UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI MEDICINA E CHIRURGIA

Dottorato di ricerca in EPIDEMIOLOGIA, AMBIENTE E SANITA' PUBBLICA

XXX Ciclo

Syncope risk stratification in the Emergency Department: comparison of different prediction models and the possible role of attribute matching

Tesi di dottorato di:

Monica Solbiati

Matricola: R10909

Tutor: Prof. Piergiorgio Duca

Coordinatore del corso: Prof. Carlo La Vecchia

Anno accademico 2016-2017

TABLE OF CONTENTS

English abstract **1**

Italian abstract **2**

Background **3**

Syncope 3

Definition 3

Classification and pathophysiology 3

Epidemiology and prognosis 5

Risk stratification 6

Prognostic models 6

Logistic regression 8

Recursive partitioning 9

Artificial neural networks 10

Clinical decision tools in syncope 10

OESIL risk score 10

EGSYS score 11

San Francisco Syncope Rule 11

ROSE rule 12

Boston syncope criteria 12

Canadian Syncope Risk Score 12

Limits of the currently available CDTs 13

Aims **14**

Materials and methods **15**

Data collection 15

Individual patients' dataset 15

Prospective dataset 15

Definitions 16

Data analysis 17

Logistic regression analysis 17
Artificial neural networks 17
Results 19
Retrospective analysis 19
Individual patient data dataset 19
Logistic regression 23
Artificial Neural Networks 25
Prospective validation 26
Prospective database 26
Logistic regression 30
Artificial Neural Networks 31
Possible future perspectives: attribute matching 31
Comparison with the previously derived methods 34
Prospective validation of the attribute matching 39
Some practical examples 41
Discussion 45
Study limitations 47
Conclusions and future perspectives 48
References 49

ENGLISH ABSTRACT

Clinical prediction tools have failed in the correct risk stratification of syncope patients in the emergency department. To assess the possible strengths and weaknesses and to compare the different statistical methodologies to derive prediction tools, we decided to derive both a multivariate logistic regression model and an artificial neural network (ANN) on a large retrospective database and to prospectively validate them in a new dataset of 354 patients. The area under the ROC curve of multivariate regression and ANN in the validation cohort were 0.726 and 0.694, respectively. Since the poor predictive accuracy of the analyzed models, we tried to identify alternative methods. We hypothesized that accurate pretest probability assessments can be obtained by matching an individual patient to a group of previously studied patients who shared the same clinical characteristic and determining the percentage of these previously studied patients who had the outcome of interest. In theory, the ideal attribute matching system would allow a very detailed clinical profile to be matched against a very large reference database to provide accurate risk estimates. Therefore, we do not offer a clinically useful prediction tool at this stage, but this method seems promising. Future studies should focus on building large prospective datasets to assess if attribute matching adds any value to both the traditional clinical decision tools and the implicit estimate of probability from clinicians. Moreover, the introduction of new and more complex input attributes and the possibility to provide as output a detailed risk assessment will create a more specific and potentially more accurate clinical profile.

ITALIAN ABSTRACT

Gli strumenti fino ad ora disponibili hanno fallito nel tentativo di predire gli eventi avversi dei pazienti con sincope in pronto soccorso. Per valutare i possibili punti di forza e di debolezza e confrontare i diversi metodi statistici usati per la derivazione di score e scale di rischio, abbiamo deciso di derivare dei modelli basati su regressione logistica multivariata e reti neurali artificiali (ANN) a partire da un database retrospettivo e di validarli in un nuovo dataset di 354 pazienti. L'area sotto la curva ROC di regressione multivariata e ANN è risultata rispettivamente di 0.726 e 0.694 nella coorte di validazione. Vista la bassa accuratezza predittiva dei modelli analizzati, abbiamo provato a identificare metodi alternativi per predire il rischio di eventi avversi dei pazienti con sincope. Abbiamo ipotizzato che si potesse stimare una probabilità pre-test accurata appaiando ogni singolo paziente ad un gruppo di pazienti con le stesse caratteristiche cliniche, e valutando la proporzione che aveva sviluppato l'outcome di interesse (*attribute matching*). In teoria, il funzionamento ideale di tale sistema consentirebbe di confrontare un profilo clinico tanto più dettagliato e di avere stime di rischio molto precise quanto più il database di riferimento è ampio. Pertanto, questo strumento non è ancora utilizzabile nella pratica clinica, ma sembra un metodo promettente. Studi futuri dovrebbero costruire database prospettici che arruolino un gran numero di pazienti per valutare se l'*attibute matching* possa aggiungere informazioni agli strumenti predittivi tradizionali e alla stima del rischio da parte di medici esperti. Inoltre, la possibilità di introdurre variabili nuove e sempre più complesse e di poter predire nel dettaglio diversi tipi di outcome, potrebbe creare un profilo clinico più specifico e potenzialmente sempre più accurato.

BACKGROUND

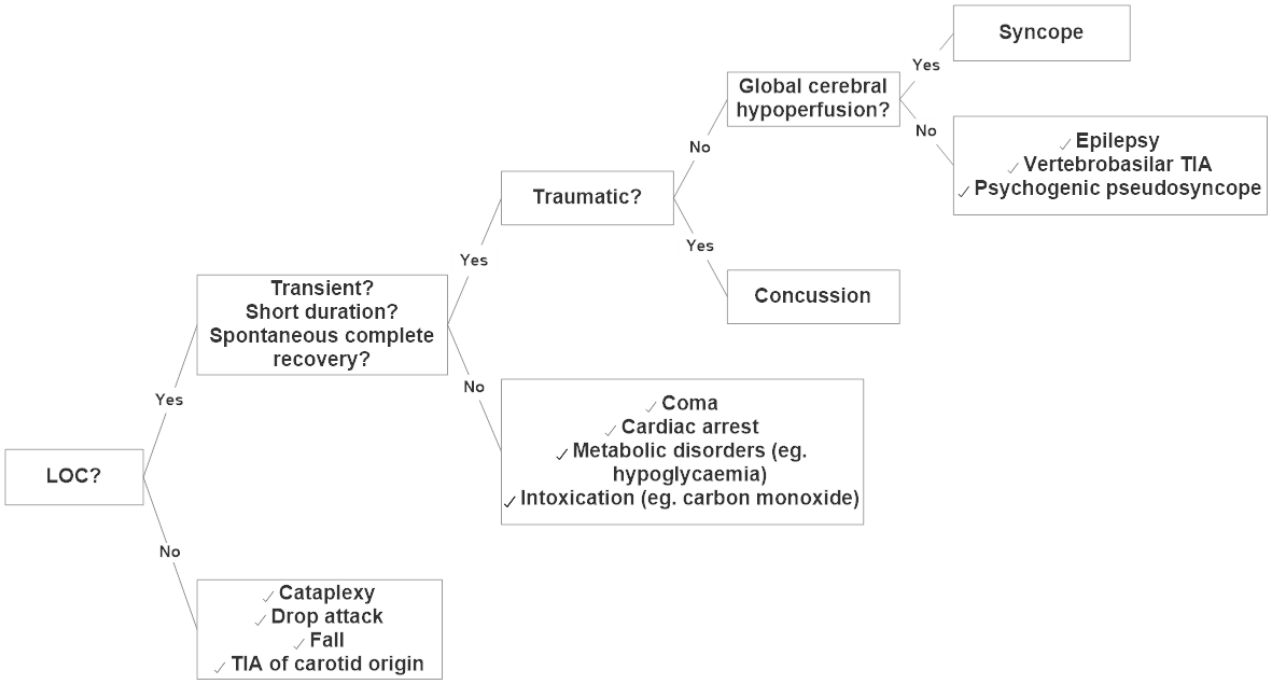
Syncope

Definition

Syncope is defined as a transient loss of consciousness (T-LOC) characterized by rapid onset, short duration, and spontaneous complete recovery. The underlying pathophysiological mechanism is transient global cerebral hypoperfusion. In some forms of syncope there may be a prodromal period in which various symptoms (e.g. lightheadedness, nausea, sweating, weakness, and visual disturbances) warn that syncope is imminent. Often, however, LOC occurs without warning. The noun “pre-syncope” or “near-syncope” is used often to describe a state that resembles the prodrome of syncope but which is not followed by LOC [1].

Several disorders may resemble syncope in two different ways. In some, consciousness is truly lost, but the mechanism is something other than global cerebral hypoperfusion. Examples are epilepsy, several metabolic disorders (including hypoxia and hypoglycaemia), intoxication, and vertebrobasilar transient ischaemic attack (TIA). In other disorders, consciousness is only apparently lost; this is the case in cataplexy, drop attacks, falls, psychogenic pseudosyncope, and TIA of carotid origin (Figure 1).

