



HAL
open science

Machine Learning Based Algorithm for the Early Prediction of Postoperative Hypocalcemia Risk After Thyroidectomy.

Olivier Muller, Pierre Bauvin, Ophélie Bacoer-Ouzillou, Théo Michailos, Maria-Vittoria Bertoni, Charles Demory, Camille Marciniak, Mikael Chetboun, Gregory Baud, Marco Raffaelli, et al.

► **To cite this version:**

Olivier Muller, Pierre Bauvin, Ophélie Bacoer-Ouzillou, Théo Michailos, Maria-Vittoria Bertoni, et al.. Machine Learning Based Algorithm for the Early Prediction of Postoperative Hypocalcemia Risk After Thyroidectomy.. *Annals of Surgery*, 2024, *Annals of Surgery*, Online ahead of print. 10.1097/SLA.0000000000006480 . hal-04706416

HAL Id: hal-04706416

<https://hal.univ-lille.fr/hal-04706416v1>

Submitted on 23 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Machine learning based algorithm for the early prediction of postoperative hypocalcemia risk after thyroidectomy

Olivier MULLER M.D*(1), Pierre BAUVIN Ph.D*(2), Ophélie BACOEUR M.D. (1), Théo MICHAÏLOS M.D. (1), Maria-Vittoria BERTONI (1), Charles DEMORY (1), Camille MARCINIAK M.D. (1,2), Mikael CHETBOUN M.D. (1,2), Grégory BAUD M.D. (1,2), Marco RAFFAELLI (3), Robert CAIAZZO M.D. (.,2), Francois PATTOU M.D (1,2)

*Equally contributed

- (1) Department of General and Endocrine Surgery, CHU Lille, Lille, France
- (2) UMR 1190, Univ Lille, Inserm, Institut Pasteur de Lille, CHU Lille, Lille, France
- (3) Endocrine Surgery, Department of Surgery, Università Cattolica della Sacro Cuore, Roma, Italy

Corresponding author : francois.pattou@univ-lille.fr

Funding:

No financial nor technical support was used for this work.

Keywords:

Total thyroidectomy, hypocalcemia, hypoparathyroidism, machine learning, prediction, parathormone, PTH, precision medicine

ABSTRACT:

Objective

We used machine learning to develop and validate a multivariable algorithm allowing the accurate and early prediction of postoperative hypocalcemia risk.

Summary background

Post-operative hypocalcemia is frequent after total thyroidectomy. An early and accurate individualized prediction of the risk of hypocalcemia could guide the selective prescription of calcium supplementation only to patients most likely to present with hypocalcemia after total thyroidectomy.

Methods

This retrospective study enrolled all patients undergoing total thyroidectomy in a single referral center between November 2019 and March 2022 (derivation cohort) and April 2022 and September 2022 (validation cohort). The primary study outcome was post-operative hypocalcemia (serum calcium under 80 mg/L). Exposures were multiple clinical and biological variables prospectively collected and analyzed with various machine learning methods, to develop and validate a multi variable prediction algorithm.

Results

Among 610 / 118 participants in the derivation / validation cohorts, 100 (16.4%) / 26 (22%) presented post-operative hypocalcemia. The most accurate prediction algorithm was obtained with random forest, and combined intraoperative parathyroid hormone measurements with three clinical variables (age, sex and body mass index), to calculate a postoperative hypocalcemia risk for each patient. After multiple cross validation, the area under the receiver operative characteristic curve was 0.902 (0.829-0.970) in the derivation cohort, and 0.928 (95% CI : 0.86; 0.97) in the validation cohort. Postoperative hypocalcemia risk values of 7% (low threshold) and 20% (high threshold) had respectively a sensitivity of 92% and a negative likelihood ratio of 0.11, and a specificity of 90% and a positive of 7.6 for the prediction of postoperative hypocalcemia.

Conclusion.

Using machine learning, we developed and validated a simple multivariable model which

allowed the accurate prediction of postoperative hypocalcemia. The resulting algorithm could be used at the point of care to guide clinical management after total thyroidectomy.

ACCEPTED

INTRODUCTION

Post-operative hypocalcemia is frequent after total thyroidectomy (TT)¹. In most cases, hypocalcemia results from hypoparathyroidism secondary to the peroperative devascularization of the parathyroid glands. Its occurrence can be associated with severe clinical symptoms, requiring the need for calcium supplementation², and prolonged hospitalization. Routine post-operative calcium supplementation after total thyroidectomy has been proposed by some authors to reduce the burden of hypocalcemia^{3,4}. This practice is however not currently recommended in international guidelines, as it may mask the onset of severe, long-lasting hypocalcemia, and delay the diagnosis of definitive hypoparathyroidism⁵

An early and precise prediction tool could help to guide the prescription of calcium supplementation to patients most likely to present with postoperative hypocalcemia. This is particularly useful in the context of outpatient thyroid surgery⁶, which requires minimization of potential complications.

There is currently no consensus on the best predictor of postoperative hypocalcemia. Several pre- and per-operative clinical and biological variables have been associated with the risk of postoperative hypocalcemia. Among them, immediate postoperative serum intact parathyroid hormone (PTH) level has recently gained a lot of attention^{7,8}. PTH is biochemically directly associated to blood calcium level and its short half-life, provides an early insight into the patient's postoperative parathyroid function.

