Atrial fibrillation prediction with longitudinal data

Alexander HUYBRECHTS, Universiteit Gent

Promotor: prof. dr. Bert VAES, KULeuven

Co-promotor: Shauni NUYTS, KULeuven

Master of Family Medicine

Masterproef Huisartsgeneeskunde
Academiejaar: 2019 – 2020

Deze masterproef is een examendocument dat niet werd gecorrigeerd voor eventueel vastgestelde fouten. Zonder voorafgaande schriftelijke toestemming van zowel de promotor(en) als de auteur(s) is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wendt u tot de universiteit waaraan de auteur is ingeschreven.

Voorafgaande schriftelijke toestemming van de promotor(en) is eveneens vereist voor het aanwenden van de in dit afstudeerwerk beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Alexander Huybrechts

Abstract

Introduction The CHARGE-AF score is currently the most accurate predictor of atrial fibrillation (AF). The CHARGE-AF score is a static model, based on the last known data of a patient. However, electronic health records contain a wealth of information, including the evolution of these data over time. This paper attempted to create a better model through analysis of the longitudinal aspect of risk factors. Furthermore, this study examined whether combining the health insurance data with data from the electronic health record provides a benefit regarding the prediction of AF.

Methods The study population was based on the Intego database. In addition, data from the Inter-Mutualistic Agency (IMA) were used. Three models were executed and compared. A fixed model was created, based on the CHARGE-AF variables. In the first longitudinal model (M1), the exact same variables from the CHARGE-AF were used but longitudinally. In the second longitudinal model (M2) the CHARGE-AF variables were included in conjunction with the extra variables from the IMA. The likelihood ratio test (LRT) and Harrell's C-statistic were conducted to evaluate the goodness of fit for these models. The predictive performance was compared using the sensitivity, specificity, balanced accuracy, area under the curve, net reclassification index, integrated discrimination index and net benefit.

Results The p-value of the LRT was smaller than 0,0001 for all models, indicating that the added covariates in the models can be useful. When looking at Harrel's C-index, all models performed well. M2 with the added IMA variables had a higher C-index. When looking at the confusion matrices, M1 performed better when compared with the fixed model. When comparing M2 with the fixed model, similar results were found. According to the decision curve analysis, M2 performs better for a thresholds probability between 0,05 and 0,09.

Conclusion For the first time, data from Intego and IMA were integrated in the search for a better risk model. The new model had a high C-statistic and performed better than the fixed model when looking at the net reclassification index, integrated discrimination index and net benefit. Further research is needed to confirm these results. Parameter selection for the new model with IMA data was made based upon clinical expertise. Several limitations were met, which will need to be addressed in further research.

Introduction

Atrial fibrillation (AF) is the most common type of cardiac arrhythmia, and is associated with serious complications such as stroke and heart failure. (1). These complications can occur even before the diagnosis, because AF can be present asymptomatically for a long time. (2) This emphasizes the importance of the development of screening tools in order to detect affected patients in time. Age is an important risk factor and the increasing life expectancy will cause the population at risk to grow over the coming decades. At the same time complications less often end in fatalities due to better treatment options and therefore patients with AF live longer than they used to. (3)

The CHARGE-AF score is currently the most accurate predictor of AF. (4, 5) It is based on data from prospective cohort studies performed in the US and validated in two European cohort studies.

Based on the last known data of a patient, the risk to develop AF in the upcoming five years can be calculated. (6) If the score shows a high risk for a certain patient, a physician may decide to perform screening ECGs in order to discover a silent arrhythmia. (7, 8) Currently, the CHARGE-AF is not commonly used by general practitioners (GPs) in Belgium, probably due to the rather complex calculation and the lack of implementation in the electronic health record. However, the CHARGE-AF score is a useful research tool for risk stratifying individuals for inclusion in primary prevention trials and evaluating novel risk factors for AF. (5)

The CHARGE-AF score is a static model, based on the last known data of a patient. However, electronic health records contain a wealth of information, including the evolution of these data over time. These data may well hide a pattern that can predict the risk to develop AF even more accurately. Some research has already been done in this field: a recent study showed an association between the change in NT-proBNP over time, and the development of AF. (9, 10)

Furthermore, attempts have been made to develop a better model through analysis of the longitudinal aspect of risk factors. (11, 12, 13) However these studies had some limitations, like the low reporting of AF and missing data regarding mortality.