Figure 1 - Syncope in the context of T-LOC



Classification and pathophysiology

The European Society of Cardiology suggests a pathophysiological classification of the main causes of syncope (Table 1) [1].

Table 1 - Pathophysiological classification of the main causes of syncope

Classification of syncope
Reflex (neuromediated) syncope
Vasovagal
Mediated by emotional stress (fear, pain, blood, instrumentation)
Mediated by orthostatic stress
Situational
Cough, sneeze
Gastrointestinal stimulation (swallow, defecation, visceral pain)
Micturition
Post-exercise
Post-prandial
Others (i.e. laugh, brass instrument playing, weightlifting)
Carotid sinus syncope
Atypical forms (without apparent triggers and/or atypical presentation)
Syncope due to orthostatic hypotension
Primary autonomic failure
Pure autonomic failure, multiple system atrophy, Parkinson's disease with autonomic failure, Lewy body dementia
Secondary autonomic failure
Diabetes, amyloidosis, uremia, spinal cord injuries
Drug-induced orthostatic hypotension
Alcohol, vasodilators, diuretics, phenotiazines, antidepressants
Volume depletion
Hemorrhage, diarrhea, vomiting, etc
Cardiac syncope (cardiovascular)
Arrhythmia as primary cause
Bradycardia
Sinus node dysfunction (including bradycardia/tachycardia syndrome)
Atrioventricular conduction system disease
Implanted device malfunction
Tachycardia
Supraventricular
Ventricular (idiopathic, secondary to structural heart disease or channelopathies)
Drug-induced bradycardia and tachyarrhythmias

Structural disease	
	Cardiac: cardiac valvular disease, acute myocardial infarction/ischemia, hypertrophic cardiomyopathy, cardiac masses, pericardial disease/tamponade, congenital abnormalities of coronary arteries, prosthetic valves dysfunction
	Others: pulmonary embolus, acute aortic dissection, pulmonary hypertension

Modified from: "Moya A, et al. Guidelines for the diagnosis and management of syncope (version 2009)"

From a pathophysiological point of view, syncope is caused by a reduction global cerebral perfusion provoked by a drop in systemic blood pressure. A sudden cessation of cerebral blood flow for as short as 6–8 s or a decrease in systolic blood pressure to 60 mmHg or lower has been shown to be sufficient to cause complete LOC. Systemic blood pressure is determined by cardiac output and total peripheral vascular resistance, and a fall in either can cause syncope, but a combination of both mechanisms is often present, even if their relative contributions vary considerably [2].

A low or inadequate peripheral resistance can be due to inappropriate reflex activity, causing vasodilatation and bradycardia manifesting as vasodepressor, mixed, or cardioinhibitory reflex syncope. Other causes of a low or inadequate peripheral resistance are functional and structural impairments of the autonomic nervous system with drug-induced, primary and secondary autonomic failure. In autonomic failure, sympathetic vasomotor pathways are unable to increase total peripheral vascular resistance in response to the upright position. Gravitational stress, in combination with vasomotor failure, results in venous pooling of blood below the diaphragm, causing a decrease in venous return and consequently in cardiac output.

The causes of transient low cardiac output are 3-fold. The first is a reflex causing bradycardia, known as cardioinhibitory type of reflex syncope. The second is cardiovascular causes, due to arrhythmia and structural disease including pulmonary embolism/hypertension. The third is inadequate venous return, due to volume depletion or venous pooling.

Epidemiology and prognosis

Syncope is a common and generally benign symptom. However, the high incidence of mortality and adverse events in patients with cardiovascular causes of syncope warrants the important use of resources to try to identify high risk patients [3,4]. Moreover, the risk of trauma following the loss of postural tone, and the psychological impact of multiple recurrences increase the disabling potential of syncope at low risk from a clinical standpoint [5].

Syncope true incidence is difficult to estimate because only a small percentage of patients with syncope seek medical advice [6]. The estimated prevalence varies according to the setting and the population studied, but it is likely that up to 40% of people faint at least once in a life and prevalence is higher in females. Syncope incidence shows a bimodal distribution, with two peaks: before 20 and after 65

years old. While vasovagal syncope is the more likely cause of syncope in young people, cardiovascular diseases, orthostatic hypotension and multiple causes (reflecting a higher frailty) are more prevalent in the elderly [7–10].

Syncope is a frequent cause of emergency department (ED) admission: about 1%-3% of ED visits occur after a syncopal episode and an extremely variable percentage of patients (13%-83%) is hospitalized [11–17]. Hence the need for tools to properly assess and risk stratify patient in order to both reduce inappropriate admissions and avoid adverse events following discharge.

Risk stratification

If the cause of syncope remains unexplained after the first clinical evaluation, the physician should assess the risk of major cardiovascular event or sudden cardiac death. To help the clinician, many clinical decision tools (CDT) have been created in the last years.

The currently available CDT for risk stratification of syncope patients are the San Francisco Syncope Rule (SFSR), the OESIL (Osservatorio Epidemiologico sulla Sincope del Lazio) risk score, the ROSE (Risk Stratification of Syncope in the Emergency Department) rule, the EGSYS (Evaluation of Guidelines in Syncope Study) risk score, the Boston Syncope Criteria, and the Canadian Syncope Risk Score.[12–14,18–20]

Prognostic models

A CDT can be defined as a clinical tool that quantifies the individual contributions that various components of the history, physical examination, and basic laboratory results make toward the diagnosis, prognosis, or likely response to treatment in an individual patient. This estimate relies either on the total points scored by the patient (score) or on the binary yes/no classification (rule) based on the presence of at least one predictor. Clinical decision rules attempt to formally test, simplify, and increase the accuracy of clinicians' diagnostic and prognostic assessments and are most likely to be useful in situations where decision making is complex, the clinical stakes are high, or there are opportunities to achieve cost savings without compromising patient care. Available CDTs include guides for establishing a pretest probability, providing screening tests for common problems that frequently go undetected, estimating the risk of developing a disease or complication [21].

Developing and testing a CDT involves 3 steps: creating or deriving the rule, testing or validating the rule, and assessing the impact of the rule on clinical behavior (impact analysis). The validation process may require several studies to fully test the accuracy of the rule at different clinical sites. Table 2 presents a hierarchy that can guide clinicians in assessing the full range of evidence supporting use of a CDR in their practice.

Table 2 - Hierarchy of evidence for clinical decision tools

Level 1	Rules that can be used in a wide variety of settings with confidence that they can change clinical behavior and improve patient outcome	At least 1 prospective validation and 1 impact analysis, demonstrating change in clinician behavior with beneficial consequences
Level 2	Rules that can be used in various settings with confidence in their accuracy	Demonstrated accuracy in either 1 large prospective study including a broad spectrum of patients and clinicians or validated in several smaller settings that differed from one another
Level 3	Rules that clinicians may consider using with caution and only in patients in the study are similar to those in the clinician's clinical setting	Validated in 1 narrow prospective sample
Level 4	Rules that need further evaluation before they can be applied in the clinical setting	Derived but not validated or validated in split samples, large retrospective databases, or by statistical techniques

From: "McGinn T, et al. Users' Guides to the Medical Literature XXII: How to Use Articles About Clinical Decision Rules. JAMA 2000"

Investigators who develop a CDT begin by constructing a list of potential predictors of the outcome of interest. The list typically includes items from the history, physical examination, and basic laboratory tests. The investigators then examine a group of patients and determine if the candidate clinical predictors are present and the patient's status on the outcome of interest. Statistical analysis reveals which predictors are most powerful and which predictors can be omitted from the rule without loss of predictive power. Typically, the statistical techniques used in this process are based on logistic regression [22]. Other techniques that investigators sometimes use include discriminant analysis, which produces equations similar to regression analysis; recursive partitioning analysis, which builds a tree in which the patient populations are split into smaller and smaller categories based on risk factors; and neural networks [23–25].

Even rigorously derived CDT should not be used in clinical practice without further validation. Indeed, CDT may reflect associations between given predictors and outcomes that are due primarily to chance. If that is so, a different set of predictors will emerge in a different group of patients, even if the patients come from the same setting. Moreover, predictors may be peculiar to the population or to the clinicians using the tool, and the rule may fail in a new setting. Finally, clinicians may, because of problems in the feasibility of rule application in the clinical setting, fail to implement a rule comprehensively or accurately. The result would be that a rule succeeds in theory but fails in practice. Statistical methods can deal with the first of these problems. For instance, investigators may split their population into 2 groups and use one to develop the rule and the other to test it. Alternatively, they may use more sophisticated statistical methods built on the same logic. Conceptually, these approaches involve removing 1 patient

from the sample, generating the rule using the remainder of the patients, and testing it on the patient who was removed from the sample. This procedure, sometimes referred to as a bootstrap technique, is repeated in sequence for every patient being studied [26].

