Original research

Patient-specific workup of adrenal incidentalomas

Romy R. de Haan[a], Johannes B.R. Visser[b], Ewoud Pons[a,c], Richard A. Feelders[d], Uzay Kaymak[b], M.G. Myriam Hunink[e,f,g], Jacob J. Visser[a,*]

[a] Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands
[b] Industrial Engineering & Innovation Sciences (IE & IS), Eindhoven University of Technology, Eindhoven, The Netherlands
[c] Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands
[d] Department of Internal Medicine, Division of Endocrinology, Erasmus Medical Center, Rotterdam, The Netherlands
[e] Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
[f] Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands
[g] Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, USA

ARTICLE INFO

ABSTRACT

Purpose: : To develop a clinical prediction model to predict a clinically relevant adrenal disorder for patients with adrenal incidentaloma.

Materials and methods: : This retrospective study is approved by the institutional review board, with waiver of informed consent. Natural language processing is used for filtering of adrenal incidentaloma cases in all thoracic and abdominal CT reports from 2010 till 2012. A total of 635 patients are identified. Stepwise logistic regression is used to construct the prediction model. The model predicts if a patient is at risk for malignancy or hormonal hyperfunction of the adrenal gland at the moment of initial presentation, thus generates a predicted probability for every individual patient. The prediction model is evaluated on its usefulness in clinical practice using decision curve analysis (DCA) based on different threshold probabilities. For patients whose predicted probability is lower than the predetermined threshold probability, further workup could be omitted.

Results: : A prediction model is successfully developed, with an area under the curve (AUC) of 0.78. Results of the DCA indicate that up to 11% of patients with an adrenal incidentaloma can be avoided from unnecessary workup, with a sensitivity of 100% and specificity of 11%.

Conclusion: : A prediction model can accurately predict if an adrenal incidentaloma patient is at risk for malignancy or hormonal hyperfunction of the adrenal gland based on initial imaging features and patient demographics. However, with most adrenal incidentalomas labeled as nonfunctional adrenocortical adenomas requiring no further treatment, it is likely that more patients could be omitting from unnecessary diagnostics.

1. Introduction

An adrenal incidentaloma is an adrenal mass detected on imaging studies performed for indications other than to evaluate the adrenal gland [1]. Advances in imaging and the widespread availability of imaging technology has resulted in detection of an increasing number of incidental findings [2]. The prevalence of adrenal incidentalomas found on computed tomography (CT) scans varies from 2.5% to 4% for abdominal CT, and 4.2% for thoracic CT, in adult populations [3–5]. This prevalence increases with age, up to 10% at an age of 70 or above [1,6–8]. As the population ages and the use of imaging technology intensifies, these incidentally discovered tumors would become a more prevalent diagnostic challenge [1,2,9].

Although most adrenal incidentalomas are nonfunctional adrenocortical adenomas requiring no further treatment, diagnostic workup is needed to determine whether these incidental findings are malignant or cause hormonal hyperfunction [2,6]. According to current endocrine and surgical guidelines, every patient is adhered to the same expensive cascade of tests and procedures [10,11]. This results in significant cost. Berland et al. [20,10] recommend biochemical evaluation only in patients with clinical signs or symptoms of adrenal hyperfunction [11]. Furthermore, by evaluation of these incidental findings, patients are exposed to radiation from the CT, which increases the probability of cancer [9]. Each patient has a 1 in 430–2170 chance of developing cancer [9]. Cawood and colleagues showed that this potential risk of cancer is similar to that of the adrenal becoming malignant during the average recommended CT scan follow-up of 3 years.

It is important to identify all adrenal incidentalomas that are

malignant and/or cause hormonal hyperfunction, but at the same time avoiding patients from unnecessary workup. Therefore, clinical management of adrenal incidentaloma should be tailored to the individual patient. Personalized medicine will help produce more efficient and effective diagnoses and treatment, and will lead to better prognoses for patients at both the individual and population level. The availability of patient- and disease-related data in today's healthcare workforce is a significant resource in the development and application of this individual-based approach [12]. Therefore, by using these data preventive or therapeutic interventions can be concentrated on the patients who will benefit, and at the same time sparing expense and side effects for those who will not [13].

The purpose of this retrospective study was to develop a clinical prediction model to predict clinically relevant adrenal disorder for patients with adrenal incidentaloma.