Numerous studies have analyzed PTH serum levels at different time points following thyroidectomy to predict the onset of postoperative hypocalcemia, from skin closure in the operating room to the first postoperative day^{9,10}. There appears to be no clear difference in the prediction values of these different time points^{11–13}. The percentage decay between preoperative and postoperative levels of serum PTH (%PTH) has been also proposed^{8,14,15}, and reported to correlate linearly with postoperative calcium levels¹⁶. Furthermore, the use of immediate postoperative serum PTH would have the advantage of easy implementation for the early initiation of calcium and vitamin supplementation therapy.

However, even if PTH seems to be the most accurate predictive factor, the potential of the combination of PTH with other predictive factors to reinforce the accuracy of prediction is unclear. Recent methods of artificial intelligence such as machine learning, could be useful to identify such favorable combinations.

In this study, we leveraged machine learning to identify the most accurate predictive factors for postoperative hypocalcemia after total thyroidectomy and define an early surrogate marker of the risk of developing postoperative hypocalcemia after total thyroidectomy.

MATERIALS AND METHODS

1. Population and preoperative data collection

In this retrospective study, we enrolled every patient who underwent a total thyroidectomy at Lille University Hospital, first between 1st November 2019 and 31 March 2022 (derivation cohort), and second between 1st april 2022 and 13 sept 2022 (validation cohort), in absence of associated hyperparathyroidism and/or planned parathyroidectomy.

Preoperative data consisted of demographic data (age, sex, height, weight, and body mass index (BMI)), surgical indication, history of surgical neck dissection or hyperthyroidism, and biological data (serum protein or albumin level, serum creatinine level and serum vitamin D level before supplementation). History of vitamin D supplementation was also collected.

2. Surgery, PTH dosage protocol and intraoperative data

Thyroidectomy was performed or supervised by 5 experimented senior surgeons, following codified dissection which aimed at formal identification and in situ preservation of superior parathyroid glands, as well as at preservation of the vascularization of all four parathyroid glands. Recurrent nerve function was monitored via endotracheal neuromonitoring (NIM), and confirmed, when necessary, by postoperative laryngoscopy. Every patient had a serum PTH dosage at the time of anesthetic induction (T0) and at the time of skin closure once total thyroidectomy was performed (T1). PTH serum levels were considered null when not detectable (inferior detection threshold 8 pg/mL).

Intraoperative data consisted of operating time, performance of a central neck dissection, number of parathyroid glands identified, number of parathyroid glands reimplemented, and the weight of the thyroid specimen.

3. Postoperative management, postoperative hypocalcemia definition and exclusion criteria

Postoperative management consisted of a serum level dosage of calcium, phosphorus and PTH at postoperative day (POD) 1 of surgery. Discharge was allowed at POD 1 when serum calcium was ≥ 80 mg/dL and POD 1 PTH was above 8 pg/mL, in absence of clinical

symptoms of hypocalcemia (paresthesia, myoclonia) or other surgical complication (dysphonia, dysphagia, or dyspnea).

When these criteria were not met, patients remained hospitalized, and serum phosphorus, calcium and PTH levels were evaluated daily, or in case of symptoms consistent with hypocalcemia. Postoperative hypocalcemia was defined by a serum calcium level inferior to 80 mg/L, measured at any postoperative time. Patients were excluded from the analysis if they received calcium and/or vitamin D in absence of biological evidence for hypocalcemia, regardless of PTH serum levels or symptoms occurrence.

All patients experiencing postoperative hypocalcemia were seen at one month after thyroidectomy, for serum calcium and PTH levels evaluation, 5 days after the interruption of supplementation. Supplementation treatment was suspended when calcium levels were normal ($>80\text{mg/L}$) in absence of hypocalcemia symptoms.

4. *Univariate statistical analysis and postoperative hypocalcemia prediction*

Qualitative variables were expressed as numbers and percentages of the study population, and quantitative variables were expressed as mean and standard deviation. Statistical comparison was performed between patients with or without postoperative hypocalcemia, using a Mann-Whitney test for quantitative variables and a Chi-2 test for qualitative variables. The predictive performance of significantly different variables was evaluated by the area under the receiver operator characteristic curve (AUROC curve), sensitivity, specificity, and positive and negative likelihood ratios for predicting the onset of postoperative hypocalcemia.

5. *Identification of predictive variables of interest and multivariate hypocalcemia prediction via machine learning*

We first conducted a multivariable analysis in the derivation cohort for identifying predictive variables of interest among the following candidate variables: sex, age, BMI, surgical indication, preoperative serum vitamin D level, preoperative serum creatinine, history of hyperthyroidism, operating time, number of parathyroids seen, parathyroid reimplantation, thyroid weight, central neck dissection, serum PTH at T0, PTH T1, and %PTH. Squared terms, as well as interaction terms, were added for the a priori most important variables: %PTH, neck dissection, surgical indication, sex, age.

