Published in final edited form as:

Surgery. 2017 April ; 161(4): 1113–1121. doi:10.1016/j.surg.2016.09.044.

Improving Diagnostic Recognition of Primary Hyperparathyroidism with Machine Learning

Yash R. Somnay, BS¹, Mark Craven, PhD³, Kelly L. McCoy, MD⁴, Sally E. Carty, MD⁴, Tracy S. Wang, MD, MPH⁵, Caprice C. Greenberg, MD MPH², and David F. Schneider, MD MS FACS¹

¹Section of Endocrine Surgery, Department of Surgery, University of Wisconsin, Madison, WI

²Wisconsin Surgical Outcomes Research Program, Department of Surgery, University of Wisconsin, Madison, WI

³Department of Biostatistics and Medical Informatics, and the Department Computer Science, University of Wisconsin, Madison, WI

⁴Division of Endocrine Surgery, Department of Surgery, University of Pittsburgh, Pittsburgh, PA

⁵Division of Surgical Oncology, Department of Surgery, Medical College of Wisconsin, Milwaukee, WI

Abstract

Importance—Parathyroidectomy offers the only cure for primary hyperparathyroidism (PHPT), but today only 50% of PHPT patients are referred for surgery, in large part because the condition is widely under-recognized. PHPT diagnosis can be especially challenging with mild biochemical indices. Machine learning (ML) is a collection of methods in which computers build predictive algorithms based on labeled examples.

Objective—With the aim of facilitating diagnosis, we tested the ability of ML to distinguish PHPT from normal physiology using clinical and laboratory data.

Design—This is a retrospective cohort study using a labeled training set and 10-fold cross-validation to evaluate algorithm accuracy. Measures of accuracy included area under the ROC curve, precision (sensitivity), and positive and negative predictive value. Several different ML algorithms and ensembles of algorithms were tested using the Weka platform.

Setting—3 high-volume endocrine surgery programs

Participants—Among 11,830 patients managed surgically from March, 2001 to August 2013, 6,777 underwent parathyroidectomy for PHPT, and 5,053 control patients without PHPT underwent thyroidectomy.

Corresponding Author: David F. Schneider, MD, MS, Assistant Professor of Surgery, Director, Endocrine Surgery Research, University of Wisconsin, Department of Surgery, 600 Highland Avenue, K4/728 CSC, Madison, WI 53792, 608-263-1387, schneiderd@surgery.wisc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Main Outcomes and Measures—Test-set accuracies for ML models were determined using 10-fold cross-validation. Age, gender, preoperative calcium, phosphate, PTH, Vitamin D, and creatinine were defined as potential predictors of PHPT. Mild PHPT was defined as PHPT with normal preoperative calcium or PTH levels.

Results—After testing a variety of ML algorithms, Bayesian network models proved most accurate, correctly classifying 95.2% of all PHPT patients (area under ROC=0.989). Omitting PTH from the model did not significantly reduce the accuracy (area under ROC = 0.985). However, in mild disease cases, the Bayesian network model correctly classified 71.1% of patients with normal calcium and 92.1% with normal PTH levels preoperatively. Bayesian networking + AdaBoost improved the accuracy to 97.2% correctly classified (area under ROC=0.994) cases, and 91.9% of PHPT patients with mild disease. This was significantly improved relative to Bayesian networking alone ($p<0.0001$).

Conclusions and Relevance—ML can accurately diagnose PHPT without human input, even in mild disease. Incorporation of this tool into electronic medical record systems may greatly aid in recognition of this under-diagnosed disorder.

INTRODUCTION

Primary hyperparathyroidism (PHPT) is the state of autonomous parathyroid hormone (PTH) hypersecretion by one or more abnormal parathyroid glands. Over-secretion of PTH causes hypercalcemia and accelerates bone loss, resulting in skeletal abnormalities such as osteopenia/osteoporosis, nephrolithiasis, renal insufficiency, peptic ulcer disease, acid reflux, cardiovascular disease and hypertension. Neurocognitive changes like fatigue, depression, anxiety, memory loss, and irritability are also associated with PHPT^{1–3}. Surgical excision of hyperfunctioning parathyroid glands remains the only curative treatment for PHPT patients, often with greater than 95% success rates^{4–6}.

Today however, less than half of biochemically proven PHPT patients are referred for surgery^{7,8}. PHPT is the most common etiology for elevated calcium in the ambulatory population, but documented hypercalcemia often exists for many years before PHPT is formally diagnosed and evaluated^{8–10}. Numerous potential reasons could explain why a knowledgeable provider might delay or forgo further workup for PHPT. In one study of a large health system's EMR, only 32% of patients with hypercalcemia had a PTH measured, suggesting that under-recognition rather than conscious observation accounts for much of the under-treatment for PHPT⁸.