2. Materials and methods

2.1. Patient identification

This retrospective study is approved by the institutional review board, with waiver of informed consent. For filtering of potential adrenal incidentaloma cases, natural language processing (NLP) is used in all thoracic and abdominal CT reports from 2010 till 2012 from a query of a searchable database of a major university-affiliated hospital. All patients with initial presentation or imaging workup for adrenal incidentaloma during this period are identified. Reports containing key sentences that mention the adrenal are selected. Non-relevant anatomic identifiers are removed from key sentences, and frequent negative patterns are subsequently filtered. A machine learning selection algorithm is trained using an annotated set of 500 key sentences and applied as final filter. By applying the NLP pipeline as proposed by Pons and colleagues [14], 139,329 radiology reports were removed from the dataset. Through this pipeline, we managed to increase finding prevalence of potential adrenal incidentaloma [14]. This corresponds with 2122 unique potential adrenal incidentaloma patients (Fig. 1).

2.2. Study population

All adult patients (≥18 years) with an adrenal mass incidentally found on CT are included in this study. Patients with recorded complaints possibly indicating adrenal disease at the moment of finding, patients with a vague described adrenal disorder presumably smaller than 1 cm in diameter, patients with expected metastasis, and patients with a history of metastatic disease are not considered to be incidentalomas and are therefore excluded from this study. Song et al., found no malignant mass among adrenal incidentalomas in low-risk patients, therefore patients with a history of malignancy are included [15]. Patients who were referred from another hospital are excluded, as no data regarding potential work- and follow-up is available in the electronic health record at our hospital. After re-measurements of the remaining incidentalomas, patients with an adrenal mass smaller than 1 cm are excluded.

2.3. Data collection

Patient workup data of all adrenal incidentaloma cases are collected from the electronical medical record by manual investigation of radiology reports, clinical letters and pathology reports. In addition, medication, lab tests and patient data are acquired from the respective hospital databases.

2.4. Data analysis: descriptive statistics

For patients included in the study demographical data, the radiological characteristics of the lesion, values of the biochemical

evaluation, and adrenal disorder are recorded. Because tumor size appeared to be lognormally distributed it is log-transformed. Surface of the nodule is calculated as the surface of an ellipse by using the dimensions of the long en short axis and laterality of the nodule. All patients outcome are verified with the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA).

2.5. Data analysis: predictive modeling

A clinical prediction model is developed to predict if an individual patient is at risk for a clinically relevant adrenal disorder. Stepwise logistic regression is used to construct the prediction model by using age and sex of the patient, the use of anti-hypertensive drugs, history of malignancy, surface, size and laterality of the nodule. For a total of 10 iterations a validation set is constructed based on a 15% hold-out sample with the original class distribution. The remaining data is used for training and tuning of the model using 10-fold cross-validation. The dataset is heavily imbalanced given that just 5% is clinically relevant. To overcome the problem of unbalanced data, the synthetic minority over-sampling technique (SMOTE) is used [16]. Hence a 50-50 class distribution of the target variable is generated by interpolation between existing cases. The prediction model predicts, at the presentation of an adrenal incidentaloma, if a patient is at risk for malignancy or hormonal hyperfunction of the adrenal gland. The model generates a predicted probability for every individual patient.

2.6. Model validation

For validation of the prediction model 10-fold cross-validation is used to better estimate the performance of the model. The performance metrics used are area under the curve (AUC), cohen's kappa and area under kappa (AUK). Cohen's kappa is chosen for two reasons: [1] it takes into account the agreement occurring by chance, and [2] it favors correct classification of the minority class over that of the majority class. These are properties not accounted for by AUC [17]. This is useful when dealing with an imbalanced dataset where it is more important to correctly predict the minority class. The performance of the model is evaluated based on the over-sampled test sets, using a threshold probability of 50%, as well as the validation dataset with original class distribution of the target variable.

2.7. Decision curve analysis

The prediction model is evaluated on its usefulness in clinical practice using decision curve analysis (DCA). DCA is a method for evaluating the benefits of a diagnostic test across a range of patients preferences for accepting risk of under- and overtreatment to facilitate decisions about test selections and use [18]. The prediction model generates a probability of having a clinically relevant adrenal disorder for every individual patient. Additional workup is needed to determine the diagnosis if the likelihood is near one. At some probability between 0 and 1, patients are unsure whether or not the be treated [18]. Threshold probability are determined by a panel of experts and are set on 1%, 1,5% and 2%. For patients whose predicted probability of the prediction model is lower than the predetermined threshold probability, diagnostic workup could be omitted.