While statistical validation within the same setting or group of subjects reduces the likelihood that the rule reflects the play of chance rather than true associations, it fails to address the other 2 threats to validity. The success of the CDT may be peculiar to the populations of patients and clinicians involved in the derivation study. Even if this is not so, clinicians may have difficulties using the rule in practice, difficulties that compromise its predictive power. Validation of a CDT involves demonstrating that its repeated application as part of the process of clinical care leads to the same results. Ideally, a validation entails the investigators applying the rule prospectively in a new population with a different prevalence and spectrum of disease from that of the patients in whom the rule was derived. One key issue is to be sure that the CDT performs similarly in a variety of populations and in the hands of a variety of clinicians working in a variety of institutions. A second issue is to be sure that the CDT works well when clinicians are applying it consciously as a rule, as opposed to a purely statistical validation.

Use of a CDT involves remembering predictor variables and often entails making calculations to determine a patient's probability of having the CDT's target outcome. Pocket cards and computer algorithms can facilitate the task of using complex CDTs. Nonetheless, CDTs demand clinician time and energy, and their use is warranted only if they change physician behavior and if that behavior change results in improved patient outcomes or reduced costs while maintaining quality of care. If these conditions are not met, whatever the accuracy of a CDT, attempts to use it systematically will be a waste of time. Many reasons might explain why an accurate CDT may not produce a change in behavior or an improvement in outcomes. First, clinical judgement may be as good as, if not better than, the CDT. If this is so, CDT information will not improve practice. Second, application of the tool might not be straightforward, and clinicians may, as a result, not use it. Finally, there may be practical barriers to acting on the results of the CDT.

Ideally, an impact study would randomize patients, or larger administrative units, to the application or nonapplication of the CDT and follow up patients for all relevant outcomes (including quality of life, morbidity, and resource utilization).

Logistic regression

The logistic regression model is the most widely used statistical technique to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. The model is flexible as it can incorporate categorical and continuous predictors,

non-linear transformations, and interaction terms. Logistic regression generates the coefficients β (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest [27].

In medical and epidemiological studies, the Cox proportional hazard model is the most often used method for survival outcomes. It is the natural extension of the logistic model to the survival setting. Indeed, the Cox model is equivalent to conditional logistic regression, with conditioning at times where events occur. In the logistic model, we use an intercept in the linear predictor, while in the Cox model a baseline hazard function is used. The hazard function indicates the risk of the outcome during follow-up [28].

Recursive partitioning

Recursive partitioning or Classification And Regression Tree ("CART") are statistical methods to construct binary trees [29,30]. The method is based on statistically optimal splitting ("partitioning") of the patients into pairs of smaller subgroups. Splits are based on cut-off levels of the predictors, which produce maximum separation among two subgroups and a minimum variability with these subgroups with respect to the outcome. The predictor causing the largest separation is situated at the top of the tree, followed by the predictor causing the next largest separation, and so on. Splitting continues until the subgroups reach a minimum size or until no improvement can be obtained. Several variants of recursive partitioning algorithms are available which use different criteria to construct a tree [31,32].

An advantage of a tree is its simple presentation. Some claim that a tree represents how physicians think: starting with the most important characteristic, followed by another characteristic depending on the answer on the first, etc. Indeed, humans are remarkably quick in pattern recognition based on a few clues. A true advantage may be that interaction effects are naturally incorporated in a tree, while a standard logistic regression model usually starts with main effects. So, when multiple, high-order interactions are expected in a huge data set, and only categorical predictors are considered, a tree might be a good choice [33].

Disadvantages of trees can be noted by considering a tree as a special case of linear logistic regression. First, all continuous variables have to be categorized, which implies a loss of information. Also, the tree assumes interactions between all predictors. In regression analysis, it is common practice to include main effects of predictors when interactions are considered; this principle is not followed in tree modelling. A higher-order interaction term is included to model the effect of a predictor in a specific branch, and simply omitted from the other branches. A predictor is typically selected in one branch of the tree and not in another. This poses a clear risk of bias: predictors are selectively considered when their effects are relatively large, and not if their effects are small.

Even if we could use a stepwise selection method in a logistic model, a tree needs to be selective in the inclusion of predictors, as it might quickly run out of cases within branches. Limited power and model instability is a major problem in the development of trees.

Artificial neural networks

Artificial neural networks (ANNs) are complex non-linear models inspired by the working of biological neural networks (i.e. the central nervous system). They are used to estimate complex functions (i.e. non-linear) that require a large number of inputs [34]. ANNs are presented as systems of layers (multilayer) composed of neurons (also called perceptrons) which exchange messages between each other by synapsis (weights). The values of input variables (patient characteristics) are imported into the network via the input layer and multiplied with the weights of the connections. These multiplied values constitute the input of the next (hidden) layer, from where the process is continued to produce the output variables in the output layer. The synapses have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning to minimize the error [35]. The number of hidden layers and number of nodes are chosen by the analyst [36].

The use of ANNs has already shown promising results in emergency medicine. For example, ANNs have been developed to reduce computed tomography imaging for suspected craniocervical junction injury in major head trauma patients [37]. Artificial neural networks have also been used to predict the risk of myocardial infarction in patients with chest pain [38]. As one of the major problems in syncope risk stratification is that syncope itself can be the final common presentation of several conditions which are very heterogeneous in terms of prognosis, the absence of linearity in such a context could make the application of ANNs appealing.

Clinical decision tools in syncope

OESIL risk score

The OESIL risk score was derived and validated in an Italian prospective multicenter study to identify predictors of 12-month death [12] The derivation cohort included 270 patients. To find independent predictors of mortality, 9 variables from history, physical examination and ECG were analyzed with *t test* (continuous variables) and χ^2 (categorical variables). 8 of them were found to be significantly associated with the outcome ($p<0.10$) and further analyzed with a Cox proportional hazards regression method. Hence the authors identified 4 independent predictors (with similar Hazard Ratios): age >65 years, absence of prodromes, abnormal ECG and history of cardiovascular disease. Each predictor was assigned 1 point. Death rate increased proportionally with ascending OESIL risk score, ranging from 0% among those patients with a score of 0 to 57.1% among those with a score of 4. The Area Under the Receiver Operating Characteristics (ROC) curve (AUC) in the derivation cohort was 0.897.

Although the authors used a robust methodology to derive the rule, the 1-year follow-up limits its practical utility in the ED. Indeed, the emergency physician is interested in identifying, and possibly avoiding, short-term adverse events. Also, how the 9 predictors were chosen by the authors is unclear.

EGSYS score

The EGSYS score was derived and validated in an Italian prospective multicenter study to identify predictors of 24 months cardiogenic syncope [18]. The derivation cohort included 260 patients, 44 of which had a diagnosis of cardiogenic syncope at follow-up. To find the independent predictors of cardiogenic syncope, 52 predictors were analyzed with a χ^2 test. 14 of them were found to be significantly associated with the outcome ($p<0.10$) and further analyzed with logistic regression. Each of the 6 independent predictors identified was assigned a score between -1 and 4 based on the multivariate logistic regression coefficient. A score ≥3 had a sensitivity, specificity and AUC of 95%, 61% and 0.904, respectively.

EGSYS score has never been externally validated, therefore its use cannot be recommended in clinical practice. Moreover, the subjectivity of the outcome considered and its distance from the index event make the score less relevant to the emergency physician.

San Francisco Syncope Rule

The SFSR was derived in a prospective monocenter study in the United States [14]. The aim of the study was to find predictors of 7-day adverse events and death. The derivation cohort included 684 patients, 79 of whom had serious outcomes. To find independent predictors of death or adverse events, 50 candidate predictors were analyzed with the Mann-Whitney (continuous variables) and χ^2 (categorical variables) test. The 26 predictors that were found to be significantly related the outcome ($p<0.10$) and with a Cohen's kappa coefficient >0.5 (for subjectively assessed variables) were analyzed with recursive partitioning. The aim was to develop a model that would maximize the prediction of serious outcomes. 5 independent predictors were selected: abnormal ECG, a complaint of shortness of breath, hematocrit less than 30%, systolic blood pressure less than 90 mm Hg, and a history of congestive heart failure. The rule is considered as positive in the presence of at least one predictor. Sensitivity and specificity were 96% and 62%, respectively in the derivation cohort.