The current study has attempted to avoid these pitfalls by the imputation of missing data and by establishing a link with a national health insurance database. (14) This study examined whether combining the health insurance data with data from the electronic health record provides a benefit regarding the prediction of AF.

Methods

1. Ethics

The Ethics Committee Research UZ / KU Leuven approved the project "The development of risk prediction algorithms based on real life longitudinal data" on the 15th of October 2018 (reference S61973). This thesis is part of the aforementioned project.

2. Study population

The study population was based on the Intego database. Intego is a general practice-based morbidity registration network by KU Leuven. (11) It contains medical data of a few dozen GPs in Flanders, Belgium, including diagnosis linked to the International Classification of Primary Care – 2nd Edition (ICPC-2). Prescribed drugs were classified according tot the WHO's Anatomical Therapeutic Chemical (ATC) classification system.

In addition, data from the Inter-Mutualistic Agency (IMA) were used. IMA is a joint venture of all Belgian sickness funds, and collects all data on reimbursed healthcare expenses and dispensed medication from public pharmacies, organised according to the ATC classification. Because healthcare insurance is obligatory in Belgium, all patients in Intego have corresponding records in IMA.

The Sectoral Committee on Social Security and Health granted permission to link the data of Intego and IMA. A complex procedure ensured anonymity, which is detailed in the report of the committee. (15)

Only data from 2011 onwards were used, because IMA data before this date were not available for the project. All patients born before 1952, who visited their GP at least once in, or before, 2011 were included.

Three exclusion criteria were used:

- a prior diagnosis of AF (ICPC: K78)
- unaccounted use of anticoagulation, defined as at least two prescriptions of vitamin K antagonists or directly acting oral anticoagulants, without diagnosis of AF, pulmonary embolism, deep-vein thrombosis or mechanical prosthetic heart valve. These patients were possibly being treated for AF without an encoded diagnosis in Intego, and were excluded to prevent bias in the non-AF dataset.
- a stay in a nursing home of any duration. Patients that stayed at least once in a nursing home ("Rustoord voor bejaarden", ROB, or "Rust- en verzorgingstehuis", RVT) between 2011 and 2015 were excluded because these health care records

are often registered on paper and would therefore cause a lot of missing data. The admittance to a nursing home was registered in IMA.

The follow-up period for each patient ended after five years, the censor date or at the time of AF diagnosis.

3. Selected variables

The variables to be tested in the longitudinal model, were mostly based on the variables used in the CHARGE-AF risk score: weight (last of each year), systolic and diastolic blood pressure (BP), use of antihypertensive medication (at least one prescription with ATC code C03, C07, C08 or C09), diabetes (ICPC-2 T89 and T90), heart failure (ICPC-2 K77), and ischaemic events (ICPC-2 K74-76). For blood pressure values, the average of the two lowest values of the last three measurements were selected for each year. Race was not available in Intego or IMA.

In addition, several variables were added to investigate their predictive value in a longitudinal model. The presence of an arterial bypass, balloon angioplasty, total number of GP contacts, total number of days in the hospital, and number of hospital admittances of each year were available for analysis through the IMA database. The total number of days in hospital was categorized in four groups: no admittance in hospital, one day, two to five days, more than five days. Because of the limited research on this topic, these variables were selected after consulting clinicians. These variables may give an impression of the patients' vascular status and general health condition.

4. Data analysis

4.1 Incomplete data

As mentioned before, Intego is a database of electronic health records collected by GPs. This implicates that it is prone to human error and inevitably contains a lot of missing data. To compensate for the missingness, a technique called multiple imputation was used. (16) This technique uses the observed data of patients to estimate realistic numbers for missing values, which can then be used in further analyses. Ideally, these imputations are expected to be done multiple times to correct for estimation bias. However, this is very time consuming and outside the scope of this thesis. Therefore, only a single imputation was done and results should be interpreted with caution.

One can with relative safety assume that the IMA database has no missing values because it is based on financial records of all Belgian health insurers and is therefore expected to be complete and accurate.

4.2 Development of risk prediction models using Cox regression

Three models were executed and compared. A fixed model was created, based on the CHARGE-AF variables. In the first longitudinal model (M1), the exact same variables from the CHARGE-AF were used but longitudinally. In the second longitudinal model (M2) the CHARGE-AF variables were included in conjunction with the extra variables from the IMA, as discussed above. The models used a time-dependent Cox regression. This type of regression is a semi-parametric model and often used for handling survival data. (17) With this model, the dependency of multiple variables on the hazard function was checked. The fixed CHARGE-AF model was fitted using a Cox regression in order to compare the fixed model with the two longitudinal models.