We used a repeated stratified k-fold cross-validation ($k=10$, $n=100$) on our center data to find the best model, according to the AUROC metric, among six different mathematical models. Notably, cross-validation is a method to evaluate a model by partitioning the total sample into

a training set to train it, and a validation set to evaluate it. The six candidate models were linear model with %PTH only, logistic regression with or without l1-regularization, decision tree, random forest, and neural network. We specifically choose a heterogeneous set of candidate models: logistic regressions exploiting the linearity assumption, tree-based models learning simple decision rules to distinguish meaningful subgroups of patients, and neural networks using the Multi-layer Perceptron to differentiate data that is not linearly separable. All models included a first step of feature selection using Least Absolute Shrinkage and Selection Operator (LASSO), to extract the most statistically relevant variables concerning outcome prediction. All models' hyperparameters were tuned using a nested cross-validation grid search approach. Finally, the selected model corresponded to the best model from the previous benchmark, retrained on the whole dataset to increase precision of predictions. It is used to find clinically relevant levels of sensitivity and specificity¹⁷.

The feature importance was calculated as the mean accumulation of the impurity decreases. The two chosen thresholds were probability thresholds of postoperative hypocalcemia risk, expressed as percentages, estimated from clinically relevant levels of sensitivity and specificity. Analysis was performed using python 3.10 and the Scikit-learn library¹⁸. To ensure the reproducibility of our findings, the predictive performances of the selected model were eventually measured in the validation cohort.

6. Ethics statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review board and the register of data analysis of the University Hospital of Lille.

RESULTS

The derivation cohort gathered 646 patients who underwent a total thyroidectomy in our center during the study period. Among them, 36 patients (5,6%) were excluded from the analysis, either because they were supplemented for symptoms without biological hypocalcemia (16 patients), for low serum levels of PTH without biological hypocalcemia (17 patients), or abusive supplementation (3 patients). Among the 610 patients enrolled in the present study, 100 (16,4%) presented with hypocalcemia. Overall 9 patients (1.5% of the overall population and 9% of patients with postoperative hypocalcemia) presented with persistent hypocalcemia, i.e ; still requiring calcium and/or vit D treatment 6 months after surgery.

Table 1 presents demographic, preoperative, intraoperative, and postoperative characteristics of patients with or without postoperative hypocalcemia. Hypocalcemia patients were significantly younger (median age 44.5 years vs 52 years, $p=0.0006$), more likely to be a female (85% vs 73.3%, $p=0.015$) and had a significantly lower BMI (24.2 vs 27.6, $p=0.001$). Both groups were comparable in terms of preoperative history, surgical indications, preoperative biological data and intraoperative data. Total thyroidectomy was indicated in 219 patients for a suspicious node, for the presence of a Bethesda score of 5 or 6 at cytoponction, and/or a size above 4 cm. Malignancy was confirmed at pathology in 181 cases (82.6%).

Postoperative data analysis showed that median values of serum PTH at skin closure, %PTH and PTH at POD 1 were significantly lower in hypocalcemia group (0 pg/ml vs 26 pg/ml, 100% vs 51.4%, and 5 pg/ml vs 26 pg/ml respectively, $p<0.0001$). Serum calcium level at POD 1 was also significantly lower in hypocalcemia group (80 mg/L vs 88 mg/L, $p<0.0001$). Hospital stay was significantly higher in hypocalcemia group (3 days vs 1 day, $p<0.0001$).

The validation cohort gathered 118 patients operated during the second study period, including 26 (22.0 %) individuals who presented postoperative hypocalcemia. Patient characteristics at baseline were similar to the derivation cohort (supplemental Table 1, Supplemental Digital Content 1, <http://links.lww.com/SLA/F251>).

As shown in **figure 1**, the AUROC curve of %PTH for predicting postoperative hypocalcemia in the derivation cohort was 0.901 (95% confidence interval 0.863-0.939, $p<0.0001$). We also evaluated the performances of %PTH to predict the risk of postoperative hypocalcemia according to three threshold values (**Table 2**). A %PTH superior to 60% had a sensitivity of 92 % (CI: 85%-96%) and a specificity of 61% (CI: 57%-65.3%) for hypocalcemia prediction, with a positive likelihood ratio (LR+) of 2.5 and a negative likelihood ratio (LR-) of 0.14. A %PTH superior to 75% had a sensitivity of 82% (CI: 73.3% - 88.3%) and a specificity of 81.6% (CI: 78.0% - 84.7%) for hypocalcemia prediction, with a LR+ of 4.45 and a LR- of 0.22. A %PTH superior to 85% had a sensitivity of 74% (CI: 64.6% - 81.6%) and a specificity of 90% (CI: 87.1% - 92.3%) for hypocalcemia prediction, with a LR+ of 7.4 and a LR- of 0.29.

Table 3 shows the performances of the six models tested for multivariable analysis. The best model according to AUROC curves, evaluated on repeated cross-validation, was the model with feature selection followed by random forest. This model achieves an AUROC of 0.902

(CI: 0.829-0.970), versus 0.889 (CI: 0.806-0.966) for the %PTH alone model. The AUROC curves of both models are displayed in **Figure 1**, using the chosen model retrained on the whole dataset.