The increasing availability of data through electronic medical record (EMR) systems offers a valuable opportunity to improve PHPT recognition. Federal incentives have led to the rapid proliferation of EMR systems, but the use of these tools to analyze the vast amounts of data contained within EMRs still lag behind their widespread adoption^{11,12}. Computer-based decision support tools have improved healthcare delivery in domains including preventative care and drug interactions^{13–15} to guide clinical decision making. Decision-support tools commonly exist as alerts to prompt or alter physician order entry¹⁶. Supervised machine learning (ML) methods are focused on inferring predictive models from existing datasets that consist of labeled examples,^{17,18} and in clinical medicine, can classify or predict future

outcomes for a given patient^{19,20}. ML approaches have been employed to predict risk of breast cancer from mammographic findings²¹ or to risk stratify patients for venous thromboembolism from EMR data²². In this study, we hypothesized that ML could accurately diagnose PHPT based on readily available clinical and laboratory indicators. We describe the development and evaluation of a ML method that accurately predicts PHPT using a large, multi-institutional dataset. Improving recognition of the disease and distinguishing mild disease from normal variations drove the work described here

METHODS

Study Sample and Definitions

The Institutional Review Boards of the University of Wisconsin School of Medicine and Public Health, the University of Pittsburgh School of Medicine, and the Medical College of Wisconsin approved the collection of data. We compiled a dataset from the existing, prospectively collected perioperative datasets of 11,830 patients from three high-volume academic centers. Of these patients, 6,777 patients (STUDY) with biochemical PHPT underwent an initial parathyroid operation for sporadic disease. Only patients confirmed to have PHPT at the time of operation were included in the STUDY group. The remaining 5,053 patients (CONTROL) underwent thyroidectomy without parathyroid disease as determined by preoperative laboratory indices. The dates of operations spanned from March 2001 to August 2013. All patients undergoing parathyroidectomy were diagnosed preoperatively to have biochemical evidence of PHPT and control patients did not have biochemical evidence of PHPT, as both groups had a minimum of serum calcium and PTH levels measured preoperatively.

ML is a collection of different methods, or algorithms, each with differing classification mechanisms. The typical approach for employing ML for clinical problems is to test a variety of different algorithms to find the most accurate. To accomplish this task, we used the ML software platform available from the Waikato Environment for Knowledge Analysis (Weka Version 3.6.11). The Weka software suite contains a library of algorithms that build predictive models by learning from examples provided in user-supplied datasets. Each algorithm was trained using our patient dataset, which included 7 attributes that we considered most likely to be relevant to PHPT diagnosis: age at surgery, gender, preoperative serum calcium, paired preoperative serum PTH, serum 25-hydroxy vitamin D, serum creatinine, and serum phosphate values. Predictor variables chosen for two reasons: 1) they were consistently available within the databases from the three institutions, and 2) represent initial, routinely available data within a given patient's electronic records that could indicate PHPT without an exhaustive or complete workup for PHPT. The studied laboratory values were those obtained closest in time to the date of surgery.

As outlined in the recent surgical guidelines on management of hyperparathyroidism (21), calcium and PTH levels characteristically vary quite a bit during the biochemical diagnosis of PHPT. In normocalcemic hyperparathyroidism, the calcium can be intermittently normal. Alternatively, the PTH level can fall within the normal range, but still be inappropriately high for the calcium level. For the purposes of this study, 'mild' PHPT was defined to be present when either the calcium or PTH level that was recorded closest to the operative date,

was within the normal range²³. Correct diagnostic classification of mild PHPT was defined to be present when a patient's ML-predicted disease status agreed with his physicians' preoperative diagnostic impression based on biochemical indices.

Classification Methods

We began by testing several categories of supervised machine learning classifiers: 1) rule-based classifiers, 2) logistic regression based classifiers, 3) tree based classifiers, and 4) Bayesian network classifiers. After testing over 20 different algorithms, we selected the Bayesian Network (BayesNet) for further study as it maximized accuracy and transparency. A Bayesian Network uses a graphical representation to represent a compact encoding of the joint distribution of input variables and the outcome(s) of interest^{24,25}. Although more complex Bayesian network structures were evaluated, the naïve Bayes models proved as accurate, and more comprehensible, than the complex networks. A naïve Bayesian network is one in which the graph structure posits that each input variable is independent of the others when conditioned on the outcome variable. The BayesNet method handles missing data through a process called imputation where missing features are inferred from other cases where those features are present. Therefore, the algorithm does not require complete datasets for either training or prediction²⁶.

We then explored means to improve the accuracy of the diagnostic model for patients with mild biochemical disease. The method selected is known as adaptive boosting, or AdaBoost, a meta-algorithm that can be coupled with other classifiers to improve performance²⁷ by placing more weight on previously misclassified instances, thereby 'boosting' previously weak classifiers. AdaBoost was chosen to place more weight on the misclassified, mild biochemical cases. An ensemble of BayesNet in conjunction with AdaBoost was then run on the established datasets.

Statistical Analysis

All ML classifiers were tested using 10-fold cross validation. The performance of each classifier was evaluated based on sensitivity (recall), specificity, positive predictive value (precision), area under the receiver operating characteristic (ROC) curve, and overall accuracy. Continuous variables are represented by their mean $\pm$ standard deviation. Statistical comparisons of continuous variables were made using Student's T-test. Categorical variables were statistically compared using Pearson's Chi-square test.