Before starting the analyses and evaluating the models, the dataset was split into a training and test dataset. The training dataset contained 70% off all observations whereas the test data set contained 30%. The training dataset was used to fit the models whereas the test dataset was used to make predictions for individuals.

4.3 Evaluation and comparison of predictive models

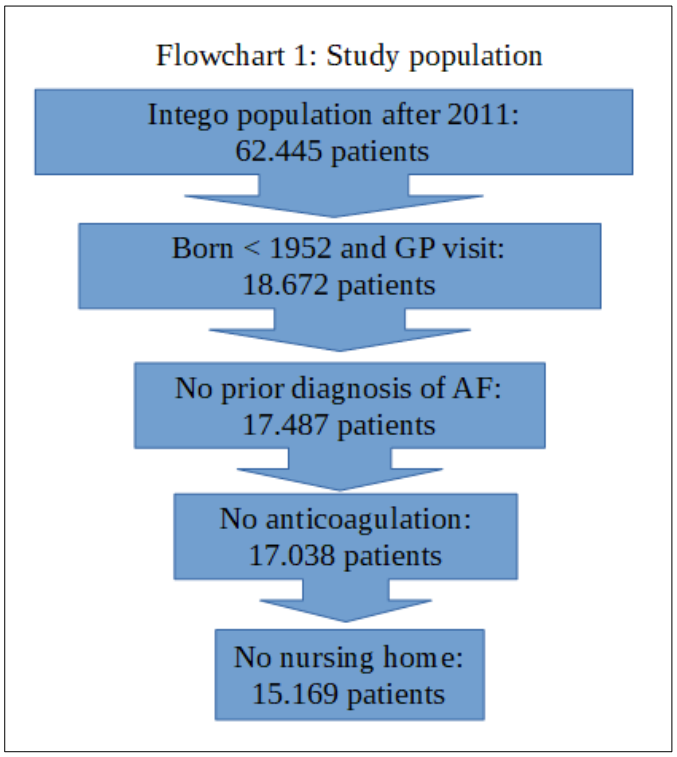
Two tests were conducted to evaluate the goodness of fit for these models. Firstly the likelihood ratio test (LRT) checks the general statistical significance of both models versus the situation where no covariates are included in both models. Secondly, Harrell's C-statistic estimates the probability that when one compares two individuals, the one with the shortest survival time has the highest event hazard. This gives a good indication of the discriminating power of the model. Values between 0.50 and 1 are favourable for survival models. (17)

After fitting and evaluating these models, the predictive performance was compared. The sensitivity, specificity, balanced accuracy and area under the curve (AUC) were calculated in order to evaluate whether adding extra IMA variables increased the predictive power. Because the AUC shows several pitfalls, the net reclassification index (NRI) and integrated discrimination index (IDI) was also calculated to optimally compare the models. The NRI attempts to quantify how well M2 correctly reclassifies subjects, compared to M1. The IDI is the change in sensitivity minus the change in specificity of both models. (18, 19)

Lastly, a decision curve analysis was made. This curve shows the corresponding net benefit (NB) for each threshold probability. The NB takes into account the clinical implications that finding a new event should be weighted against the cost or harm of a wrong diagnosis. For example, by increasing the number of patients to screen for AF, the cost for patients and society increases as well. How many patients would you submit to extra tests to find one more case of AF? For each threshold probability, the NB can be

calculated, resulting in a decision curve analysis. The curves of the different models can be compared.

Results



Study population

18.672 patients form the Intego population were eligible for inclusion. Based on the criteria mentioned above, 3.503 patients were excluded, resulting in a study population of 15.169 patients (Flowchart 1).

The baseline characteristics of the study population are displayed in table 1. This shows that 609 patients (4,40%) developed AF within five years.

Time-dependent Cox regression

A regression analysis was performed, using the Cox proportional hazards model. This was accomplished through the PHREG procedure of SAS. The results for M1 and M2 are displayed in table 2. The p-value of the LRT was smaller than 0,0001 for both models, indicating that the added covariates in both models can be useful. When looking at the results of M2 in detail, many parameters have a significant influence. For example, age has a p-value < 0,0001 and a hazard ratio of 1,062. Therefore, an increase in age of one year will result in a 6,2% increase in the expected hazard (assuming all other variables remain constant).