Among all variables included in the model, only 5 selected variables were identified as predictive factors for hypocalcemia: %PTH squared, PTH at anesthetic induction, interaction between %PTH and sex, BMI, and age squared. These variables are reported in Table 4, by order of feature importance. Noteworthy, feature importance is only relevant for the model itself and its value cannot be used to tell if a given feature has a negative or positive impact on postoperative hypocalcemia occurrence. This five variables model was used to determine for each patient of the dataset, a value of postoperative hypocalcemia risk (PHR), expressed in percent. We then identified two thresholds of PHR which were associated with clinically useful levels ($\geq 90\%$) of sensitivity or specificity, respectively (**Table 5**). A PHR of 7% (low threshold) was associated with a sensitivity of 92% resulting in a LR- of 0.11 for the prediction of postoperative hypocalcemia. A PHR of 20% (high threshold) was associated with a specificity of 90, resulting in a LR+: 7.6 for postoperative hypocalcemia prediction. Supplemental Figure 1, Supplemental Digital Content 2, <http://links.lww.com/SLA/F252> illustrates the relation between postoperative hypocalcemia risk according to the random forest model (PHR) and %PTH for 4 virtual individual patients with distinct characteristics. Finally, the predictive performance of the selected random forest model were tested in the validation cohort, resulting in an AUROC curve value of 0.928 (95% CI : 0.86; 0.97), well within the confidence interval of the AUROC curve value estimated through repeated cross-validation in the initial dataset.

Eventually, the model's mathematical algorithm was implemented in a shinyapp, available online (**Figure 2**) and allowing the automated calculation of PHR of any given patient, at the point of care, according to the 5 variables of interest (<https://lille-model.shinyapps.io/Hypocal/>).

DISCUSSION

Predicting the risk of postoperative hypocalcemia after total thyroidectomy has been an important research question for several decades¹⁹. In the present study, we used machine learning to developed and validate a multivariable model allowing the early and accurate

prediction of postoperative hypocalcemia, based on age, sex, BMI, preoperative PTH, and %PTH.

Serum PTH has a 4 minutes half-life and directly reflects parathyroid function. The early measurement of this direct biomarker of parathyroid function has been increasingly proposed to predict hypocalcemia after total thyroidectomy^{27, 28}. In 2018, a literature review by Marthur and al. concluded that the best PTH based predictor of postoperative hypocalcemia was PTH decay between pre- and post-operative levels⁸. Furthermore, Liu and al. described a linear correlation between percentage of PTH decay and postoperative calcium levels¹⁶. In addition, other variables have been proposed to assess the risk of hypocalcemia^{16,21}, and improve prediction. This suggests that even if, at first, PTH appears as a near perfect biomarker for hypocalcemia prediction, there might be other independent and clinically relevant predictive factors. Here, we initially evaluated the performance of intraoperative PTH measurement alone to predict the onset of hypocalcemia after total thyroidectomy. We found that both preoperative PTH and %PTH measured immediately after surgery were independent predictors of postoperative hypocalcemia. In addition, our multivariable analysis also identified three other preoperative variables (age sex and body mass index) that were independently associated with the risk of postoperative hypocalcemia in our population. We then apply random forest, a machine learning method used for classification and regression models, operating by constructing a multitude of decision trees at training time. Using this method, we found that even if %PTH was a major predictive factor, prediction could be further improved when other variables of interest were added in the model. Following multiple cross validation, the algorithm resulted in a PHR score with an AUROC curve of 0.902 (95% CI 0.829; 0.970). We also identified two low and high thresholds of the PHR score with very good positive (>5) and negative (<0.2) likelihood ratios to predict hypocalcemia. The five simple variables selected being all available, immediately after surgery, the resulting model has been incorporated in an online shinyapp that can be used at the point of care, to guide early postoperative clinical management.

Several studies have also shown the impact of sex^{22–24} and age²⁴ as a predictive factor for postoperative hypocalcemia. In line with our results, two studies^{29,30} have previously identified higher body mass index as a potential protective factor for transient hypocalcemia after thyroidectomy. Furthermore, serum PTH level has been shown to be linked with obesity in the general population³¹, providing a potential explanation of this apparent "obesity paradox" for postoperative hypocalcemia after total thyroidectomy. Of note, other variables,

which have been previously associated with the risk of hypocalcemia like lymphadenectomy, hyperthyroidism, parathyroid reimplantation^{22–24}, were not found to be independent predictor in the present study. This suggests that postoperative PTH, and hence %PTH, already capture most of the information needed to estimate the risk of hypocalcemia. On the other hand, our results suggest that female sex, younger age and low body mass index may be associated with hypocalcemia independently of hypoparathyroidism. In contrast, other pre-operative characteristics or intra-operative events increase risk only through a decrease in parathyroid function. Finally, hyperthyroidism did not have any influence on risk prediction in our population, in contrast with previous report describing its association with hypocalcemia through the “hungry bone syndrome”²⁵. Routine medical control of hyperthyroidism before thyroidectomy in our center may explain this apparent discrepancy^{25,26}.

The main strength of our study lies in the mathematical approach used, based on machine learning with multiple cross-validations. The resulting model allowed an individualized prediction of the postoperative hypocalcemia risk (PHR) based only on five simple pre- and intraoperative variables, that can be easily incorporated in the online calculator provided. To ensure strong external validity, our definition of post-thyroidectomy hypocalcemia was deliberately very broad, including all patients with biological hypocalcemia, irrespective of its symptomatic nature, depth, or duration. Future work may focus on the potential determinants of symptomatic hypocalcemia or prolonged hypoparathyroidism. Another limitation of our work lies in the absence of model testing on an independent external cohort. Therefore, the generalizability of our findings and algorithm warrants further validation in other centers. In conclusion, in the present study, we leveraged machine learning to develop and validate a multivariable algorithm allowing the early and accurate prediction of postoperative hypocalcemia, combining three simple preoperative variables with intraoperative PTH measurement, which may guide early post-operative clinical management after total thyroidectomy. The proposed approach could help to individualize postoperative care and facilitate the implementation of outpatient thyroidectomy.