RESULTS

Baseline Patient Characteristics and Accuracy of Bayesian Network Based Model

This sample study included 6,777 biochemically confirmed PHPT patients and 5,053 control patients, of whom 76.5% and 80.1% were female, respectively. PHPT patients (mean age, 59 $\pm$ 0.2 years) were older than control patients (mean age, 49 $\pm$ 0.2 years; $p < 0.0001$). Serum levels of calcium, PTH, phosphate, vitamin D, and creatinine were all consistent with the biochemical definition of PHPT (Table 1)²³. 8,450 cases had at least one missing feature, but 79.1% of the cases had only 1 missing feature. The most common missing features, in descending order, were phosphate, creatinine, and vitamin D.

After testing a variety of ML methods, we found that a naïve Bayesian network most accurately predicted patient diagnoses (PHPT vs. control). A graphical representation of the network structure is shown in Figure 1, with each feature contributing independently to the probabilistic classification of patients. Here, the Bayesian network is “naïve” because each feature is conditionally dependent on the outcome alone without any arcs or probabilistic relationships between the predictor features. The Bayesian networks resulted in a 95.2% test set accuracy, and an area under the ROC curve of 0.989 (Table 2).

Patients with mild biochemical indices for PHPT represent a diagnostic challenge because they must be distinguished from those with secondary causes of elevated serum PTH such as vitamin D deficiency. Hence, we further tested the accuracy of our model in making diagnostic predictions that were consistent with our specialists’ clinical diagnoses specifically in these subpopulations (Table 2). The accuracy of our model did not change for patients with low vitamin D (<30 ng/mL), correctly classifying 95.6%. Among patients with mild biochemical disease (N=1,738) however, we observed a reduction in the accuracy of classification to 86.0%. These findings suggest that the accuracy of the BayesNet predictive model may be more limited when it comes to categorizing patients with milder biochemical disease.

We further investigated all the misclassified PHPT cases and compared them to the correctly classified instances. When we compared PHPT patients who were correctly diagnosed (true positives) to those mistaken as normal (false negatives), the latter group was on average significantly younger ($p=6.7E-9$), had higher serum phosphate levels ($p=1.2E-8$), and had higher serum vitamin D levels ($p=0.0002$). Much of the misclassification occurred in mild biochemical disease; 99% of the false negatives had mild disease, while only 23% of the true positives had mild disease. False negatives had a mean serum calcium level within normal range (9.7 ± 0.04 mg/dL) and a much lower mean serum PTH level (110.9 ± 10.6 pg/mL) compared to that of true positives (181.0 ± 7.1 pg/mL, $p=1.6E-8$).

Control patients who were correctly classified (true negatives) were on average younger than those incorrectly classified as having PHPT (false positives) ($p=3.8E-5$). The false positive controls exhibited significantly higher mean calcium (10.1 ± 0.04 mg/dL), and PTH levels (74.03 ± 4.2 pg/mL) ($p=8.0E-32$). 6.3% of control patients were mistaken to have PHPT.

Adaptive Boosting of BayesNet

Although the BayesNet algorithm alone was highly accurate, this ML tool becomes even more useful to experts if it can reliably distinguish cases of mild hyperparathyroidism. To reduce the rate of misclassification among the mild cases, we employed the meta-algorithm known as adaptive boosting in conjunction with BayesNet, and observed that doing so enhanced our model’s classification accuracy from 95.2% to 97.2% with an area under the ROC of 0.994 (Table 3). The total number of misclassified patients decreased from 563 (4.76%) using BayesNet alone, to 329 (2.78%) using the BayesNet+AdaBoost model (Table 3). Since many of the previously misclassified patients had mild PHPT, the use of adaptive boosting improved classification accuracy among those with mild disease.

With the same patient dataset used for the original BayesNet model, we once again compared the biochemical profiles of those misclassified to those correctly classified after BayesNet+AdaBoost. The boosted model significantly reduced the number of false negatives from 246 (2.07%) to 149 (1.26%, $p < 0.0001$, Table 3). Importantly, AdaBoost reduced the number of false negatives with mild disease (14.0% in BayesNet alone compared to 8.1% in BayesNet+AdaBoost, $p < 0.0001$, Table 3). The improvement in classification of mild cases was associated with a small increase in false positives in the control group, specifically for patients with serum PTH levels above the upper limit of normal (2.7% vs. 3.6% false positives, $p < 0.0001$).

Alternative ML Approach

Since the observed accuracy of the BayesNet + AdaBoost was higher than in many clinical studies using machine learning methods^{22,28}, we sought to test its classification accuracy using an alternate approach. Rather than employing 10-fold cross-validation (randomly chosen test sets), we used each institutional data as a separate test set allowing assessment of any institutional variations in accuracy that may indicate different referral or diagnosis patterns. The model was trained in the UW dataset and then tested on the other two institutional datasets. Although there were slight differences in sensitivity and specificity, the Bayesian network had similar overall accuracy when tested in this way, and the area under the ROC curve was nearly identical to our original analysis (Table 4).