When looking at Harrel's C-index, all models performed well (C-index with 95% confidence interval respectively 0,705; 0,717 and 0,806 for the fixed model; M1 and M2). M2 with the added IMA variables had a higher C-index, indicating a better discrimination between AF and non-AF cases.

When looking at the predictive performance as displayed in table 3, the accuracy of all models appears high. The balanced accuracy holds into account the large number of non-AF cases and gives a better representation. The balanced accuracy was slightly lower in the fixed model compared to M1 and M2, but the AUC was roughly equivalent in all models.

When looking at the confusion matrices in table 4, M1 performed better when compared with the fixed model: both NRI and IDI were positive (values 95% confidence interval [CI] 0,0141–0,1438 and values 95% CI 0,0091–0,0237, respectively). When comparing M2 with the fixed model, similar results were found (values 95% CI 0,0069–0,1695 and values 95% CI 0,0255–0,0694). The difference in performance of M1 and M2 was less clear: M2 appeared better, but the NRI was not significant (values 95% CI -0,0719–0,0904).

Image 1 shows the decision curve analysis. With thresholds probability between 0,03 and 0,09, all three models perform better than the two naive models (treat all patients and treat no patients). For thresholds probability between 0,02 and 0,05 the fixed model appears to have a higher NB. For higher thresholds probability, M2 performs better. Above a threshold probability of 0,09 the NB becomes lower than zero in M1 and M2 so it would be better not screen patients at this point.

Discussion

The CHARGE-AF score is currently the most accurate predictor of AF, but it is limited by the fact that it is a static risk model. Two new models were created using longitudinal data and new variables. These models showed good results compared to the fixed model.

1. Development of a longitudinal model (M1) and integration of IMA variables (M2)

Existing risk models like CHARGE-AF calculate the risk to develop AF over the course of five years, based on the status of risk factors at baseline. These risk factors evolve over time and some might occur right before the onset of AF, such as heart failure or ischemic heart disease. For the first time, data from Intego and IMA were integrated in the search for a better risk model. The incidence rate in this population was 9/1.000 person-years, similar to what previous studies found for this age group. (20, 21, 22) Parameter selection for the new model with IMA data was rather arbitrary, but nonetheless the new longitudinal model showed good results with a higher C-index (0,806; standard error [SE] 0,01) than the fixed model (0,705; SE 0,008) and M1 (0,717; SE 0,011). Similar results were found in a previous study to validate CHARGE-AF in Intego with a fixed (C-index 0,681) and longitudinal model (C-index 0,736). (12) Compared to this previous study, patients with unaccounted use of anticoagulation were excluded. This might explain the slightly better performance of the fixed model. All clinical evaluation methods (NRI, IDI, NB) showed that M2 performed better than the fixed model, at least with a threshold probability between 0,05 and 0,08. AF is asymptomatic and often paroxysmal, so screening for AF implies prolonged or repeated heart rhythm monitoring

in patients without symptoms at that time. Therefore it is important to make a good selection of patients to screen for AF. These results showed that it might be beneficial to use longitudinal data to identify patients at risk. The risk calculation is rather complicated compared to easy to remember mnemonics like the CHA_2DS_2-VASC score for stroke risk in AF. (5) However, it could be implanted in the electronic health record to facilitate the use in clinical practice. The role of the additional IMA data is less clear. Further research is needed to evaluate the impact.

2. Limitations

Several limitations were confronted during this study. First, not all CHARGE-AF variables were available from Intego (or IMA), so the reported models are an approximation. For example, race is not registered. Smoking was registered but due to the amount of missing values, it was not reliable as a parameter. This should be addressed in future research. Second, Intego is a primary care registry and it is prone to human error and contains a lot of missing data despite the effort of the participating GP's. As explained before, these missing data can be imputed. In order to correct possible biases in the imputation process, the imputations should be done multiple times, creating multiple datasets and resulting in a more reliable model. Due to human error it is for example possible that patients with AF do not get a registered diagnosis and end up in the control group. The exclusion criterium of unaccounted use of anticoagulation tries to compensate for this error, but it is not possible to catch all mistakes. Furthermore, patients that were registered with unaccounted use of anticoagulation during the follow-up period were not excluded. In further research, a procedure to handle these cases should be investigated.