SFSR is the most externally validated among syncope CDTs. Validation studies did not confirm the promising data of the rule derivation and this might be due to the use of recursive partitioning that could have led to overfitting and poor external validity. Moreover, after the publication of the study, the definition of “abnormal ECG” was much discussed. The authors stated that they considered all the ECGs performed in the ED, including telemetry. However, the detection of an arrhythmia during ECG monitoring should be considered a diagnosis, rather than an abnormal ECG.

ROSE rule

The ROSE rule was derived and validated in a prospective monocenter UK study [13]. The aim was to find predictors of 30-day death and adverse events. The derivation cohort included 550 patients, 40 of which had serious outcomes. To find the independent predictors of death and adverse events variables from history, physical examination and ECG were analyzed with *t test* (continuous variables) and χ^2 (categorical variables). Those found to be significantly correlated with the outcome ($p<0.10$) were analyzed with logistic regression. 9 independent predictors were therefore identified and subsequently analyzed with recursive partitioning, thus leaving 7 predictors for inclusion in the rule. The ROSE rule was considered as positive in the presence of at least 1 predictor. Sensitivity and specificity in the derivation cohort were 93% and 74%, respectively.

ROSE was the first syncope CDT considering a biomarker (the Brain Natriuretic Peptide) among the predictors. However, we must acknowledge some weaknesses: 1) the rule has never externally validated; 2) the sample size was calculated based on the planned multivariate regression analysis. However, the authors decided to further reduce the number of predictors with recursive partitioning. As this had not been planned in advance, the methodological validity of the study might be compromised.

Boston syncope criteria

The authors of this score did not derive it based on prospectively collected data, but they considered as predictors, variables identified by guidelines or experts in the field [19]. The Boston syncope criteria have indeed scarce clinical utility, as it is based on a high number of variables. Also, some of them (i.e. brain ischemia) are considered both risk factors and outcome measures.

Canadian Syncope Risk Score

The Canadian Syncope Risk Score was derived and internally validated in a prospective multicenter Canadian study [20]. The aim was to find predictors of 30-day adverse events and death. Of the 4030 patients enrolled, 254 had serious outcomes. To find independent predictors of death or adverse events 43 variables from history, physical examination, ECG, vitals, and blood tests were assessed. Those with less than 5 associate events, a *variance inflation factor* for multicollinearity >5, more than 25% of missing values, and a low interrater agreement ($\kappa <0.4$) were excluded. Then a univariate analysis with χ^2 of Fisher exact test (categorical variables) and *t test* (continuous variables) was performed. Continuous predictors were dichotomized using a combination of clinical rationale and analysis of receiver operating characteristic curves, which identified the optimal cut-off point based on measures of sensitivity, specificity and the Youden Index. After categorization, the 23 predictors selected by bivariable analysis were included in a multivariable logistic regression model, which was reduced by stepwise backward elimination with a 5% significance level to stay in the model. The score was created multiplying the regression coefficients by the

shrinkage factor. The score ranges from −3 to 11, with a shrinkage-adjusted expected risk ranging from 0.4% to 83.6%, respectively. Internal validation was carried out with the use of 500 bootstrap samples. The sensitivity and specificity were 99.2% and 25.4%, respectively with a cut-off ≥-2, and 97.7% and 45.1%, respectively with a cut-off ≥-1. The AUC was 0.88.

The Canadian Syncope Risk Score is the CDT on syncope that was derived in the largest available dataset, therefore the high number of serious outcomes should give stability to the multivariate model. Moreover, the choice not to establish an arbitrary cut-off, but rather to provide the probability of adverse events based on the score, could increase its clinical utility. However, we must acknowledge some limitations: 1) as no external validation exists, it should not be used in clinical practice; 2) data on one of the predictors (troponin) is missing in about a half of the enrolled patients and the authors arbitrarily supposed that a missing value implied a normal value; 3) although the observation of serious outcomes in the ED was one of the exclusion criteria, two of the predictors included in the model were an ED diagnosis of vasovagal or cardiogenic syncope.

Limits of the currently available CDTs

Although the derivation studies showed promising results for a possible use of the above CDT in patients presenting to the ED for syncope, some limitations have to be acknowledged. Only two scores (SFSR and OESIL) have been externally validated, and no CDT should be used in clinical practice without an external validation [21]. External validation studies and systematic reviews have failed to confirm the data of derivation studies [16,39]. Moreover, an individual-patient data meta-analysis showed that the sensitivity, specificity or prognostic yield of the currently available CDTs is not better than clinical judgment in predicting short-term serious outcome after syncope [40]. Therefore, the use of CDT is no more recommended in the assessment of ED syncope patients [41].

AIMS

Aim of the present work was to compare multivariate logistic regression and ANN to analyze the possible similarities, differences, strengths and weaknesses of the two methods in predicting serious adverse events in patients presenting with syncope to the ED. In case none of such methods succeeded in an accurate prediction, we aimed at finding possible alternative methods to risk stratify syncope patients.

MATERIALS AND METHODS

Data collection

Individual patients' dataset

To retrospectively derive and compare a multivariate logistic regression model and a ANN, we built an individual-patient dataset of patients prospectively included in studies enrolling syncope patients in the ED and for whom a 7-10 days follow-up was available.

Therefore, we performed a systematic literature search in Embase and PubMed looking for the words "syncope", "emergency service" and "clinical prediction guides" and their synonyms. We imposed no language restriction. We then wrote the corresponding author of each study to ask for individual patients' data.

Prospective dataset

Patients for the prospective validation of the above tools were enrolled in a prospective multicenter study that was held in 6 hospitals in northern Italy: Niguarda, Policlinico and Sacco Hospitals in Milano, Humanitas Research Hospital in Rozzano, Alessandria Hospital and Santa Croce Hospital in Moncalieri.

Inclusion criteria

We enrolled adult (i.e. >18 years) patients that were admitted after syncope in the ED of one of the participating centers between September 2015 and February 2017.

Exclusion criteria

- LOC following head trauma;
- Non-spontaneous consciousness recovery;
- Episodes of ground fall, dizziness or lightheadedness without LOC;
- LOC associated with alcohol or drugs abuse;
- Pregnancy or breastfeeding;
- Inability to provide informed consent to the study or to complete follow-up;
- Syncope as underlying symptom of acute conditions that were diagnosed in the ED or requiring therapeutic intervention irrespective of syncope (acute myocardial infarction, pulmonary embolism, aortic dissection, cerebral hemorrhage, arrhythmias diagnosed before ECG monitoring in the ED);
- Non-syncopal LOC (i.e. history of epilepsy);
- Poor prognosis in the next 30 days.

Data collection

Personal data, past medical history and features of the syncopal episode were collected for each patient according to a previous consensus [42]. The ED physicians managed the patients irrespective of the study inclusion and were asked about the perceived risk of adverse events (based on clinical judgement). In case of admission, a copy of the hospital discharge letter was retrieved. A 7 and 30-day telephone follow-up was performed to assess the occurrence of any adverse events.

The study was approved by the Ethical Committee on Human Research of the Coordinating Centre (Ospedale "L.Sacco"), and participants provided written consent. Oral consent was obtained in patients discharged from ED that were interviewed by phone.

Definitions

Syncope was defined as a transient loss of consciousness with loss of postural tone, likely due to transient global cerebral hypoperfusion and characterized by rapid onset, short duration, and spontaneous complete recovery [1].

ECG monitoring was considered positive in the presence of any of the following [41]:

- Sinus arrest with cardiac pause >3 seconds;
- Sustained or non-sustained ventricular tachycardia, whether symptomatic or asymptomatic;
- High grade atrioventricular (AV) block (second-degree type 2 or third-degree AV block);
- Bradycardia <30 beats per minute (bpm) whether symptomatic or asymptomatic;
- Bradycardia <50 bpm in a symptomatic patient;
- Tachycardia >120 bpm in a symptomatic patient.

We defined as abnormal ECG the presence of any of the following abnormalities at ED presentation [41,42]:

- Non-sinus rhythm (including paced rhythm);
- Sinus bradycardia ≤40 bpm;
- Left bundle branch block;
- Delta waives;
- Prolonged QRS (>120 milliseconds);
- Prolonged QTc (>450 milliseconds);
- Brugada pattern;
- Q/ST/T changes consistent with acute or chronic ischemia.