ACKNOWLEDGMENTS AND CONFLICTS OF INTEREST

Authors have no acknowledgments nor conflicts of interest to declare.

References

1. Del Rio P, Carcoforo P, Medas F, et al. Adverse events in thyroid surgery: observational study in three surgical units with high volume/year. *BMC Surg.* 2021;21:352.
2. Kazaure HS, Zambeli-Ljepovic A, Oyekunle T, et al. Severe Hypocalcemia After Thyroidectomy: An Analysis of 7366 Patients. *Ann Surg.* 2021;274:e1014–e1021.
3. Sessa L, De Crea C, Zotta F, et al. Post-thyroidectomy hypocalcemia: Is a routine preferable over a selective supplementation? *Am J Surg.* 2021;S0002-9610(21)00605-X.
4. Mercante G, Anelli A, Giannarelli D, et al. Cost-effectiveness in transient hypocalcemia post-thyroidectomy. *Head Neck.* 2019;41:3940–3947.
5. Frey S, Van Den Heede K, Triponez F, et al. Prevention of hypocalcemia and hypoparathyroidism after total thyroidectomy. Recommendations of the Francophone Association of Endocrine Surgery (AFCE) with the French Society of Endocrinology (SFE) and the French Society of Nuclear Medicine (SFMN). *J Visc Surg.* 2023;160:S95–S109.
6. Segel JM, Duke WS, White JR, et al. Outpatient thyroid surgery: Safety of an optimized protocol in more than 1,000 patients. *Surgery.* 2016;159:518–523.
7. Stein DJ, Noordzij JP, Kepchar J, et al. Use of Parathyroid Hormone Assay after Thyroidectomy: A Survey of US and European Surgeons. *Clin Med Insights Endocrinol Diabetes.* 2013;6:39–45.
8. Mathur A, Nagarajan N, Kahan S, et al. Association of Parathyroid Hormone Level With Postthyroidectomy Hypocalcemia: A Systematic Review. *JAMA Surg.* 2018;153:69–76.
9. Saba A, Podda M, Messina Campanella A, et al. Early prediction of hypocalcemia following thyroid surgery. A prospective randomized clinical trial. *Langenbecks Arch Surg.* 2017;402:1119–1125.
10. Paladino NC, Guérin C, Graziani J, et al. Predicting risk factors of postoperative hypocalcemia after total thyroidectomy: is safe discharge without supplementation possible? A large cohort study. *Langenbecks Arch Surg.* 2021;406:2425–2431.
11. Lee DR, Hinson AM, Siegel ER, et al. Comparison of Intraoperative versus Postoperative Parathyroid Hormone Levels to Predict Hypocalcemia Earlier after Total Thyroidectomy. *Otolaryngol Head Neck Surg.* 2015;153:343–349.
12. Barczyński M, Cichoń S, Konturek A. Which criterion of intraoperative iPTH assay is the most accurate in prediction of true serum calcium levels after thyroid surgery? *Langenbecks Arch Surg.* 2007;392:693–698.
13. Grodski S, Serpell J. Evidence for the role of perioperative PTH measurement after total thyroidectomy as a predictor of hypocalcemia. *World J Surg.* 2008;32:1367–1373.

14. Schlottmann F, Arbulú ALC, Sadava EE, et al. Algorithm for early discharge after total thyroidectomy using PTH to predict hypocalcemia: prospective study. *Langenbecks Arch Surg.* 2015;400:831–836.
15. Alía P, Moreno P, Rigo R, et al. Postresection parathyroid hormone and parathyroid hormone decline accurately predict hypocalcemia after thyroidectomy. *Am J Clin Pathol.* 2007;127:592–597.
16. Liu C, Tang L, Goel P, et al. A Practical Mathematic Method to Predict and Manage Hypocalcemia After Parathyroidectomy and Thyroidectomy. *Ann Otol Rhinol Laryngol.* 2020;129:70–77.
17. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research.* 2010;11:2079–2107.
18. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON.*
19. Pattou F, Combemale F, Fabre S, et al. Hypocalcemia following thyroid surgery: incidence and prediction of outcome. *World J Surg.* 1998;22:718–724.
20. Rosa KM, Matos LL de, Cernea CR, et al. Postoperative calcium levels as a diagnostic measure for hypoparathyroidism after total thyroidectomy. *Arch Endocrinol Metab.* 2015;59:428–433.
21. Soares CSP, Koga KH, Moriguchi SM, et al. Development of a tool to calculate the probability of hypocalcemia after total thyroidectomy: a prospective study. *Langenbecks Arch Surg.* 2024;409:33.
22. Puzziello A, Rosato L, Innaro N, et al. Hypocalcemia following thyroid surgery: incidence and risk factors. A longitudinal multicenter study comprising 2,631 patients. *Endocrine.* 2014;47:537–542.
23. Chen Z, Zhao Q, Du J, et al. Risk factors for postoperative hypocalcaemia after thyroidectomy: A systematic review and meta-analysis. *J Int Med Res.* 2021;49:300060521996911.
24. Eismontas V, Slepavicius A, Janusonis V, et al. Predictors of postoperative hypocalcemia occurring after a total thyroidectomy: results of prospective multicenter study. *BMC Surg.* 2018;18:55.
25. Karunakaran P, Maharajan C, Ramalingam S, et al. Is hungry bone syndrome a cause of postoperative hypocalcemia after total thyroidectomy in thyrotoxicosis? A prospective study with bone mineral density correlation. *Surgery.* 2018;163:367–372.
26. Fields T, Ramonell K, Fazendin J, et al. Postoperative Hypocalcemia in Hyperthyroid Patients: The Parathyroids Aren't Always to Blame. *J Surg Res.* 2023;288:202–207.
27. Lombardi CP, Raffaelli M, Princi P, et al. Early prediction of postthyroidectomy hypocalcemia by one single iPTH measurement. *Surgery.* 2004;136:1236–1241.