3. Strengths

Intego is a large registry, accounting for 1,9% of the Flemish population and representative for age and sex. (11) The combination with the IMA data is unique and offers many more and accurate data, which may improve the predictive value of the model.

Conclusion

This study used Intego and IMA data to develop a longitudinal risk prediction model for AF. This model has a high C-statistic and performs better than the fixed model when looking at the net reclassification index, integrated discrimination index and net benefit. Further research is needed to confirm these results.

Samenvatting

Achtergrond De CHARGE-AF score is momenteel de meest accurate voorspeller van voorkamerfibrillatie (VKF). De CHARGE-AF score is een statisch model, gebaseerd op de laatst bekende data. Echter bevat het elektronisch patiëntendossier een schat aan informatie die een meerwaarde kan vormen, inclusief de evolutie van parameters in de tijd. Deze studie probeerde een beter model te ontwikkelen door het analyseren van het longitudinaal aspect van risicofactoren. Verder werd onderzocht of het combineren van data uit het patiëntendossier met data van de ziekenfondsen een betere predictie van VKF oplevert.

Methoden De studiepopulatie is gebaseerd op de Intego database. Ook data van het InterMutualistisch Agentschap (IMA) werd gebruikt. Drie modellen werden opgesteld: een statisch model gebaseerd op de CHARGE-AF variabelen, een eerste longitudinaal model (M1) met exact dezelfde variabelen maar dan longitudinaal, en een tweede longitudinaal model (M2) met extra variabelen van IMA. De *likelihood ratio test* (LRT) en *Harrell's C-statistic* werden berekend om de *goodness of fit* te bepalen. De *predicitive performance* werd vergeleken aan de hand van de sensitiviteit, specificiteit, gebalanceerde accuraatheid, *area under the curve*, *net reclassification index* (NRI), *integrated discrimination index* (IDI) en *net benefit* (NB).

Resultaten De p-waarde van de LRT was kleiner dan 0,0001 voor alle modellen, wat er op wijst dat de toegevoegde covariaten nuttig kunnen zijn. *Harrel's C-index* was goed in alle modellen en M2 had de hoogste C-index. Volgens de berekende *confusion matrices* was de performantie van M1 en M2 afzonderlijk beter dan het statische model. Volgens de *decision curve analysis* was M2 beter voor een *threshold probability* tussen 0,05 en 0,09.

Conclusie Voor het eerst werden data van Intego en IMA gecombineerd in de zoektocht naar een beter risicomodel. Dit nieuwe model had een hoge C-index en presteerde beter dan het statisch model volgens de NRI, IDI en NB. Verder onderzoek is nodig om deze resultaten te bevestigen. De selectie van parameters uit IMA werd gebaseerd op klinische expertise. Er waren wel enkele beperkingen aan deze studie, die in verder onderzoek aangepakt moeten worden.

Table 1: Baseline characteristics in 2011

	Overall (n = 15.169)	**AF** (n = 609)	**Non-AF** (n = 14.560)
Age (years)	70,3 (7,7)	74,0 (7,8)	70,1 (7,7)
Height (cm)	165,6 (9,3)	166,6 (9,4)	165,6 (9,3)
Weight (kg)	75,3 (15,0)	77,9 (16,2)	75,2 (14,9)
Systolic BP (mmHg)	131,4 (14,0)	132,0 (15,2)	131,3 (13,9)
Diastolic BP (mmHg)	77,6 (8,7)	76,0 (8,4)	77,6 (8,7)
HDL cholesterol (mg/dL)	58,0 (16,4)	55,9 (16,4)	58,1 (16,4)
LDL cholesterol (mg/dL)	117,3 (35,1)	110,2 (35,5)	117,6 (35,1)
Triglyceride (mg/dL)	128,4 (73,2)	127,5 (70,2)	128,4 (73,3)
Total cholesterol (mg/dL)	199,8 (40,1)	191,0 (41,2)	200,2 (40,0)
Creatinine (mg/dL)	0,9 (0,3)	1,0 (0,3)	0,9 (0,3)
Glucose (mg/dL)	101,1 (23,8)	105,0 (26,5)	101,0 (23,7)
Heartbeat (bpm)	73,0 (11,9)	71,9 (12,6)	73,0 (11,9)
Hospital visits	0,6 (1,7)	0,7 (1,2)	0,6 (1,7)
GP visits	2,0 (5,1)	3,6 (7,1)	0,7 (1,2)
Diabetes, n (%)	2.339 (15,4)	136 (22,3)	2.203 (15,1)
Ischemic heart disease, n (%)	1.544 (10,2)	110 (18,1)	1.434 (9,8)
Heart failure, n (%)	269 (1,8)	34 (5,6)	233 (1,6)
Valvular heart disease, n (%)	522 (3,4)	54 (8,9)	468 (3,2)
Stroke, n (%)	442 (2,9)	41 (6,7)	401 (2,8)