We considered as adverse events any of the following [42]:

- Cardiac death and syncope-related death;

- Ventricular fibrillation;
- Sustained ventricular tachycardia and symptomatic non-sustained ventricular tachycardia;
- Paroxysmal or new onset atrial fibrillation;
- Sinus arrest with cardiac pause >3 seconds;
- Sick sinus syndrome with alternating bradycardia and tachycardia;
- Second-degree type 2 or third-degree AV block;
- Permanent pacemaker (PM) or implantable cardioverter defibrillator (ICD) with cardiac pauses;
- Aortic stenosis with valve area ≤1 cm2;
- Hypertrophic cardiomyopathy with outflow tract obstruction;
- Left atrial myxoma or thrombus with outflow tract obstruction;
- Myocardial infarction;
- Pulmonary embolism;
- Aortic dissection;
- Occult hemorrhage or anemia requiring transfusion;
- Syncope or fall resulting in major traumatic injury (trauma that requires admission or procedural⁄surgical intervention);
- PM or ICD implantation;
- Cardiopulmonary resuscitation;
- Syncope recurrence and syncope recurrence with hospital admission;
- Cerebrovascular events.

Data analysis

Descriptive statistics for continuous and categorical variables were used to summarize the baseline characteristics of patients enrolled.

Logistic regression analysis

Potential predictors of short-term severe outcomes were first individually evaluated and then analyzed by multivariate logistic regression analysis with and without interaction variables with a stepwise selection strategy. In case of one predictor was missing in one patient, it was considered as absent.

Artificial neural networks

Prior to ANN development, we set optimal desired values of sensitivity and specificity for ANN training. To improve the robustness of results, a bootstrap resampling procedure for 100 iterations was performed on the full data set. Each random resampling was divided into training and validation sets. In the first phase, elements from the training samples (both clinical and demographic characteristics, and outcome information) were introduced into the model using the back-propagation supervised training algorithm. The

ANN calculated weight matrices until a predefined error threshold was reached and the best weight matrix (in terms of sensitivity and specificity) was identified. Once the weights and ANN framework were defined (i.e. type of activation functions between layers; number of hidden layers and their nodes; other technical parameters, such as the search function for the optimal gradient), these parameters were applied to the validation samples, which were introduced into the ANN without any information on the outcomes. The ANN applied the framework to the new data in order to test the ability of the model to provide the correct classification (in our case, to correctly identify patients with a high risk of a serious outcome). We used a conventional approach which used 4/5 of the population for the training set and 1/5 for the validation set. In the derivation setting, both the sensitivity and specificity thresholds were set to 1.

The overall diagnostic performance of both multivariate logistic regression and ANN was assessed with ROC curves and their AUC. To test the predictive accuracy of the models if used to either discharge or admit patients, we pre-defined three cut-offs: 2% (probability of adverse events below which a patient should be discharged), 30% and 50% (probabilities of adverse events above which a patient should be considered for admission). For each cut-off we calculated sensitivity, specificity, positive and negative predictive values.

Analyses were performed using the SAS (release 9.4) and MATLAB (release R2012b, 64bit) statistical software.

RESULTS

Retrospective analysis

Individual patient data dataset

We selected 13 studies (published by 11 authors) meeting our inclusion criteria. Six authors agreed to provide data, for a total of 3681 patients [13,14,18,43–46]. Here the details of the included studies:

- Grossman 2007: 293 patients, prospective validation of the Boston Syncope criteria [43];
- Del Rosso 2008/Ungar 2010: 465 patients, EGSYS derivation and internal validation/prognosis of the patients enrolled in the EGSYS study [18,44];
- Reed 2010: 1067 patients, derivation and validation of the ROSE rule [13];
- Quinn 2004: 684 patients, derivation of the SFSR [14];
- Costantino 2008: 695 patients, STePS study to analyze the short and long-term prognosis of syncope [45];
- Sun 2007: 477 patients, external validation of the SFSR [46].

Each database was assessed to make data homogeneous. Abnormal ECG and short-term (i.e. 7-10 days) adverse events were codified according to the above criteria [42]. Outcomes happening in the ED were also included.

The common dataset included all the following variables:

- Age;
- Sex;
- Peripheral oxygen saturation;
- Hematocrit;
- Syncope during exertion, in supine or sitting position;
- Syncope associated with chest discomfort;
- Syncope associated with shortness of breath;
- Syncope associated with palpitations;
- Syncope associated with trauma;
- Syncope without any prodromes;
- History of congestive heart failure;
- History of cardiovascular disease;
- History of cerebrovascular disease;
- History of arterial hypertension
- Systolic blood pressure below 90 mmHg;

- Abnormal ECG;
- Admission to hospital vs ED discharge;
- Death and adverse events in the ED, at 10 and 30 days.

We selected as candidate predictors the variables that were present in at least 5 of the 6 datasets. We identified 10 common variables: age, sex, syncope during exertion, syncope associated with trauma, abnormal ECG, history of cardiovascular disease, history of cerebrovascular disease, history of previous syncope, absence of prodromes, and history of arterial hypertension.

Due to the lack of data on some of the predictors, one dataset was subsequently excluded [43], leaving 3388 patients for the analysis. Table 3 shows the main characteristics of the included patients.

Table 3 - Characteristics of the included patients

Variables	EGSYS	SFSR	STePS	ROSE	Sun 2007	Total
Total number of patients	465	684	695	1067	477	3388
Age, median (IQR)	70 (45-81)	70 (42-81)	64 (41-78)	69 (48-81)	58 (35-79)	67 (43-80)
N of admitted patients (%)	178 (38)	364 (53)	265(38)	538 (50)	286 (60)	1631 (48)
N of men (%)	253 (54)	281 (41)	306 (44)	480 (45)	210 (44)	1530 (45)
N of patients with history of syncope (%)	195 (42)	124 (18)	389 (56)	176 (16)	160/457	1044/2931 (36)
N of patients affected by hypertension (%)	184 (40)	204 (30)	276 (40)	409 (38)	185/470 (37)	1258/2918 (43)
N of patients without prodromes (%)	122 (26)	260 (38)	195 (28)	410 (38)	141 (30)	1128 (33)
N of patients with syncope during exertion (%)	15 (3)	53 (8)	15 (2)	61 (6)	74/466 (16)	218/2922 (7)
N of patients with trauma following syncope (%)	133 (29)	45 (7)	162 (23)	316 (30)	n.a.	656/2911 (23)
N of patients with abnormal ECG (%)	178 (38)	222 (32)	202 (29)	665 (62)	170 (36)	1437 (42)
N of patients with a history of cardiovascular disease (%)	153 (33)	139 (20)	178 (26)	284 (27)	150 (31)	904 (27)
N of patients with a history of cerebrovascular disease (%)	166 (36)	115(17)	227(33)	n.a.	169 (35)	677/2321 (29)

N of patients with serious outcomes at 10 days (%)*	93 (20)	81 (12)	44 (6)	49 (5)	62 (13)	329 (10)
N of deaths	6	6	7	6	1	26 (1)
N of arrhythmias	31	30		20	32	
N of cardiopulmonary resuscitations			5	2		
N of myocardial infarctions	6	33			1	
N of structural cardiopulmonary diseases	9	10		14	6	
N of PM insertions or malfunctions	43		25	11	2	
N of ICD insertions or malfunctions	5		2			
N of haemorrhages		24		7	8	

*Table legend: IQR: interquartile range; ECG: electrocardiogram; PM: pacemaker; ICD: Implantable Cardioverter Defibrillator; n.a.: not available. *Some patients had more than one outcome.*

Logistic regression

At univariate analysis, the risk factors significantly associated with severe short-term outcomes were: age, male gender, syncope during exertion, abnormal ECG, history of cardiovascular disease, history of cerebrovascular disease, absence of prodromes, and history of arterial hypertension (Table 4).