3. Results

A total of 2122 potential adrenal incidentalomas are identified by using NLP. Of these, 1487 patients are excluded after manual inspection based on the inclusion and exclusion criteria (Fig. 1). Table 1 displays patient demographics and diagnostic workup features of the final cohort of 635 adrenal incidentaloma patients. Patients are predominantly older, averaging 63 years of age, with more females than males. Almost one-fifth of these patients (N = 118) used anti-hypertensive drugs at

Fig. 1. Flowchart study design.
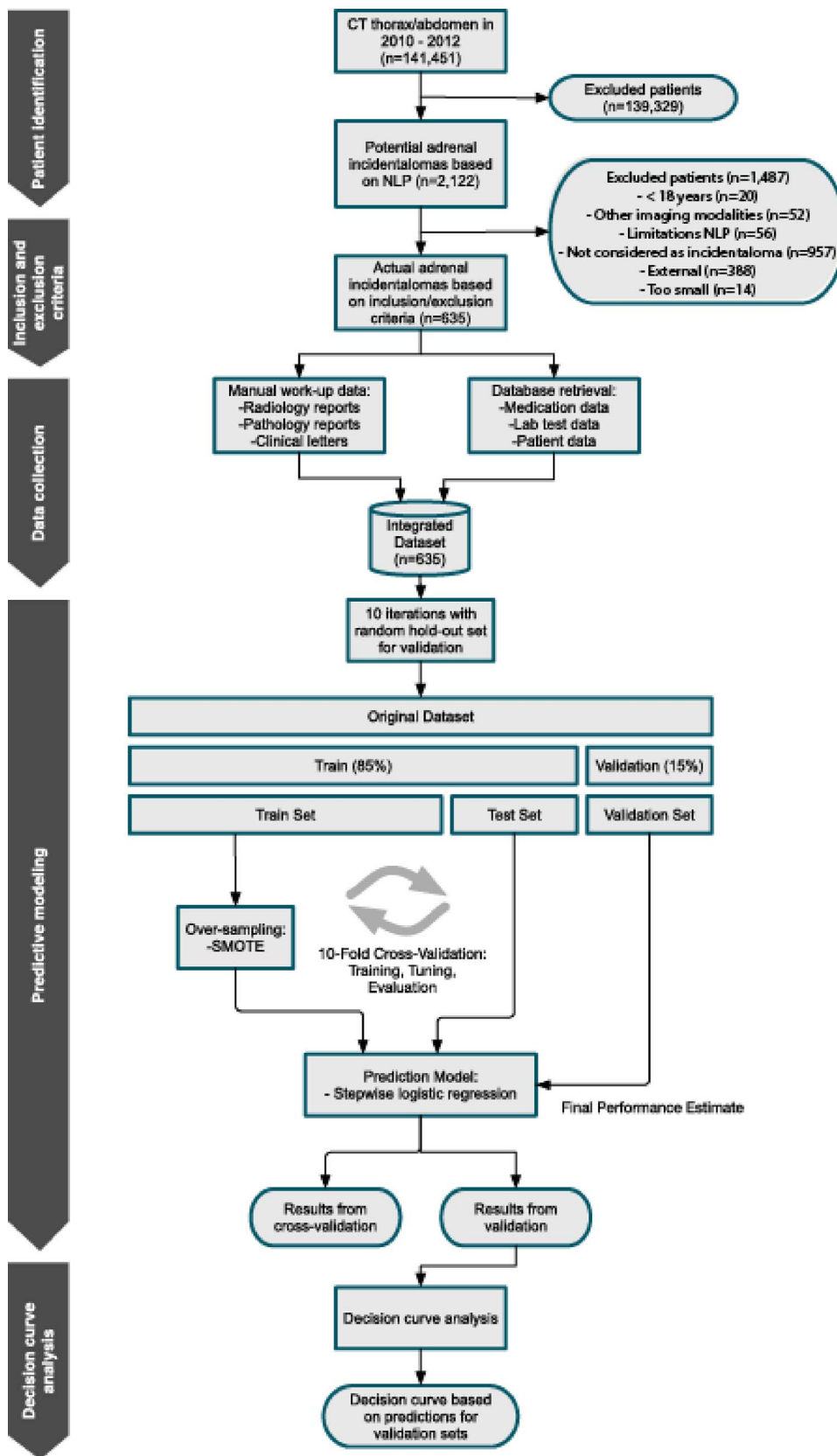
the moment of adrenal incidentaloma presentation, and almost 54% of the patients (N = 341) had a history of malignancy. The median diameter of the lesion was 13 mm with a range of 10–141 mm. Only half of the 187 patients that received workup, received both biochemical screening and imaging workup. 32 of 635 patients (5%) had a clinically relevant adrenal disorder. This included 1 patient with an adrenalcortical carcinoma, 2 patients with a pheochromocytoma, 9 patients with a hormonal hyperfunction, 16 patients whereby growth of ≥ 1 cm of the