Omission of PTH

The set of laboratory values included as predictive features for the above BayesNet and boosted algorithm includes a serum PTH level. However, for a clinician to order a serum PTH level means that he or she is already considering the diagnosis of PHPT. To assist the physician who either lacks awareness of PHPT or is not yet considering that particular diagnosis, we also tested several ML models omitting PTH as a predictive feature. Again, BayesNet proved the most accurate algorithm, correctly classifying 95.6% of all STUDY cases. The false positive rate was 5.4%, and the area under the ROC curve was 0.985. Meta algorithms like adaptive boosting could not improve upon the accuracy of the BayesNet alone in this particular setting.

DISCUSSION

In summary, we trained a machine learning algorithm to diagnose PHPT based on biochemical and demographic data using a multi-institutional dataset. The algorithm was highly accurate, with over 95% accuracy tested by two different techniques. The boosted algorithm performed remarkably well even for patients with mild disease. Additionally, the boosted algorithm appeared to be useful in distinguishing mild hyperparathyroidism from normal, physiologic alterations in serum calcium and/or PTH. The BayesNet model without PTH remained highly accurate even without PTH as a predictive feature. Therefore, we provide here different algorithms useful in different scenarios, depending on the informational needs of the provider: the BayesNet + AdaBoost is useful for the expert trying to distinguish mild disease from normal physiology, while the BayesNet without serum PTH can assist the clinician who lacks diagnostic conviction and/or awareness of the disease.

Either or both algorithms could serve as the underlying code for a clinical decision support tool that prompts physicians to refer patients for further evaluation when labs are consistent with primary hyperparathyroidism.

The clinical diagnosis of PHPT is primarily based on interpretation of laboratory indices including PTH, calcium and vitamin D^{29,30}. Emerging evidence indicates that many with biochemical evidence of this disease remain undiagnosed, resulting in a vast, unrecognized population of afflicted patients^{7,8}. Within the Kaiser healthcare system (3.5 million patients), over one half of those with biochemical evidence of PHPT did not receive a parathyroidectomy over a 13 year period, regardless of symptoms or fulfillment of international consensus criteria for treatment⁷. However, population level studies cannot indicate why patients go untreated. Several other studies had similar findings: for patients with elevated serum calcium, unfortunately serum PTH levels are frequently not obtained despite PHPT being the most common etiology of outpatient hypercalcemia^{7–10}. To address situations where serum PTH is never ordered, we tested ML algorithms without PTH as a predictive feature, and found BayesNet without PTH is still over 95% sensitive for detecting PHPT.

Even for patients who meet objective diagnostic criteria, several studies have found that parathyroidectomy is significantly underutilized,^{8,31} and the under-treatment appears to hinge on at least two factors: first, a lack of diagnostic awareness and second, deferment of evaluation³². The informatics tools reported here can assist with recognition, and integration into the EMR can prompt timely referral for further evaluation and treatment. Importantly, this tool utilizes readily available clinical data, and provides automated diagnosis without additional cost or invasive testing.

Parathyroidectomy benefits PHPT patients by not only correcting their biochemical disease, but also, preventing the progression of PHPT-related complications, including worsening bone mineral density and fragility fractures^{33–36}. Compared to observation, Almqvist et al. determined that early parathyroidectomy reduced patients' long-term risk of hip and other bone fractures³⁷. Longstanding PHPT has also been associated with poorer quality of life, psychiatric illnesses, left ventricular hypertrophy and heightened mortality risk from cardiovascular disease^{38–45}. Additionally, parathyroidectomy is more cost-effective than observation⁴⁶. The tool developed here promises to aid both practitioners and patients recognize and treat the disease earlier to reduce long-term morbidity.

It is important to identify mild PHPT because patients without significant biochemical disease still experience symptoms and a decline in bone density^{47,48}. Accumulating studies provide evidence that mild PHPT patients benefit from parathyroidectomy despite not meeting formal consensus criteria for surgical intervention^{23,33,34,42,47,49,50} as the new AAES surgical guidelines discuss in detail (21). Moreover, the incidence of mild PHPT is on the rise⁵¹. Diagnosis of mild biochemical PHPT becomes challenging for primary providers, endocrinologists, and surgeons. Aside from difficulty distinguishing the mild biochemical PHPT from normal physiology or secondary causes of hyperparathyroidism, the symptoms associated with PHPT are often non-specific. Symptoms like fatigue or musculoskeletal pain are common in PHPT and numerous other conditions^{7–9,52}. Therefore, physicians may focus

on other etiologies of reported symptoms prior to fully evaluating an elevated serum calcium or PTH. Adaptive boosting of the Bayesian classifier improved the classification mild PHPT, and in doing so, increased the overall accuracy of the model. For clinicians with less awareness of PHPT, the model without serum PTH can prompt further workup. For instance, a “pop-up” alert might prompt the provider to order a serum PTH. Hence, this tool can aid *both* the expert *and* the non-specialist distinguish PHPT from normal variants, or at least prompt earlier biochemical evaluation by raising awareness for clinicians less familiar with PHPT.

