (n) is very low (152 patients), and it is also very biased, because all patients were 40 years old and older, white/Caucasian ethnicity, and with moderate-to-high income.

CONCLUSIONS

This innovative non-invasive blood-based biomarker algorithm holds promise in providing timely and accurate assistance to doctors in the basic assessment –as well as screening–, of thyroid diseases –hypothyroidism and hypothyroidism– of primary and secondary origin –even in early stages–, before symptoms and signs appear and when treatment is most likely to be successful³⁸. It is also possible to define the subtype of these diseases –primary or secondary–, depending on if the affection is in the thyroid gland or in the pituitary gland. In this process, the price should be a very important variable since this panel should be the cheapest one to enable universal and quality access to healthcare.

- This RCT shows the achievement of a routine, affordable –accessible to low-income and underserved populations–, and high accurate blood test. The panel maximizes diagnostic yield while minimizing costs. The selected analytes can detect early signs of thyroid diseases, without requiring expensive or invasive procedures. Besides, by focusing on high-prevalence diseases and selecting analytes that are both low-cost and available, this panel is particularly feasible on a large scale. Automated laboratory systems for blood tests help reduce labor costs and increase throughput, further driving down overall costs.
- Given the concerning problem about the healthcare costs in the U.S. –but also worldwide–, and its access for a considerable portion of its population, this algorithmic test can also be a solution for this, costing only EUR 1.80 (USD 1.97) for Roche Diagnostics systems for immunoassay.
- By addressing the most common and preventable diseases at a fraction of the cost of traditional healthcare models, this initiative aligns with global health goals such as universal health coverage (UHC)³⁹ and the Sustainable Development Goals (SDG)⁴⁰, particularly those focused on reducing premature mortality and morbidity from thyroid diseases⁴¹.
- As the other hard-to-treat and/or chronic diseases, the need to encourage healthcare professionals to explore how Evidence-Based Laboratory Medicine (EBLM)²⁸ and new technologies –mainly machine learning (ML) algorithms–, can help them to improve diagnosis accuracy, reduce medical errors