Table 1

Patient demographics and initial imaging features (n = 635).

Variable	N (%)
Patient demographics	
Age (years)	62,8 (10.9)*
Range	19–93*
Sex	
Female	294 (53.7)
Male	341 (46.3)
Use of anti-hypertensive drugs	118 (18.6)
History of malignancy	341 (53.7)
Radiological characteristics	
Surface nodule (mm)	267,8 (538,2)*
Range	57,7–12556*
Nodule size (mm)	19,6 (8,9)*
Range Categories nodule size	10–141*
< 40 mm	626 (98.6)
40–60 mm	6 (0,9)
> 60 mm	3 (0,5)
Laterality	
Unilateral	527 (83)
Bilateral	108 (17)
Diagnostic workup	
Biochemical screening	21 (3,3)
Imaging workup (non-contrast CT)	78 (12.3)
Biochemical screening and imaging workup	88 (13.9)
Adrenal disorders	
Adrenalcortical carcinoma	1 (0,2)
Pheochromocytoma	2 (0,3)
Subclinical Cushing's syndrome	4 (0,6)
Cushing's syndrome (subclinical)	4 (0,6)
Primary aldosteronism	1 (0,2)
Growth of ≥ 1 cm	17 (2,7)
Adrenalectomy due to tumor size	3 (0,5)

**Surface of the nodule is calculated as the surface of an ellipse by using the dimensions of the long en short axis.

* These values are presented as mean (± SD).

adrenal gland was determined, and 4 patients who underwent adrenalectomy due to tumor size or radiological characteristics.

3.1. Results of prediction model

Stepwise logistic regression resulted in two different prediction models. The number of variables selected for each of the 10 folds of the 10-fold cross-validation are presented in Fig. 2. All 10 folds contained the variables laterality and surface of the nodule. Half of the 10 folds contained a third variable, i.e. age. The average coefficients are presented in Table 2.
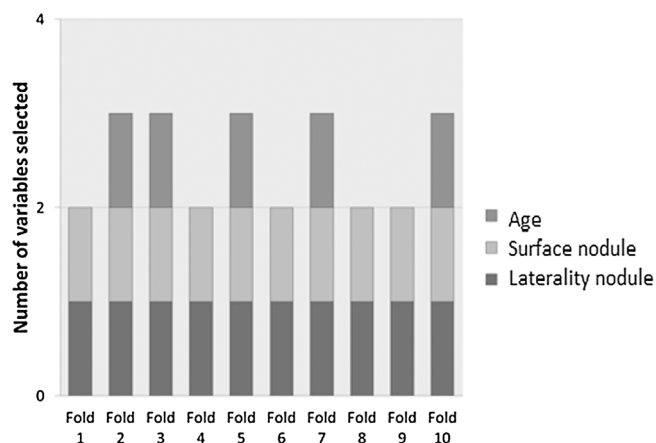

Fig. 2. Selected variables for 10-fold cross-validation on the validation data.

Table 2

Coefficients 10-fold cross-validation on the validation data.

Model	B	Std. Error	t	Sig.
1 (Constant)	−2.550	0.380	−6.710	0.000
laterality	−1.224	0.420	−3.396	0.000
Surface	0.002	0.001	3.488	0.001
2 (Constant)	0.264	3.313	0.199	0.825
laterality	−1.454	0.362	−3.340	0.000
Surface	0.002	0.001	3.415	0.001
Age	−0.043	0.020	−2.146	0.033

a. Dependent variable: clinically relevant adrenal disorder.