Emerging computer decision support tools that use machine learning reliably assist physicians in a variety of other domains such as predicting venous thromboembolism, assessing breast cancer risk, or detection of myocardial infarction^{21,28,53,22}. However, few decision support aids for diagnosing PHPT currently exist and are limited mainly to graphical representations of where abnormal calcium and PTH values intersect in standard laboratory reporting. These graphs use arbitrary cutoffs whereas the Bayesian Network utilized in this study was trained on nearly 12,000 examples of normal and abnormal patient laboratory results.

The tool described here uses readily available clinical data to automatically alert the provider when his or her patient exhibits certain laboratory features of PHPT, thereby prompting further evaluation or referral. Such automated alerts can become incorporated into existing electronic medical record systems as “best practices alerts.” Clinical decision-making is a complex task, requiring the physician to integrate multiple pieces of diverse data types, synthesize all this information, and bridge the “inferential gap” between the data at hand and the clinical knowledge needed to make the correct diagnosis or treatment decision^{54,55}. Computerized decision-support tools overcome many of these challenges, and can improve the quality of care through prompts or second opinions in real-time, at the point of care. The other advantage of incorporating ML algorithms into EMRs is the ability to continuously improve with the addition of more data. Hence, our Bayesian Network to recognize PHPT could become even more accurate as it is “trained” on more patients. Clinicians rarely work with complete data sets. Since our classification method displayed excellent accuracy despite the high frequency of missing features, we expect it to maintain its accuracy in real-world, clinical settings.

Limitations

This current study is limited by its retrospective nature; we tried to address this with the large, multi-institutional sample used to train our classifier and, in addition, used three different institutions to investigate for any institutional biases. The study patients also comprised only those referred for surgery which introduces a potential selection bias because our sample could represent those with a more severe presentation that prompted the physicians to recommend surgery rather than continued observation. A third potential limitation is the determination of test-set accuracies by cross-validation within the dataset, thus making it difficult to predict how the algorithms would perform in a different population, although the model was observed to have similar overall accuracy when evaluated across multiple institutions. The STUDY group only included patients from

surgical databases confidently diagnosed with PHPT. Future testing of the tool should include those patients being observed or worked up further for PHPT. Hence, the methods described here will require prospective testing in different patient populations prior to robust clinical implementation. Another potential flaw is our decision to use in just one set of biochemical indices, taken at the time point close to surgery, which was a practical decision for evaluation of the huge masses of laboratory data in this pilot study. Future analysis of this tool should include laboratory values taken at multiple time points to reflect actual clinical practice and potentially increase the accuracy for the most challenging cases. Finally, since this is a surgical cohort, most patients affected with urinary calcium leak or familial hypocalciuric hypercalcemia (FHH) were not included, so there is no way to determine here how well the algorithm would perform on these uniquely challenging cases. However, due to the geographical location of the three examined centers and the case mix of their controls, the training set did include many patients with vitamin D deficiency and hyperthyroidism, which are two common secondary causes of elevations in PTH (21).

CONCLUSIONS

ML is a reliable adjunct for detecting PHPT to aid both specialists and non-specialists alike based on a dataset of 7 readily available clinical variables. As described above, the studied algorithm performed least well on the mild PHPT cases, but the boosted algorithm successfully reduced the misclassification of these most challenging cases. However, our purpose in this initial study was to improve upon the overall recognition of the disease rather than diagnose the most challenging cases that should probably undergo further workup regardless. Similarly, even though the model without PTH was slightly less accurate, it can facilitate the intended goal of prompting further clinical workup including PTH levels. Future studies will prospectively evaluate its performance in real time with the goal of incorporation into electronic medical record systems as clinical decision support for further evaluation or referral when results indicate PHPT is likely.

Acknowledgments

This work was supported by NIH UL1TR000427 and NIH KL2TR000428.

REFERENCES

1. Coker LH, Rorie K, Cantley L, et al. Primary hyperparathyroidism, cognition, and health-related quality of life. *Ann Surg*. 2005; 242(5):642–650. [PubMed: 16244536]
2. Mihai R, Wass JA, Sadler GP. Asymptomatic hyperparathyroidism--need for multicentre studies. *Clin Endocrinol (Oxf)*. 2008; 68(2):155–164. [PubMed: 17651448]
3. Rodgers SE, Lew JI, Solórzano CC. Primary hyperparathyroidism. *Curr Opin Oncol*. 2008; 20(1): 52–58. [PubMed: 18043256]
4. Doppman JL, Miller DL. Localization of parathyroid tumors in patients with asymptomatic hyperparathyroidism and no previous surgery. *J Bone Miner. Res*. 1991; 6(Suppl 2):S153–S158. discussion S159. [PubMed: 1763668]
5. Grant CS, Thompson G, Farley D, van Heerden J. Primary hyperparathyroidism surgical management since the introduction of minimally invasive parathyroidectomy: Mayo Clinic experience. *Archives of surgery*. 2005; 140(5):472–478. discussion 478–479. [PubMed: 15897443]
6. Chen H, Zeiger MA, Gordon TA, Udelsman R. Parathyroidectomy in Maryland: effects of an endocrine center. *Surgery*. 1996; 120(6):948–952. discussion 952–943. [PubMed: 8957479]