28. Noordzij JP, Lee SL, Bernet VJ, et al. Early prediction of hypocalcemia after thyroidectomy using parathyroid hormone: an analysis of pooled individual patient data from nine observational studies. *J Am Coll Surg*. 2007;205:748–754.
29. Blanchard C, Bannani S, Pattou F, et al. Impact of body mass index on post-thyroidectomy morbidity. *Head Neck*. 2019;41:2952–2959.
30. Remer LF, Linhares SM, Scola WH, et al. Transient Hypocalcemia After Total Thyroidectomy: The Obesity Paradox at Work? *J Surg Res*. 2022;278:93–99.
31. Kamycheva E, Sundsfjord J, Jorde R. Serum parathyroid hormone level is associated with body mass index. The 5th Tromsø study. *Eur J Endocrinol*. 2004;151:167–172.

ACCEPTED

Figure 1 : AUROC curve of %PTH alone and random forest model vaue to predict the risk of postoperative hypocalcemia. *Abbreviations.* %PTH: Percentage decay between parathormone at anesthetic induction and parathormone at skin closure; AUROC: area under the receiver operating characteristic curve.

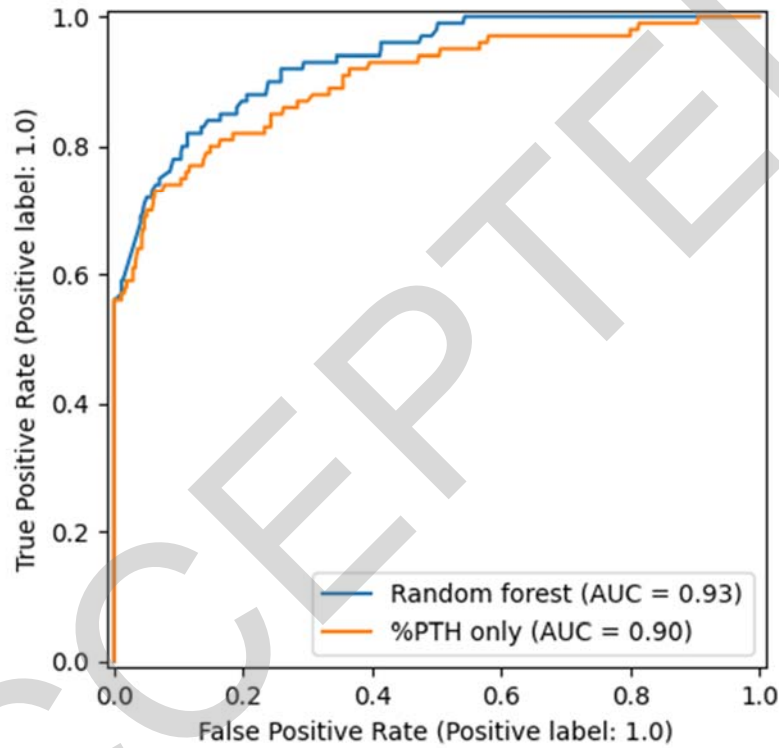


Figure 2: Screen capture of the shinyapp allowing the automated calculation of postoperative hypocalcemia risk of an individual patient, according to the 5 variables of interest (<https://lille-model.shinyapps.io/Hypocal/>). The colour code indicates the risk of experiencing postoperative hypocalcemia as low (green $\leq 7\%$), intermediate (orange) and high (red, $\geq 20\%$).

Post operative risk of hypocalcemia

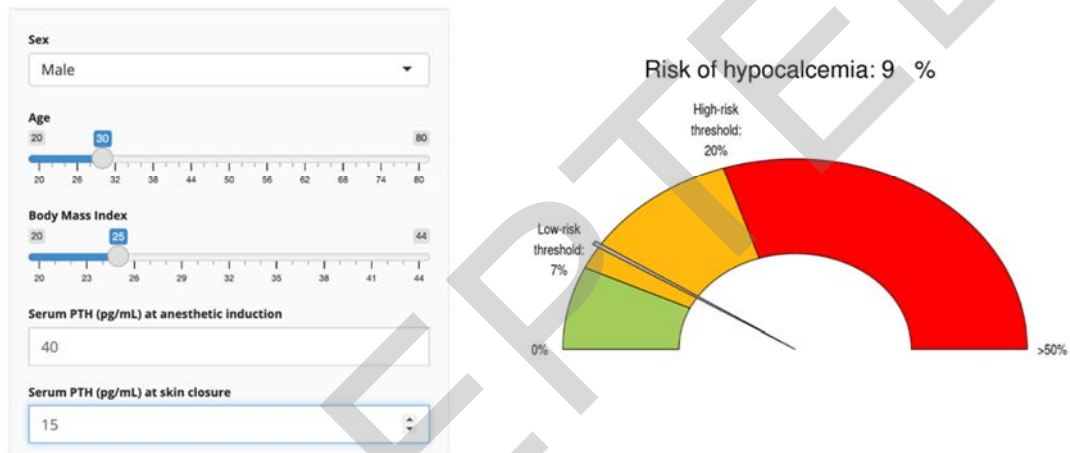


Table 1: Comparison of pre-, intra- and post-operative characteristics in patients with or without postoperative hypocalcemia