3.2. Model validation

ROC and AUK curves are determined and presented in Fig. 3. The area under the curve (AUC) is 0.78 on both the cross-validation and validation data. The results of the area under kappa (AUK) and the Cohen's Kappa are comparable for both the models. The Cohen's kappa of the cross-validation shows that the model performs 13,7% better than if the classification was done only by chance. The average performance metrics after 10-fold cross-validation are presented in Table 3.

3.3. Results of decision curve analysis

The prediction model is evaluated on its usefulness in clinical practice using DCA based on different threshold probabilities. Results of the DCA are presented in Table 4. A higher threshold probability leads to higher savings. However, by increasing the threshold probability from 1,5% to 2%, unnecessary workup is avoided in 23% of the patients, but patients with clinically relevant adrenal disorder are missed. Without omitting any clinically relevant patient, the highest possible threshold probability is 1,7%. Using this threshold probability, unnecessary workup is avoided in 11% of patients with an adrenal incidentaloma, with a sensitivity of 100% and specificity of 11%. Fig. 4 shows an added value of the prediction model at 3%. At a lower threshold probability the model does not differ from treating all patients.

4. Discussion

This research resulted in a prediction model for patient-specific workup of adrenal incidentalomas. Our findings demonstrate that a prediction model for predicting a clinically relevant outcome of adrenal incidentalomas is a solution for avoiding patients from unnecessary workup. Hereby, management of this growing groups of patient can be tailored to the individual patient what will lead to reduced risks and monetary saving [11]. However, the number of patients avoiding from unnecessary workup depends on the predetermined threshold probability of the decision curve analysis (DCA). A higher threshold probability leads to higher savings, but will increase the chance of omitting patients with adrenal disorder, adrenalcortical carcinoma, pheochromocytoma or hyperfunction adrenal lesion, from workup.

A prediction model is successfully developed with an AUC of 0.78, including age and sex of the patient, the use of anti-hypertensive drugs, history of malignancy, surface, size and laterality of the nodule. Results of the DCA indicate that up to 11% of patients with an adrenal incidentaloma can be avoided from unnecessary workup, with a sensitivity of 100%. In contrast to the AACE/AAES guideline, these patients will not receive any workup [19]. By increasing the corresponding threshold probability of 1,7%, patients with clinically relevant adrenal disorder will be missed and omitted from necessary workup, with all its consequences. Furthermore, the DCA shows no added value of the prediction model at this threshold probability, i.e. the model does not
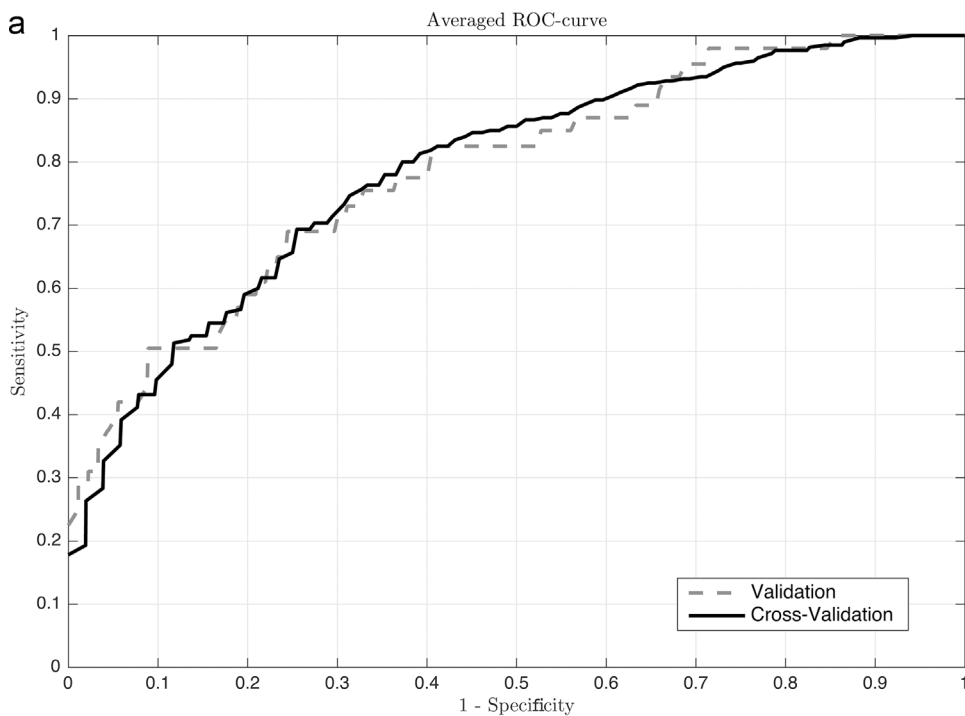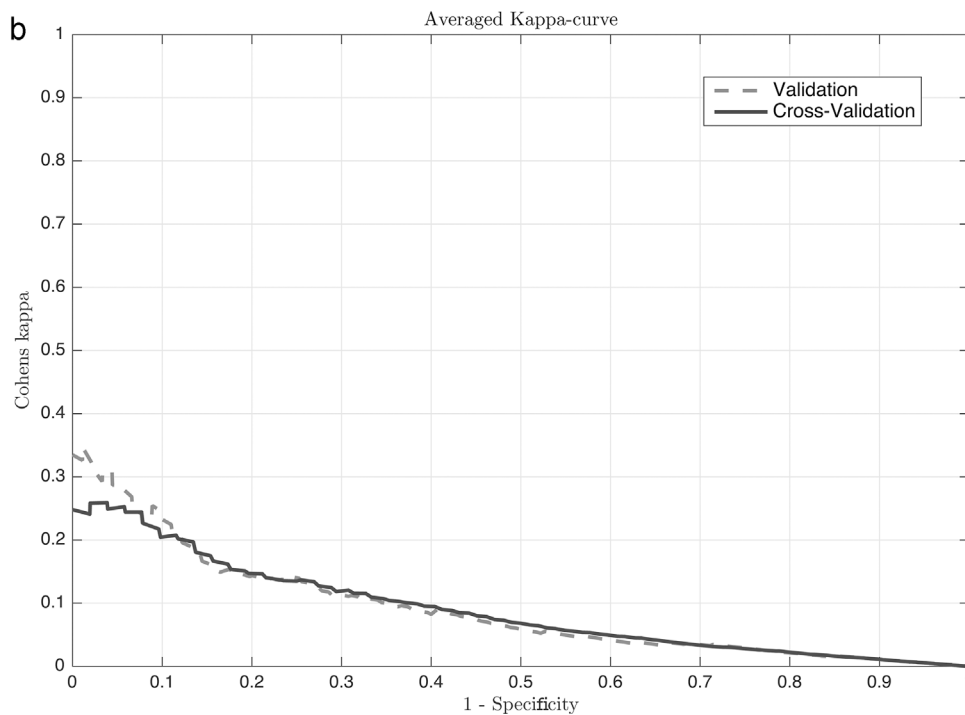
a