7. Yeh MW, Wiseman JE, Ituarte PH, et al. Surgery for primary hyperparathyroidism: are the consensus guidelines being followed? *Annals of surgery*. 2012; 255(6):1179–1183. [PubMed: 22584631]
8. Press DM, Siperstein AE, Berber E, et al. The prevalence of undiagnosed and unrecognized primary hyperparathyroidism: a population-based analysis from the electronic medical record. *Surgery*. 2013; 154(6):1232–1237. discussion 1237–1238. [PubMed: 24383100]
9. Padmanabhan H. Outpatient management of primary hyperparathyroidism. *Am J. Med.* 2011; 124(10):911–914. [PubMed: 21816381]
10. Boonstra CE, Jackson CE. Serum calcium survey for hyperparathyroidism: results in 50,000 clinic patients. *Am J Clin Pathol.* 1971; 55(5):523–526. [PubMed: 5090209]
11. [Accessed August 15, 2015] Health Information Technology for Economic and Clinical Health Act, Title XIII of the American Recovery and Reinvestment Act of 2009 HITECH Ac. 2009. <http://www.gpo.gov/fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf>
12. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J. Med.* 2010; 363(6):501–504. [PubMed: 20647183]
13. Helmons PJ, Suijkerbuijk BO, Nannan Panday PV, Kosterink JG. Drug-drug interaction checking assisted by clinical decision support: a return on investment analysis. *Journal of the American Medical Informatics Association : JAMIA.* 2015; 22(4):764–772. [PubMed: 25670751]
14. Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *Journal of the American Medical Informatics Association : JAMIA.* 1996; 3(6):399–409. [PubMed: 8930856]
15. Balas EA, Weingarten S, Garb CT, Blumenthal D, Boren SA, Brown GD. Improving preventive care by prompting physicians. *Archives of internal medicine.* 2000; 160(3):301–308. [PubMed: 10668831]
16. Denekamp Y. Clinical decision support systems for addressing information needs of physicians. *Isr Med Assoc J.* 2007; 9(11):771–776. [PubMed: 18085031]
17. O'Connell K, Yen TW, Shaker J, Wilson SD, Evans DB, Wang TS. Low 24-hour urine calcium levels in patients with sporadic primary hyperparathyroidism: is further evaluation warranted prior to parathyroidectomy? *American journal of surgery.* 2014
18. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature reviews. Genetics.* 2012; 13(6):395–405.
19. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl.* 2004; 86(5):334–338. [PubMed: 15333167]
20. Lisboa PJ. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw.* 2002; 15(1):11–39. [PubMed: 11958484]
21. Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology.* 2006; 240(3):666–673. [PubMed: 16926323]
22. Kawaler E, Cobian A, Peissig P, Cross D, Yale S, Craven M. Learning to predict post-hospitalization VTE risk from EHR data. *AMIA ... Annual Symposium proceedings / AMIA Symposium.* 2012; 2012:436–445. [PubMed: 23304314]
23. Wilhelm SM, Wang TS, Ruan DT, et al. The American Association of Endocrine Surgeons (AAES) Guidelines for Definitive Management of Primary Hyperparathyroidism. *JAMA Surgery.* 2016 in press.
24. Kahn CE Jr, Laur JJ, Carrera GF. A Bayesian network for diagnosis of primary bone tumors. *J Digit Imaging.* 2001; 14(2 Suppl 1):56–57. [PubMed: 11442121]
25. Kahn CE Jr, Roberts LM, Wang K, Jenks D, Haddawy P. Preliminary investigation of a Bayesian network for mammographic diagnosis of breast cancer. *Proc Annu Symp Comput Appl Med Care.* 1995:208–212. [PubMed: 8563269]
26. Oba S, Sato MA, Takemasa I, Monden M, Matsubara K, Ishii S. A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics.* 2003; 19(16):2088–2096. [PubMed: 14594714]

27. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*. 1997; 55
28. Dutra I, Nassif H, Page D, et al. Integrating machine learning and physician knowledge to improve the accuracy of breast biopsy. *AMIA ... Annual Symposium proceedings / AMIA Symposium*. 2011; 2011:349–355. [PubMed: 22195087]
29. Silverberg SJ. Vitamin D deficiency and primary hyperparathyroidism. *J Bone Miner. Res*. 2007; 22(Suppl 2):V100–V104. [PubMed: 18290710]
30. Bilezikian JP, Brandi ML, Rubin M, Silverberg SJ. Primary hyperparathyroidism: new concepts in clinical, densitometric and biochemical features. *J Intern. Med*. 2005; 257(1):6–17. [PubMed: 15606372]
31. Wu B, Haigh PI, Hwang R, et al. Underutilization of parathyroidectomy in elderly patients with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2010; 95(9):4324–4330. [PubMed: 20610600]
32. Mahadevia PJ, Sosa JA, Levine MA, Zeiger MA, Powe NR. Clinical management of primary hyperparathyroidism and thresholds for surgical referral: a national study examining concordance between practice patterns and consensus panel recommendations. *Endocr Pract*. 2003; 9(6):494–503. [PubMed: 14715476]
33. Applewhite MK, Schneider DF. Mild primary hyperparathyroidism: a literature review. *The oncologist*. 2014; 19(9):919–929. [PubMed: 25063228]
34. Silverberg SJ, Clarke BL, Peacock M, et al. Current issues in the presentation of asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. *The Journal of clinical endocrinology and metabolism*. 2014; 99(10):3580–3594. [PubMed: 25162667]
35. VanderWalde LH, Liu IL, O'Connell TX, Haigh PI. The effect of parathyroidectomy on bone fracture risk in patients with primary hyperparathyroidism. *Archives of surgery*. 2006; 141(9):885–889. discussion 889–891. [PubMed: 16983032]
36. Vestergaard P, Møllerup CL, Frøkjær VG, Christiansen P, Blichert-Toft M, Mosekilde L. Cohort study of risk of fracture before and after surgery for primary hyperparathyroidism. *BMJ*. 2000; 321(7261):598–602. [PubMed: 10977834]
37. Almqvist EG, Becker C, Bondeson AG, Bondeson L, Svensson J. Early parathyroidectomy increases bone mineral density in patients with mild primary hyperparathyroidism: a prospective and randomized study. *Surgery*. 2004; 136(6):1281–1288. [PubMed: 15657588]
38. Hedbäck G, Odén A, Tisell LE. Parathyroid adenoma weight and the risk of death after treatment for primary hyperparathyroidism. *Surgery*. 1995; 117(2):134–139. [PubMed: 7846616]
39. Stefenelli T, Mayr H, Bergler-Klein J, Globits S, Woloszczuk W, Niederle B. Primary hyperparathyroidism: incidence of cardiac abnormalities and partial reversibility after successful parathyroidectomy. *Am J. Med*. 1993; 95(2):197–202. [PubMed: 8356983]
40. Stefenelli T, Abela C, Frank H, Koller-Strametz J, Niederle B. Time course of regression of left ventricular hypertrophy after successful parathyroidectomy. *Surgery*. 1997; 121(2):157–161. [PubMed: 9037227]
41. Ljunghall S, Jakobsson S, Joborn C, Palmér M, Rastad J, Akerström G. Longitudinal studies of mild primary hyperparathyroidism. *J Bone Miner. Res*. 1991; 6(Suppl 2):S111–S116. discussion S121–114. [PubMed: 1763661]
42. Burney RE, Jones KR, Christy B, Thompson NW. Health status improvement after surgical correction of primary hyperparathyroidism in patients with high and low preoperative calcium levels. *Surgery*. 1999; 125(6):608–614. [PubMed: 10372026]
43. Solomon BL, Schaaf M, Smallridge RC. Psychologic symptoms before and after parathyroid surgery. *Am J. Med*. 1994; 96(2):101–106. [PubMed: 8109593]
44. Caron NR, Pasiëka JL. What symptom improvement can be expected after operation for primary hyperparathyroidism? *World journal of surgery*. 2009; 33(11):2244–2255. [PubMed: 19288279]
45. Pasiëka JL, Parsons L, Jones J. The long-term benefit of parathyroidectomy in primary hyperparathyroidism: a 10-year prospective surgical outcome study. *Surgery*. 2009; 146(6):1006–1013. [PubMed: 19958927]

46. Sejean K, Calmus S, Durand-Zaleski I, et al. Surgery versus medical follow-up in patients with asymptomatic primary hyperparathyroidism: a decision analysis. *European journal of endocrinology / European Federation of Endocrine Societies*. 2005; 153(6):915–927.
47. Schneider DF, Burke JF, Ojomo KA, et al. Multigland disease and slower decline in intraoperative PTH characterize mild primary hyperparathyroidism. *Annals of surgical oncology*. 2013; 20(13): 4205–4211. [PubMed: 23943034]
48. Bargren AE, Repplinger D, Chen H, Sippel RS. Can biochemical abnormalities predict symptomatology in patients with primary hyperparathyroidism? *Journal of the American College of Surgeons*. 2011; 213(3):410–414. [PubMed: 21723154]
49. Wallace LB, Parikh RT, Ross LV, et al. The phenotype of primary hyperparathyroidism with normal parathyroid hormone levels: how low can parathyroid hormone go? *Surgery*. 2011; 150(6): 1102–1112. [PubMed: 22136828]
50. Koumakis E, Souberbielle JC, Sarfati E, et al. Bone mineral density evolution after successful parathyroidectomy in patients with normocalcemic primary hyperparathyroidism. *The Journal of clinical endocrinology and metabolism*. 2013; 98(8):3213–3220. [PubMed: 23783096]
51. McCoy KL, Chen NH, Armstrong MJ, et al. The small abnormal parathyroid gland is increasingly common and heralds operative complexity. *World journal of surgery*. 2014; 38(6):1274–1281. [PubMed: 24510243]
52. Jin J, Mitchell J, Shin J, Berber E, Siperstein AE, Milas M. Calculating an individual maxPTH to aid diagnosis of normocalcemic primary hyperparathyroidism. *Surgery*. 2012; 152(6):1184–1192. [PubMed: 23102677]
53. Dhawan A, Wenzel B, George S, Gussak I, Bojovic B, Panescu D. Detection of acute myocardial infarction from serial ECG using multilayer support vector machine. *Conf Proc IEEE Eng Med Biol. Soc.* 2012; 2012:2704–2707. [PubMed: 23366483]
54. Stewart WF, Shah NR, Selna MJ, Paulus RA, Walker JM. Bridging the inferential gap: the electronic health record and clinical evidence. *Health Aff (Millwood)*. 2007; 26(2):w181–w191. [PubMed: 17259202]
55. Sarkar IN, Butte AJ, Lussier YA, Tarczy-Hornoch P, Ohno-Machado L. Translational bioinformatics: linking knowledge across biological and clinical realms. *Journal of the American Medical Informatics Association : JAMIA*. 2011; 18(4):354–357. [PubMed: 21561873]