N : number ; Sd : standard deviation ; BMI : body mass index ; PTH : parathormone ; POD : postoperative day ; * : Other indications consisted in one myasthenia and two drug induced hyperthyroidism ; %PTH : Percentage decay between PTH at anesthetic induction and PTH at skin closure ; IQR : interquartile range

	Hypocalcemia	Missing data, n (%)	Control	Missing data, n (%)	P
N	100	-	510	-	-
Median age, years (IQR)	44.5 (35.2-55)	0 (0)	52 (39-63)	0 (0)	0.0006
Female sex, n (%)	85 (85)	0 (0)	374 (73.3)	0 (0)	0.015
Median BMI, kg/m2 (IQR)	24.2 (22.3-31.2)	18 (18)	27.6 (23.9-32.5)	62 (12)	0.001
Preoperative history, n (%)					
- Hyperthyroidism	32 (32.0)	0 (0)	181 (35.5)	0 (0)	0.566
- Preoperative vitamin D deficiency supplementation	16 (16.0)	1 (1)	78 (15.29)	8 (1.6)	0.879
Surgical indication, n (%)		0 (0)		0 (0)	
- Grave's disease	16 (16.0)		95 (18.6)		
- Suspicious node	36 (36.0)		183 (35.9)		
- Toxic node	4 (3.0)		7 (1.37)		
- Suspicious multinodular goiter	11 (10.0)		45 (8.8)		0.530
- Toxic multinodular goiter	11 (11.0)		64 (12.5)		
- Symptomatic goiter	22 (22.0)		113 (22.2)		
- Other*	0 (0)		3 (0.59)		
Preoperative biological data, median (IQR)					
- Serum albumin (g/dL)	43 (39.5-46)	23 (23)	43 (41-46)	141 (27,6)	0.148
- Serum creatinine (mg/L)	7 (6-8)	7 (7)	7 (6-8)	57 (11.2)	0.208
- Serum Vitamin D (ng/mL)	25 (17-29)	18 (18)	25 (18-32)	102 (20)	0.462
- Serum PTH at anesthetic induction (pg/mL)	59 (38.2-81.7)	0 (0)	61 (44.7-80.2)	0 (0)	0.441
Intraoperative data					
- Median operating time, minutes (IQR)	166 (131-229)	0 (0)	159 (132-197)	2 (0.4)	0.166
- Number of parathyroids seen, median (IQR)	4 (3-4)	1 (1)	4 (3-4)	26 (5.1)	0.7743

- Neck dissection, n (%)	40 (40.0)	0 (0)	167 (32.7)	0 (0)	0.167
- Parathyroid reimplantation, n (%)	13 (13.0)	0 (0)	46 (9.0)	0 (0)	0.265
- Abnormal postoperative laryngoscopy n (%)	8 (8.0)	2 (2)	27 (5.29)	12 (2.4)	0.343
- Median thyroid weight, gram (IQR)	44 (24-90)	1 (1)	43 (22-86)	11 (2.2)	0.705
Postoperative biological data					
- Median serum PTH at skin closure, pg/mL (IQR)	0 (0-10)	0 (0)	26 (16-43)	0 (0)	<0,0001
- Median serum PTH at POD 1, pg/mL (IQR)	5 (0-12)	0 (0)	26 (18-36)	1 (0.2)	<0,0001
- Median serum calcium at POD 1, mg/L (IQR)	80 (77-83)	0 (0)	88 (86-91)	1 (0.2)	<0,0001
- Median %PTH (%) (IQR)	100 (83.2-100)	0 (0)	51.4 (30-71.6)	0 (0.2)	<0,0001
- Median hospital stay, days (IQR)	3 (2-4)	0 (0)	1 (1-2)	0 (0)	<0,0001

ACCEPTED

Table 2 : Performances of %PTH alone for predicting postoperative hypocalcemia according to various thresholds

Threshold (%)	%PTH		
	>60	> 75	> 85
Sensitivity (%)	92	82	74
Specificity (%)	61	81.6	90
LR+	2.5	4.45	7.4
LR-	0.14	0.22	0.29

. %PTH : Percentage decay between parathormone at anesthetic induction and parathormone at skin closure ; LR+ : positive likelihood ratio ; LR- : negative likelihood ratio

Table 3 : AUROC curves values for postoperative hypocalcemia prediction evaluated by repeated cross-validation for %PTH only model, and for the different machine learning models