Antihypertensive drug, n (%)	7.842 (51,7)	423 (69,5)	7.419 (51,0)
Arterial bypass, n (%)	13 (0,1)	0 (0)	13 (0,1)
Balloon angioplasty, n (%)	71 (0,5)	5 (0,8)	66 (0,5)
Days in hospital (0), n (%)	11.084 (73,1)	384 (63,1)	10.700 (73,5)
Days in hospital (1), n (%)	1.142 (7,5)	43 (7,1)	1.099 (7,5)
Days in hospital (2 – 5), n (%)	1.417 (9,3)	82 (13,5)	1.335 (9,2)
Days in hospital (> 5), n (%)	1.526 (10,1)	100 (16,4)	1.426 (9,8)

Values correspond to mean (standard error) or number of cases (percent).
AF: patients who develop AF over the course of the study; Non-AF: patients who do not develop AF over the course of the study.

Table 2: Analysis of maximum likelihood estimates

Model 1

Parameter	**Parameter estimate**	**Standard error**	**Chi-square**	**P-value**	**Hazard ratio (95% CI)**
Age	0,06949	0,00670	107,5543	<0,0001	1,072 (1,0580; 1,0861)
Diastolic BP	-0,01001	0,00680	2,1696	0,1408	0,990 (0,9769; 1,0033)
Systolic BP	-0,01712	0,00423	16,4092	<0,0001	0,983 (0,9749; 0,9912)
Height	0,03034	0,00622	23,8257	<0,0001	1,031 (1,0183; 1,0434)
Weight	0,00193	0,00375	0,2658	0,6062	1,002 (0,9946; 1,0093)
Diabetes	-0,33815	0,11370	8,8448	0,0029	0,713 (0,5706; 0,8911)
Ischemic heart disease	-0,35531	0,12190	8,4962	0,0036	0,701 (0,5520; 0,8901)
Heart failure	-1,34850	0,16466	67,0729	<0,0001	0,260 (0,1880; 0,3585)
Antihypertensive drug	-1,25693	0,13627	85,0741	<0,0001	0,285 (0,2178; 0,3716)

Model 2

Parameter	Parameter estimate	Standard error	Chi-square	P-value	Hazard ratio (95% CI)
Age	0,05987	0,00694	74,3297	<0,0001	1,062 (1,0473; 1,0762)
Diastolic BP	-0,00458	0,00677	0,4588	0,4982	0,995 (0,9868; 1,0134)
Systolic BP	-0,01393	0,00413	11,3523	0,0008	0,986 (0,9782; 0,9942)
Height	0,03108	0,00622	24,9249	<0,0001	1,032 (1,0191; 1,0442)
Weight	0,00214	0,00373	0,3286	0,5665	1,002 (0,9948; 1,0095)
Hospital visits	-0,04550	0,03293	1,9099	0,1670	0,956 (0,8958; 1,0192)
GP visits	0,01944	0,00575	11,4260	0,0007	1,020 (1,0082; 1,0312)
Diabetes	-0,24751	0,11376	4,7337	0,0296	0,781 (0,6247; 0,9758)
Ischemic heart disease	-0,19329	0,12185	2,5162	0,1127	0,824 (0,6491; 1,0466)
Heart failure	-1,06100	0,16364	42,0409	<0,0001	0,346 (0,2511; 0,4770)
Antihypertensive drug	-1,19062	0,13636	76,2396	<0,0001	0,304 (0,2327; 0,3972)
Arterial bypass	0,83577	1,00703	0,6888	0,4066	2,307 (0,3205; 16,6024)
Balloon angioplasty	-0,37714	0,27953	1,8203	0,1773	0,686 (0,3965; 1,1862)
Days in hospital (1)	1,19351	0,17718	45,3755	<0,0001	3,299 (2,3309; 4,6682)
Days in hospital (2 - 5)	1,61996	0,15205	113,5132	<0,0001	5,053 (3,7507; 6,8072)
Days in hospital (> 5)	2,19348	0,14878	217,3646	<0,0001	8,966 (6,6984; 12,0022)