Figure 1.

Graphical structure of our Bayesian network classifier. A Bayesian network is represented by a directed acyclic graph that indicates the probabilistic relationships among the input features and the outcome of interest. Each arrow, or arc, represents a conditionally dependent relationship. Associated with each node is a conditional probability distribution that represents the distribution of the node variable conditioned on its parents in the graph. In this case, the model is a “naïve” Bayesian network, meaning that each input feature is treated as being independent of the others when conditioned on the outcome variable (PHPT or normal).

Table 1

Patient Demographics and laboratory data.

	PHPT Patients (n = 6777)	Control Patients (n = 5053)	p value
<i>N</i>	6777	5053	-
<i>Age</i>	59.9 ± 0.17	49.8 ± 0.21	< 0.001
<i>Gender (% Female)</i>	76.5%	80.1%	< 0.001
<i>Calcium (mg/dL)</i>	11.3 ± 0.17	9.3 ± 0.007	< 0.001
<i>Phosphate</i>	2.9 ± 0.02	2.1 ± 0.02	< 0.001
<i>PTH (pg/mL)</i>	178 ± 7	41 ± 0.4	< 0.001
<i>Vitamin D (ng/mL)</i>	29 ± 0.17	32.4 ± 0.17	< 0.001
<i>Median Creatinine (mg/dL)</i>	1.0 ± 0.01	0.9 ± 0.01	0.430

Continuous variables are represented by their mean (or median where indicated) ± SEM

Table 2
Accuracy of Bayesian Network Modeling for Predicting Disease within Different Patient Subsets

	Correctly Classified Instances	Sensitivity	Specificity	Positive Predictive Value	Area under ROC
All Patients (N=11830)	95.20%	96.40%	93.70%	95.40%	0.989
<i>Males Only (N=2958)</i>	95.8%	96.8%	94.2%	96.4%	--
<i>Low Vitamin D (N=2613)</i>	95.6%	95.5%	95.8%	97.1%	--
<i>High Vitamin D (N=2601)</i>	93.5%	91.8%	95.3%	95.1%	--
<i>PHPT Patients with Mild Disease (N=1738)</i>	86.0%	--	--	--	--
- <i>Normal Calcium (N=760)</i>	71.1%	--	--	--	--
- <i>Normal PTH (N=1125)</i>	92.1%	--	--	--	--
<i>PHPT Patients with Non-Mild Disease (N=5039)</i>	99.9%	--	--	--	--

Mild labs are defined as a calcium level of <10.2mg/dL and/or a PTH level of <72pg/mL
Low Vitamin D is defined as <30ng/mL

Table 3

Accuracy of Bayesian Network alone versus Adaptive Boosted Modeling

	Bayesian Network Alone	Bayesian Networking + Adaptive Boosting	p-value	
<i>Correctly Classified Instances</i>	95.2%	97.2%	--	
<i>Sensitivity</i>	96.4%	97.8%	--	
<i>Specificity</i>	93.7%	96.4%	--	
<i>Positive Predictive Value</i>	95.4%	97.4%	--	
<i>Area under ROC</i>	0.989	0.994	--	
<i>All Misclassified Instances, N</i>	563/11830	329/11830	< 0.0001*	
<i>False Positives, N</i>	317	180	--	
<i>False Negatives, N</i>	246	149	--	
PHPT Patients w/ Mild Disease (N=1738)	<i>True Positives (Correctly Classified), N (%)</i>	1494 (86.0%)	1598 (91.9%)	< 0.0001*
	<i>False Negatives (Incorrectly Classified), N (%)</i>	244 (14.0%)	140 (8.1%)	

Mild disease defined as a calcium level and/or a PTH level within the normal range

Table 4
Trained adaptive boosted Bayesian network model used to test other institutional cohorts

Training Set = University of Wisconsin (UW) Cohort					
Supplied Test Set	Correctly Classified Instances	Sensitivity	Specificity	Positive Predictive Value	Area under ROC
MCW	94.5%	97.6%	90.8%	92.6%	0.987
PITT	96.8%	99.7%	91.4%	95.4%	0.995
MCW + PITT	96.2%	99.2%	91.2%	94.8%	0.993

MCW- Medical College of Wisconsin cohort
PITT- University of Pittsburgh cohort