	Mean AUC	95% CI
%PTH only	0.889	0.806-0.966
Logistic regression	0.898	0.810-0.958
Logistic regression with regularization	0.896	0.820-0.961
Decision tree	0.865	0.740-0.963
Random forest	0.902	0.829-0.970
Gradient boosting	0.901	0.866-0.964
Neural network	0.884	0.818-0.954

%PTH: Percentage decay between parathormone at anesthetic induction and parathormone at skin closure; AUROC curve: area under the receiving operator characteristics curve ; CI: confidence interval

Table 4: Feature importance of the random forest model

	Feature importance
%PTH squared	0.46
PTH T0	0.34
Interaction between %PTH and sex	0.19
BMI	0.01
Age squared	0.005

%PTH: Percentage decay between parathormone at anesthesia induction and parathormone at skin closure; PTH T0: Parathormone at anesthetic induction

Table 5: Performances of the random forest model for predicting postoperative hypocalcemia according to low and high values of postoperative hypocalcemia risk (PHR)

	Low threshold: PHR 7%	High threshold: PHR 20%
Sensitivity (%)	92	79
Specificity (%)	73	90
LR+	3.4	7.6
LR-	0.11	0.23

LR+: positive likelihood ratio; LR-: negative likelihood ratio; PHR: postoperative hypocalcemia risk, a pre-test risk for postoperative hypocalcemia calculated by the model for a given patient.

#28

ANNSURG-D- François
24-00533 Pattou

Machine learning based algorithm for the early
prediction of postoperative hypocalcemia risk ESA Paper
after thyroidectomy

First Discussant: Elizabeth Nieveen van Dijkum (Amsterdam, The Netherlands)

In this single-center retrospective study including 610 patients over a 29-month period, machine learning was used to develop a model predicting the risk of hypocalcemia after total thyroidectomy. The dynamics of perioperative PTHi measurements (pre-OP PTHi and PTHi drop in %) were found to be the most potent predictors in isolation. However, the association of perioperative PTHi with sex, age, and BMI showed the best performance for hypocalcemia prediction.

I have the following questions for the authors: First, study aims to improve risk prediction and apply specific statistical techniques. You mention the need for a robust predictor, but how does your prediction model compare to other predictors already published?

Second, the thorough machine learning statistical analysis resulted in including 5 factors, with the most important factors the drop in PTH and the PTH at the start of surgery. These factors are known and mostly used in existing models and not new. New factors, not known from earlier risk predictions are BMI and age, but these factors have very low predictive values. How valid are these factors, since the development cohort already had a difference in these BMI and age values? Additionally, why was external validation not included, especially focusing on these two factors?

Finally, the predictor indicates low (<8%) intermediate (8-19%) and high (>19%) risk for hypocalcemia. How can these values be used in patient care? Is a patient with a risk of 10% discharged? With or without calcium supplementation? What is the accepted number to treat (calcium supplementation) to prevent for example one re-admission in relation to this risk predictor? Why was not use a simple high and low risk based on an accepted NNT?

Response From François Pattou (Lille, France)

Thank you very much for these very sound, interesting, and important questions. To answer the first question regarding the different risk factors that have been described: We started with our data and identified the statistically significant risk factors through analysis. This study focused on mathematics and data evaluation. If you perform this analysis on thousands of patients, you may detect small signals of hyperparathyroidism, which could be a good sign. This is a key reason for our study.

Addressing your second question: The percentage of PTH results is crucial and significantly influences the statistics. When analyzed using linear regression, the mean values provide a strong prediction of PTH levels. However, linear regression can miss some patients, which is where non-linear mathematics excels. It can identify information in the gaps that linear regression overlooks. This is the strength of machine learning – it is very efficient in capturing these subtleties.

To answer the last question: Safely discharging patients is the most important aspect. We should be able to assure patients “You are safe to go back home and do not need to send me calcium in one month”. For patients at higher risk (red category), it’s challenging because of the low patient numbers. Clinical judgment and context play significant roles. In our practice, patients in the yellow or red categories stay an additional day, and we measure the calcium the next day. However, in other settings where patients are discharged quickly, this metric becomes crucial for patient safety.

Discussant: Antonio D. Pinna (Weston, United States)

I am very curious to know, if it is possible to include the application in an electronic medical record?

Response From François Pattou (Lille, France)

Sure! You need four entry points (BMI, age, etc.). These plugins can be easily integrated into a medical record system to provide useful tools. We use this approach for predicting weight trajectories and sell the information to those managing medical records.

Discussant: Lawrence T. Kim (Chapel Hill, United States)

I have a rather mundane question related to the previous one. One of the barriers to this kind of work is that many of us, including myself, don’t know how to do the coding and other technical aspects involved in creating these techniques. Do you do it yourself? Did you learn how to do it, and if so, how do you get it done?

Response From François Pattou (Lille, France)

Thank you for this very important question. I don’t do it by myself; I’ve been with a team for five years. This interdisciplinary work has involved a lot of discussions. In healthcare, it’s crucial to use interpretable techniques. For example, understanding why a lower BMI is a risk factor is important. Surgeons don’t need to code themselves, but spending significant time collaborating with technical experts is essential. What won’t succeed in healthcare is the “non-interpretable black box”. Unlike a phone navigation system that guides you with an explanation, healthcare requires interpretability. Health-technology assessments need to understand and evaluate these tools to provide validation and reimbursement. There’s a difference between widely available tools and those validated for real care. Interpretability is the key.