Averaged ROC-curve

b

Averaged Kappa-curve

Table 3
Performance metrics.

Observations	Cross-validation	Validation
AUC	0.776	0.776
Kappa	0.137	0.093
AUK	0.084	0.078

Table 4
Results decision curve analysis.

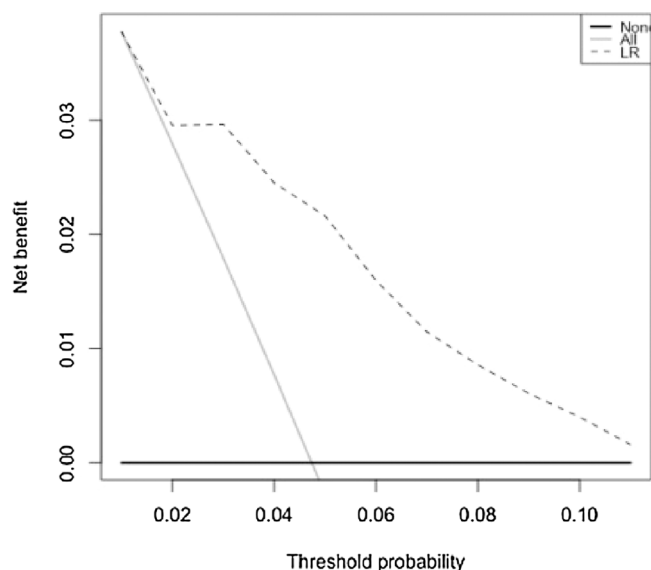| Threshold probability	Patients avoided from workup	Sensitivity	Specificity
1%	2%	100%	2%
1,5%	7%	100%	7%
2%	23%	90%	24%

Fig. 4. Decision curve analysis.
Note − At a threshold probability of 3% or lower, the model does not differ from treating all patients.

differ from treating all patients.