CI: confidential interval

Table 3: Confusion matrices

	Fixed model			**Model 1**		**Model 2**	
	Truth			Truth		Truth	
Predicted		**0**	**1**	**0**	**1**	**0**	**1**
	0	3865	132	3470	101	3415	97
	1	503	51	898	82	953	86

	Fixed model	**Model 1**	**Model 2**
Accuracy	0,8605	0,7805	0,7693
Accuracy 95% CI	(0,8501; 0,8704)	(0,7682; 0,7924)	(0,7568; 0,7815)
Sensitivity	0,8848	0,7944	0,7818
Specificity	0,2790	0,4481	0,4700
Balanced Accuracy	0,5818	0,6213	0,6259
AUC	0,6885	0,6823	0,6803

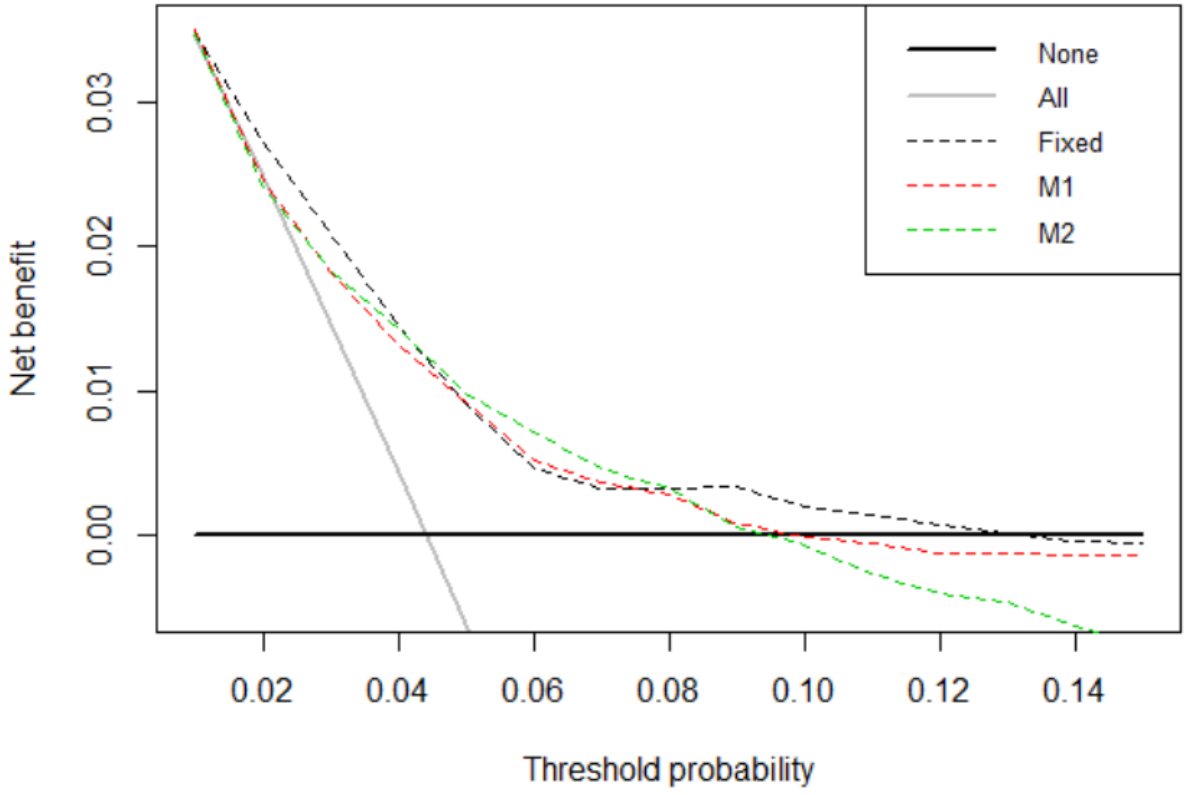
CI: confidential interval; AUC: area under the curve

Table 4: Reclassification tables

	Fixed versus M1	**Fixed versus M2**	**M1 versus M2**
NRI (95% CI)	0,0790 (0,0141 – 0,1438)	0,0882 (0,0069 – 0,1695)	0,0093 (-0,0719 – 0,0904)
NRI p-value	*0,0170*	*0,0334*	*0,8229*
IDI (95% CI)	0,0164 (0,0091 – 0,0237)	0,0475 (0,0255 – 0,0694)	0,0310 (0,0120 – 0,0501)
IDI p-value	*< 0,0001*	*< 0,0001*	*0,0014*

NRI: net reclassification index; IDI: integrated discrimination index; M1: Model 1; M2: Model 2; CI: confidence interval

Image 1: Decision curve analysis for all curves



References

1. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation. 1998 Sep 8;98(10):946-52.