Table 4 - Risk factors for severe short-term outcomes within 10 days (univariate analysis)

	Severe Outcomes		
	Yes (%) (n=329)	No (%) (n=3059)	p-value*
Male gender, n (%)	196 (60)	1334 (44)	<0.0001
Age, n (%)			<0.0001
< 45 years	24 (7)	869 (28)	
≥ 45 and < 65 years	56 (17)	658 (22)	
≥ 65 years	249 (76)	1532 (50)	
Syncope during exertion, n (%)	31 (9)	187 (6)	0.0211
Trauma following syncope, n (%)	64 (19)	592 (19)	0.9651
Abnormal ECG, n (%)	229 (70)	1208 (39)	<0.0001
Medical history, n (%)			
Cardiovascular disease	161 (49)	743 (24)	<0.0001
Cerebrovascular disease	132 (40)	545 (18)	<0.0001
Arterial hypertension	154 (47)	1104 (36)	0.0001
Previous syncope	109 (33)	964 (31)	0.5491
Absence of prodromes, n (%)	126 (38)	1002 (33)	0.0430

**Chi-square test; ECG: electrocardiogram*

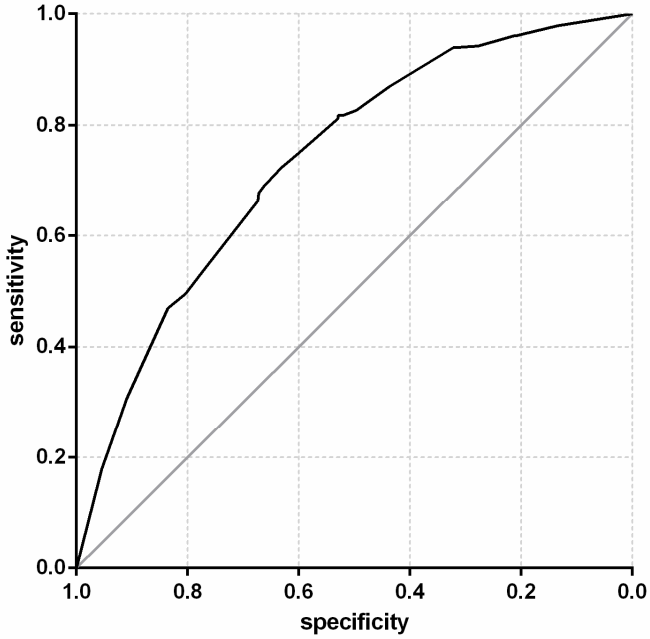
At multivariate analysis, male gender (adjusted odds ratio [OR] 1.6; 95% confidence interval [CI] 1.3 to 2.0), age between 45 and 65 years (OR 2.3, 95% CI 1.4 to 3.8), age over 65 years (OR 3.5, 95% CI 2.3 to 5.5), an abnormal ECG (OR 2.6, 95% CI 2.0 to 3.3), and a past medical history of cerebrovascular disease (OR 1.9, 95% CI 1.5 to 2.5) were independent risk factors for the development of severe adverse outcomes in the short term (Table 5). The area under the Receiver Operating Characteristics (ROC) curve (AUC) for the multivariate model was 0.736 (Figure 2).

Table 5 - Risk factors for severe short-term outcomes within 10 days at logistic multivariate regression (stepwise selection)

	Adjusted Odds Ratio	95% Confidence Interval	p-value*
Male gender	1.6	1.3 – 2.0	0.0001
Age			<0.0001
< 45 years	1.0		
≥ 45 and < 65 years	2.3	1.4 – 3.8	
≥ 65 years	3.5	2.3 – 5.5	
Abnormal ECG	2.6	2.0 – 3.3	<0.0001
Medical history of cerebrovascular disease	1.9	1.5 – 2.5	<0.0001

**Chi-square test; ECG: electrocardiogram*

Figure 2 - ROC curve for the multivariate model



The multivariate model with interaction variables showed a similar predictive accuracy, with an AUC of 0.745. The risk factors for severe short-term outcomes within 10 days at logistic multivariate regression with interaction variables are reported in Table 6.

Table 6 - Risk factors for severe short-term outcomes within 10 days at logistic multivariate regression with interaction variables (stepwise selection)

	p-value*
Age class	<0.0001
Sex	0.0002
Abnormal ECG	<0.0001
Medical history of cerebrovascular disease	<0.0001
Age class x medical history of cerebrovascular disease	0.0004
Abnormal ECG x medical history of cerebrovascular disease	0.0373

**Chi-square test; ECG: electrocardiogram*

Artificial Neural Networks

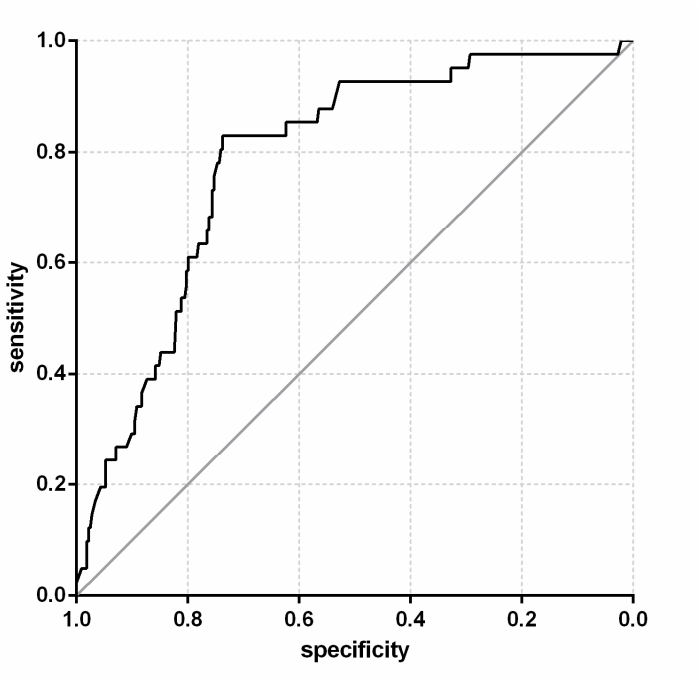
For the purpose of the derivation of ANN, two datasets were excluded from the analysis because of the absence of at least one of the above predictor variables [13,46]. The remaining three datasets had 19 patients with missing data, therefore the final analytic cohort included 1825 patients [14,18,45]. 4/5 of the population (1460 patients) were used for the ANN training, and 1/5 (365 patients) for the internal validation of the ANN. At the optimal cut-off of 0.064 identified during the training phase, the sensitivity, specificity and AUC of ANN in identifying short-term adverse events were 93%, 67% and 0.786, respectively (Table 7 and Figure 3) [47].

Table 7 - Predictive accuracy of the ANN

Validation	Patients (n)	Events (n)	FP (n)	FN (n)	Sensitivity	Specificity	AUC
1/5 of the population	365	41	100	7	0.83	0.69	0.786

Table legend: FP: false positives; FN: false negatives; AUC: Area Under the Curve

Figure 3 - ROC curve for the ANN (1/5 of the population)



Prospective validation

Prospective database

From September 2015 to February 2017 we enrolled 354 patients who presented to the ED for syncope. 179 (51%) were males and 175 (49%) females. Median age was 72 years (Interquartile range, IQR 52 to 81 years). 11 patients were enrolled at Alessandria hospital, 81 at Humanitas Research Hospital (Rozzano), 15 at Moncalieri Hospital, 62 at Niguarda Ca' Granda Hospital (Milano), 119 at Ospedale Maggiore Policlinico (Milano), and 66 at Sacco Hospital (Milano). The patients judged at low, intermediate and high risk by the ED physician were 98, 183 and 73, respectively. 107 patients were admitted, 246 were discharged and 1 patient was discharged against medical advice. As the ED physician disposition would have been the patient's admission, we considered this patient as if he/she was admitted to hospital. Therefore, the admission rate was 30.5%

The main characteristics of the population are reported in Table 8. ECG was not performed in 8 patients.