How to workup patients with adrenal incidentaloma is a controversial issue. Workup strategies range from a minimalist approach using a single scan within 6 months while most recommend annual CT scanning for up to 4 or 5 years [20]. There have been attempts to make the current management algorithm more specific by incorporating a risk stratification algorithm to leave less room for subjective decision making. Birsen and colleagues used tumor size and Houndsfield units (HU) density on noncontrast CT as parameter for decision making [21]. In our research we used a prediction model for decision making. The prediction model uses statistical and machine learning algorithms which have been applied successfully in various fields [22].

Our study had an important limitation that is inherent to retrospectively designed studies; i.e. the availability of clinical data. At Erasmus Medical Center management of the adrenal incidentaloma is highly variable in terms of guideline adherence. This resulted in only a small group of patients that received diagnostic workup. Furthermore, not all initial imaging findings were accurate reported in the electronical medical record. To reduce missing data all patients outcome were verified and all adrenal incidentalomas were remeasured. While only 5% of this patient group has a clinically relevant adrenal disorder, it is likely that more patients could be omitting from unnecessary diagnostics. With more imaging features, e.g. tumor size during follow-up and HU of the noncontrast CT, it can be determined whether and how often imaging scanning is necessary during the workup period for every individual patient. In this manner, a better individualized workup can be achieved what results in a more appropriate use of existing diagnostics. Furthermore, the performance on the validation set was highly fluctuating. This is due to the imbalanced distribution of the data. This skewed distribution is inherent in health care, where other models could be more suitable in order to deal with this problem. Moreover, the model was developed and validated using internal data from the Erasmus Medical Center. In order to ensure general applicability of the prediction model external validation is essential.

Not all adrenal incidentaloma patients should adhere to the same cascade of test and procedures. Instead, during the diagnostic workup clinicians should make considered decisions to stop or continue workup for every individual patient. By implementing the prediction model in clinical practice both radiologists and referring physicians can be supported by these decisions.

To support decision making in clinical routine, integration of the prediction model is necessary. This form of clinical decision support can be used to tailor diagnostic algorithms to individual patients and will provide actionable information and suggestions for clinicians, based on evidence rather than intuition or habit [22–25]. This is closely related to the concept of integrated diagnostics, the convergence of imaging, pathology, and laboratory tests with advanced information technology (IT), what could increase the quality and efficiency of healthcare [26]. For the model to become common clinical practice enormous advancement in the IT infrastructure and the EHR are required [23]. For example, structured reporting is needed to share information among disciplines and facilitate the development of reproducible algorithms to integrate data from diverse sources [26].

In conclusion, a prediction model can accurately predict if an adrenal incidentaloma patient is at risk for malignancy or hormonal hyperfunction of the adrenal gland based on initial imaging features and patient demographics. The model can support clinicians to tailor diagnostic workup to the individual patient. However, savings are small. With most adrenal incidentalomas labeled as nonfunctional adrenocortical adenomas requiring no further treatment, it is likely that more patients could be omitting from unnecessary diagnostics.

Acknowledgment

References

[1] G. Mansmann, J. Lau, E. Balk, et al., The clinically inapparent adrenal mass: update in diagnosis and management, Endocr. Rev. 25 (2) (2004) 309–340.

[2] R.H. Grogan, E. Mitmaker, M.R. Vriens, et al., Adrenal incidentaloma: does an adequate workup rule out surprises? Surgery 148 (2) (2010) 392–397.

[3] W.F. Young Jr., Clinical practice: the incidentally discovered adrenal mass, N. Engl. J. Med. 356 (6) (2007) 601–610.

[4] A. Comlekci, S. Yener, S. Ertilav, et al., Adrenal incidentaloma, clinical, metabolic, follow-up aspects: single centre experience, Endocrine 37 (1) (2010) 40–46.

[5] E.V. Ferreira, M.A. Czepielewski, C.S. Faccin, M.C. Accordi, A.P. Furtado, Prevalence of adrenal incidentaloma at computed tomography (chest and abdominal) in a general hospital in Brazil, Arq. Bras. Endocrinol. Metabol. 49 (5) (2005) 769–775.

[6] G. Arnaldi, M. Boscaro, Adrenal incidentaloma, Best Pract. Res. Clin. Endocrinol. Metab. 26 (4) (2010) 405–419.

[7] L. Nieman, Approach to the patient with an adrenal incidentaloma, J. Clin. Endocrinol. Metab. 95 (9) (2010) 4106–4113.