2. Friberg L, Hammar N, Pettersson H, Rosenqvist M. Increased mortality in paroxysmal atrial fibrillation: report from the Stockholm Cohort-Study of Atrial Fibrillation (SCAF). Eur Heart J. 2007 Oct;28(19):2346-53.

3. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol. 2014;6:213-20.

4. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet. 2009 Feb;2(1):73-80.

5. Christophersen IE, Yin X, Larson MG, Lubitz SA, Magnani JW, McManus DD, et al. A comparison of the CHARGE-AF and the CHA2DS2-VASc risk scores for prediction of atrial fibrillation in the Framingham Heart Study. Am Heart J. 2016 Aug;178:45-54.

6. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. J Am Heart Assoc. 2013 Mar 18;2(2):e000102.

7. Vandenberk T, Stans J, Mortelmans C, Van Haelst R, Van Schelvergem G, Pelckmans C, et al. Clinical Validation of Heart Rate Apps: Mixed-Methods Evaluation Study. JMIR Mhealth Uhealth. 2017 08 25;5(8):e129.

8. Vaes B, Stalpaert S, Tavernier K, Thaels B, Lapeire D, Mullens W, et al. The diagnostic accuracy of the MyDiagnostick to detect atrial fibrillation in primary care. BMC Fam Pract. 2014 Jun 9;15:113.

9. Li L, Selvin E, Lutsey PL, Hoogeveen RC, O'Neal WT, Soliman EZ, et al. Association of N-terminal pro B-type natriuretic peptide (NT-proBNP) change with the risk of atrial fibrillation in the ARIC cohort. Am Heart J. 2018 10;204:119-27.

10. Svennberg E, Lindahl B, Berglund L, Eggers KM, Venge P, Zethelius B, et al. NT-proBNP is a powerful predictor for incident atrial fibrillation - Validation of a multimarker approach. Int J Cardiol. 2016 Nov 15;223:74-81.

11. Department of general practice, KU Leuven. Intego-project. [Internet]. 2011 [cited 20/08/2019]. Available from: https://www.intego.be

12. Blykers J, Vaes B. Atrial fibrillation prediction with longitudinal data: insights from a general practice registry; Master of Family Medicine, KU Leuven. 2018 (Unpublished).

13. Izquierdo Juncás D, Molenberghs G, Mamouris P, Vaes B. Survival Analysis in the Intego Database. Master in statistics, KU Leuven. 2018 (Unpublished).

14. Het InterMutualistisch Agentschap. Data. [Internet] [cited 11/10/2019]. Available from: https://ima-aim.be/Data-209

15. Sectoraal Comité van de Sociale Zekerheid en van de Gezondheid. Beraadslaging nr. 17/071 van 19 september 2017 betreffende de koppeling van gecodeerde persoonsgegevens die de gezondheid betreffen tussen het huisartsen morbiditeitsregister Intego en de databank van het InterMutualistisch Agenschap (IMA). Reference SCSZG/17/167.

16. Rubin DB. Multiple imputation for nonresponse in surveys. Department of Statistics, Harvard University; 1987.

17. Harrell FE. Regression Modeling Strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer, New York; 2016.

18. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008 Jan 30;27(2):157-72; discussion 207-12.

19. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. Am J Epidemiol. 2012 Sep 15;176(6):473-81.

20. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BH, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. Eur Heart J. 2006 Apr;27(8):949-53.

21. Psaty BM, Manolio TA, Kuller LH, Kronmal RA, Cushman M, Fried LP, et al. Incidence of and risk factors for atrial fibrillation in older adults. Circulation. 1997 Oct 7;96(7):2455-61.

22. Krahn AD, Manfreda J, Tate RB, Mathewson FA, Cuddy TE. The natural history of atrial fibrillation: incidence, risk factors, and prognosis in the Manitoba Follow-Up Study. Am J Med. 1995 May;98(5):476-84.