Table 8 - Characteristics of the enrolled patients

	n (%)
Patients enrolled	**354**
Patients' characteristics	
Male sex	179 (51)
Median age (IQR)	72 (52-81)
Median systolic blood pressure, mmHg (IQR)	130 (115-150)
Median heart rate, bpm (IQR)	76 (66-86)
Characteristics of the syncopal episode	
Syncope during working activity	20/350 (6)
Syncope during exertion	5 (1)
Syncope during driving	2 (1)
In supine position	8 (2)
In sitting position	103 (29)
In orthostatic position	211 (60)
While standing from the sitting position	35 (10)
Postprandial syncope	40 (11)
Trauma following syncope	132 (37)
Syncope without prodromes	110 (31)
Syncope associated with:	
Chest pain	22 (6)
Shortness of breath	19 (5)
Palpitations	16 (5)
Lightheadedness	84 (24)
Nausea/vomiting	70 (20)
Sensation of warmth	31 (9)
Sweating	70 (20)
Blurred vision	68 (19)
Triggered by pain/stressors	24 (7)
Triggered by cough/micturition/defecation	24 (7)
Hematocrit <30%	12/350 (3)
Hemoglobin <9 g/dl	10/351 (3)
Systolic arterial blood pressure <90 mmHg	11/353 (3)
Past medical history	
Syncope in the previous year	94 (27)
Congestive heart failure	9 (3)

Ischemic cardiomiopathy	53 (15)
Structural heart disease	24 (7)
Aortic stenosis	7 (2)
Left ventricular outflow obstruction	1 (0)
Left ventricular hypertrophy	6 (2)
Left ventricular ejection fraction <40%	9 (3)
Pulmonary hypertension	11 (3)
Valvular heart disease	10 (3)
Arrhythmias	37 (10)
Previous PM implant	12 (3)
Previous ICD implant	2 (1)
Sick sinus syndrome	1 (0)
Mobitz 2 second- or third-degree AV block	3 (1)
Arterial hypertension	189 (53)
Stroke/TIA	28 (8)
Neoplasm	45 (13)
Cronic kidney disease (serum creatinine ≥2 mg/dl)	14 (4)
COPD	20 (6)
ECG findings (346 patients with ECG results available)	
Bradycardia <50 bpm	14 (4)
First-degree AV block	35 (10)
Mobitz 1 second-degree AV block	3 (1)
Right bundle branch block	37 (11)
Left bundle branch block	11 (3)
Left anterior fascicular block	26 (8)
Previous myocardial infarction	23 (7)
Left ventricular hypertrophy	5 (1)
Ventricular ectopic beats	13 (4)
Supraventricular ectopic beats	14 (4)
Atrial fibrillation	24 (7)
Sinus bradycardia <60 bpm	40 (12)
Sinus tachycardia >100 bpm	23 (7)
Prolonged QT interval	7 (2)

Table legend: n: number; IQR: interquartile range; PM: pacemaker; ICD: Implantable Cardioverter Defibrillator; AV: atrioventricular; COPD: chronic obstructive pulmonary disease; TIA: transient ischemic attack; ECG: electrocardiogram.

After 7 days of follow-up 48 patients (13.6%) had serious outcomes, 2 of whom died (Table 9).

Table 9 – Serious outcomes at the 7-day follow-up

7-day serious outcomes	**n**
Total patients with events	48 (14%)
Death	2
Ventricular fibrillation	1
Sustained ventricular tachycardia	1
Non-sustained ventricular tachycardia	2
Atrial fibrillation	12
Cardiac pause	7
Sick sinus syndrome	3
Mobitz 2 second-degree AV block	1
Third-degree AV block	4
PM disfunction with pause	2
Acute myocardial infarction	2
Pulmonary embolism	1
Acute hemorrhage	4
Syncope with trauma	1
Syncope with hospitalization	2
PM or ICD implant	20
Cerebrovascular events	1

Table legend: n: number; PM: pacemaker; ICD: Implantable Cardioverter Defibrillator; AV: atrioventricular.

The prevalence of predictors in patients with and without 7-day serious outcomes is reported in Table 10.

Table 10 - Prevalence of predictors in patients with and without 7-day adverse events

	Serious outcomes	
	Yes (%) (n=48)	No (%) (n=306)
Male gender, n (%)	25 (52)	154 (50)
Age, n (%)		
< 45 years	2 (4)	59 (19)
≥ 45 and < 65 years	6 (12)	71 (23)
≥ 65 years	40 (83)	176 (58)
Syncope during exertion, n (%)	1 (2)	4 (1)
Trauma following syncope, n (%)	21 (44)	111 (36)
Abnormal ECG, n (%)*	37 (77)	120 (39)
Medical history, n (%)		

Cardiovascular disease	20 (42)	88 (29)
Cerebrovascular disease	5 (10)	23 (8)
Arterial hypertension	30 (62)	159 (52)
Previous syncope	27 (56)	233 (76)
Absence of prodromes, n (%)	21 (44)	89 (29)

*Table legend: *ECG was not performed in 8 patients, 1 with and 7 without outcomes; ECG: electrocardiogram; n: number.*

Logistic regression

The AUC of the multivariate regression model in the prospective cohort was 0.726 (Figure 4). The predictive accuracy of the model according to the three pre-defined cut-offs is reported in Table 11.

Figure 4 - ROC curve for the validation of the multivariate model

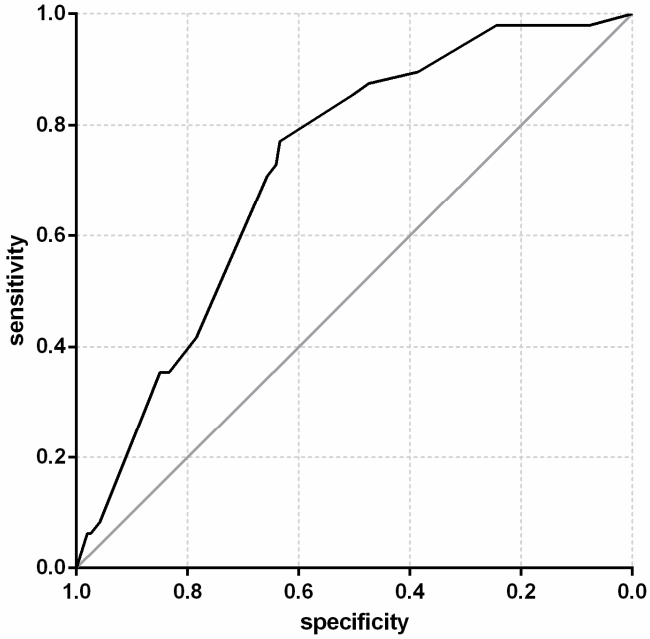


Table 11 - Predictive accuracy of the multivariate regression model according to different cut-offs

Cut-off	FP (n)	FN (n)	Sensitivity	Specificity	PPV	NPV	Patients with a decision§ (n)
2%	282	1	0.98	0.78	0.14	0.96	25 (7%)
30%	6	45	0.63	0.98	0.33	0.87	9 (3%)
50%*	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	0

*Table legend: n: number; FP: false positives; FN: false negatives; PPV: positive predictive value; NPV: negative predictive value; n.e.: not estimable; *no patient had an estimated probability above 50%; §: proportion of patients below the 2% or above the 30% and 50% thresholds for discharge and admission on the total number of patients (i.e. 354).*

Artificial Neural Networks

The AUC of the ANN in the prospective cohort was 0.694 (Figure 5). The predictive accuracy of the model according to the three pre-defined cut-offs is reported in Table 12.

Figure 5 - ROC curve for the validation of ANN

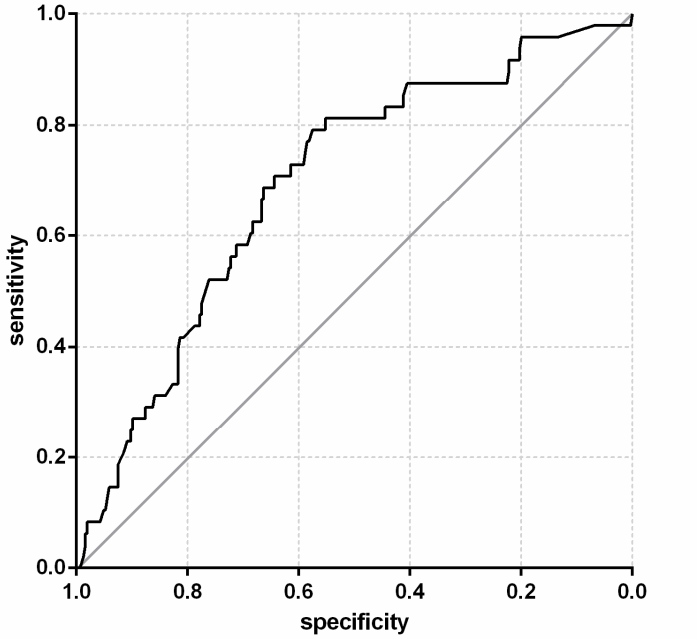


Table 12 - Predictive accuracy of ANNs according to different cut-offs

Cut-off	FP (n)	FN (n)	Sensitivity	Specificity	PPV	NPV	Patients with a decision[§] (n)
2%	297	1	0.98	0.29	0.14	0.90	10 (3%)
30%	49	33	0.31	0.84	0.23	0.89	64 (18%)
50%	18	41	0.15	0.94	0.28	0.88	25 (7%)

Table legend: n: number; FP: false positives; FN: false negatives; PPV: positive predictive value; NPV: negative predictive value; §: proportion of patients below the 2% or above the 30% and 50% thresholds for discharge and admission on the total number of patients (i.e. 354).

Possible future perspectives: attribute matching

Since the poor predictive accuracy of the analyzed models, we tried to identify alternative methods to provide an estimate of the risk of adverse events.