[8] S. Bovio, A. Cataldi, G. Reimondo, et al., Prevalence of adrenal incidentaloma in a contemporary computerized tomography series, J. Endocrinol. Invest. 29 (4) (2006) 298–302.

[9] J.T. Cawood, P.J. Hunt, D. O'Shea, D. Cole, S. Soule, Recommended evaluation of adrenal incidentalomas is costly, has high false-positive rates and confers a risk of fatal cancer that is similar to the risk of the adrenal lesion becoming malignant; time to rethink? Eur. J. Endocrinol. 161 (4) (2009) 513–527.

[10] D. Kastelan, I. Kraljevic, T. Dusek, et al., Clinical course of patients with adrenal incidentaloma: is it time to reconsider the current recommendations? Eur. J. Endocrinol. 2 (2015) 275–282.

[11] L.L. Berland, S.G. Silverman, R.M. Gore, et al., Managing incidental findings on abdominal CT: white paper of the ACR incidental findings committee, J. Am. Coll. Radiol. 7 (10) (2010) 754–773.

[12] E.S. Estape, M.H. Mays, E.A. Sternke, Translation in data mining to advance personalized medicine for health equity, Intell. Inf. Manag. 8 (1) (2016) 9–16.

[13] National Research Council (US) Committee on a Framework for Developing a New Taxonomy of Disease, Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease, National Academies Press, Washington, D.C, 2011.

[14] E. Pons, L. Braun, M. Hunink, J. Kors, Natural language processing in radiology: a systematic review, Radiology 279 (2) (2016) 329–343.

[15] J.H. Song, et al., The incidental adrenal mass on CT: prevalence of adrenal disease in 1049 consecutive adrenal masses in patients with no known malignancy, AJR Am. J. Roentgenol. 190 (5) (2008) 1163–1168.

[16] N.V. Chawla, Data mining for imbalanced datasets: an overview, Data Mining Knowl. Discov. Handbook 85 (2005) 3–86 (7).

[17] U. Kaymak, A. Ben-David, R. Potharst, The AUK. a simple alternative to the AUC, Eng. Appl. Artif. Intell. 5 (2012) 1082–1089.

[18] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making 26 (6) (2006) 565–574.

[19] M.A. Zeiger, G.B. Thompson, Q.Y. Duh, et al., American association of clinical endocrinologists and american association of endocrine surgeons medical guidelines for the management of adrenal incidentaloms: executive summary of

recommendations, Endocr. Pract. 15 (5) (2009) 450–453.

[20] S. McDermott, O.J. O'Connor, C.G. Cronin, M.A. Blake, Radiological evaluation of adrenal incidentalomas − current methods and future prospects, best practice and research, Clin. Endocrinol. Metabol. 26 (no. 1) (2012) 21–33.

[21] O. Birsen, M. Akyuz, C. Dural, et al., A new risk stratification algorithm for the management of patients with adrenal incidentalomas, Surgery 156 (2014) 959–966.

[22] L. Winters-Miner, P. Bolding, J. Hilbe, et al., Practical predictive analytics and decisioning systems for medicine, Acad. Press 18 (2014) 3–18 (4).

[23] B. Regierer, V. Zazzu, R. Sudbrak, A. Kühn, H. Lehrach, Future of medicine: models in predictive diagnostics and personalized medicine, Adv. Biochem. Eng. Biotechnol. 133 (2013) 15–33.

[24] M.G. Hunink, Dynamic Evidence-based Diagnostics. Integrated Diagnostics and Massive Computing: Convergence of Medical Imaging, Laboratory Tests, and IT Solutions, Ninth biennial symposium of the International Society for Strategic Studies in Radiology, 2011, pp. 25–27 (Dubrovnik, Croatia).

[25] J.H. Thrall, The Position of Radiology in an Integrated Diagnostic Center. Integrated Diagnostics and Massive Computing: Convergence of Medical Imaging, Laboratory Tests, and IT Solutions, Ninth biennial symposium of the International Society for Strategic Studies in Radiology, 2011, pp. 25–27 (Dubrovnik, Croatia).

[26] G.P. Krestin, P.A. Grenier, H. Hricak, et al., Integrated Diagnostics, Proceedings from the 9th biennial symposium of the International Society for Strategic Studies in Radiology Aug, Dubrovnik, Croatia, 2011, pp. 25–27.