We hypothesized that accurate pretest probability assessments can be obtained by matching an individual patient to a group of previously studied patients who shared the same clinical characteristic, and determining the percentage of these previously studied patients who had the outcome of interest. This hypothesis proposes a method of inference that differs substantially from both the logistic regression and ANNs. Instead of treating each clinical characteristic as an independent value and adding up the

coefficients, this method proposes a system that forces the probability to be computed from a dependent set of clinical characteristics. In other words, all chosen characteristics of a patient of interest (or attributes) must be matched to an identical profile of attributes shared by a group of patients with known outcomes contained in a large derivation database. This approach has already been described to assess the pre-test probability of acute coronary syndrome and pulmonary embolism in patients with chest pain [48–50]. This process results in a denominator of all matched patients and numerator of patients with serious outcomes, and the quotient reveals the pretest probability that can be expressed as a percentage value.

We used as derivation database the retrospective cohort or 3388 patients who were evaluated for syncope and for whom outcomes were known. And we considered as predictors the same 10 variables that were assessed in the derivation of the logistic and ANN models.

The attribute matching method used 515 of the 1536 possible unique pre-test probability estimates. No patient in the derivation database matched the remaining 1021 combinations of predictors. The median match size (or denominator, i.e. the number of patients with the same combination of predictors) used to compute the pretest probability from the attribute matching method was 2 (IQR 1 to 7, range 1 to 128). Only 9 of the 515 (2%) pre-test probability estimates had a match size ≥50 patients, 19 (4%) had a match size ≥30 patients, and most (420, 82%) had a match size <10 patients. The top 10 most frequent matching profiles are shown in Table 13.

Table 13 - Top 10 most frequent matching profiles

Patient count (n)*	Serious outcomes (n)	Estimated probability	Sex	Age (years)	Syncope during exertion	Trauma	Absence of prodromes	No previous syncope	Hystory of cerebrovascular disease	Hystory of cardiovascular disease	Arterial hypertension	Abnormal ECG
138	3	2.2%	F	<45	no	no	no	yes	no	no	no	no
112	0	0%	F	<45	no	no	no	no	no	no	no	no
83	2	2.4%	M	<45	no	no	no	yes	no	no	no	no
64	1	1.6%	F	<45	no	no	yes	yes	no	no	no	no
63	2	3.2%	F	≥65	no	no	no	yes	no	no	no	no
59	1	1.7%	F	45-64	no	no	no	yes	no	no	no	no
55	7	12.8%	F	≥65	no	no	no	yes	no	no	yes	yes
50	3	6.0%	F	≥65	no	no	no	yes	no	no	yes	no
50	3	6.0%	M	45-64	no	no	no	yes	no	no	no	no
46	9	19.6%	M	≥65	no	no	no	yes	no	no	no	yes

*Table legend: n: number; ECG: electrocardiogram; *number of patients in the derivation dataset matched to the profile.*

Comparison with the previously derived methods

We compared attribute matching to the previously derived prediction models. First, we calculated the Intraclass Correlation Coefficient (ICC) and its 95% confident intervals based on a mixed-effects model on the 1825 patients included in all the ANN, logistic regression and attribute matching derivation datasets [51]. The results are reported in Table 14. Figure 6, Figure 7 and Figure 8 represent the scatter plots to visually assess the correlation between the probability of adverse events predicted by logistic regression, ANN, and attribute matching.

Table 14 – ICC and their 95% Confidence Intervals between attribute matching, logistic regression and ANN

Comparison	Intraclass Correlation	95% CI	
		Lower bound	Upper bound
logistic regression and ANN	0.758	0.738	0.777
logistic regression and attribute matching	0.457	0.420	0.493
ANN and attribute matching	0.520	0.486	0.553

Table legend: CI: Confidence Interval; ANN: Artificial Neural Network.

Figure 6 – Scatter plot of the probability of adverse events predicted by logistic regression and ANN

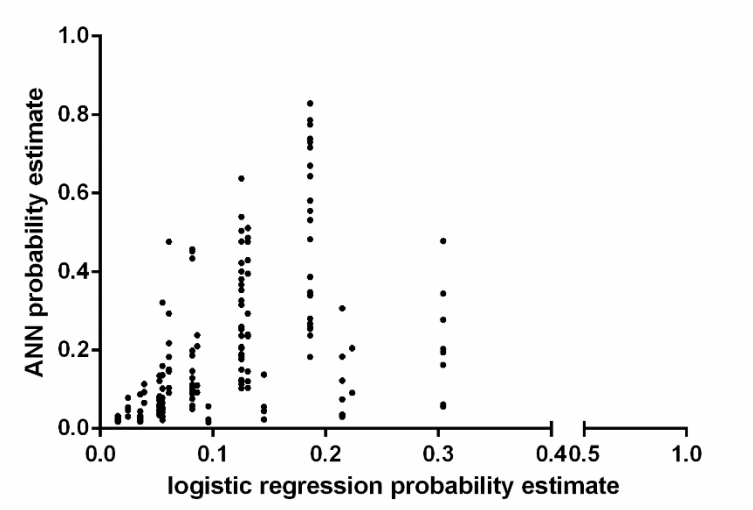


Figure 7 - Scatter plot of the probability of adverse events predicted by logistic regression and attribute matching

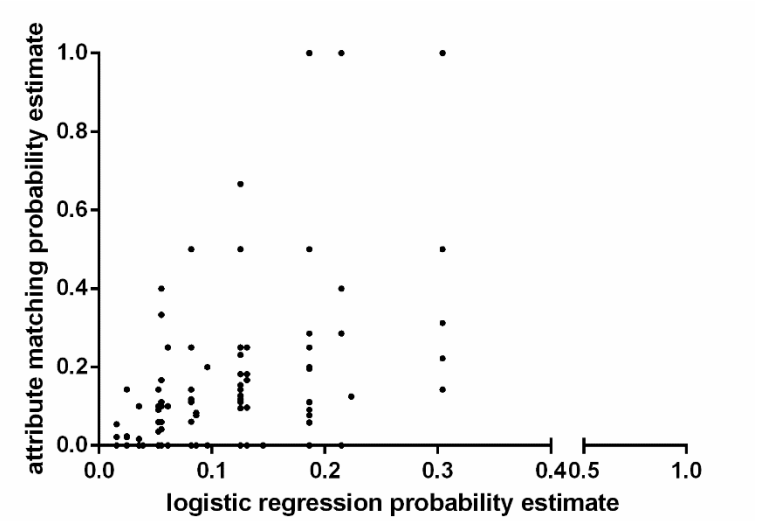
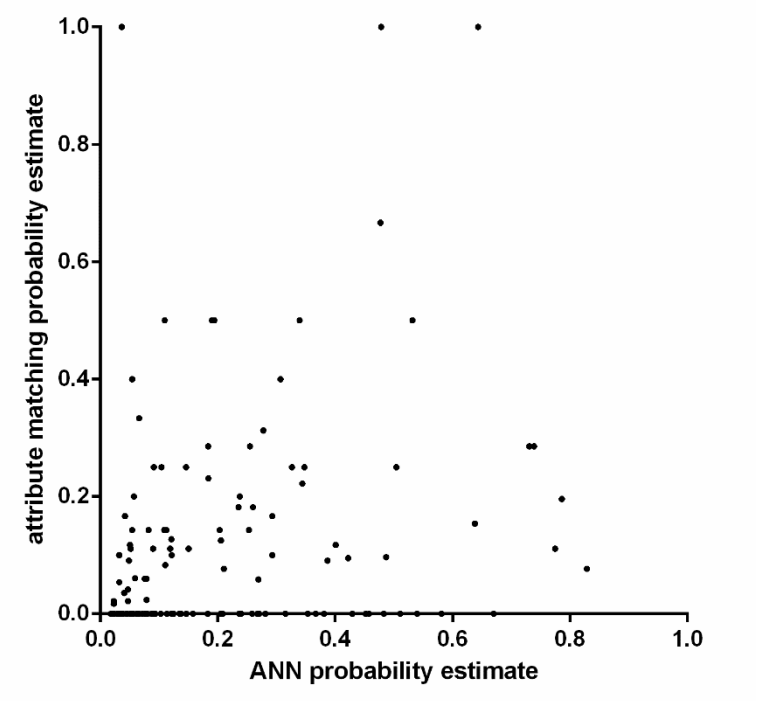


Figure 8 - Scatter plot of the probability of adverse events predicted by ANN and attribute matching



Then, we performed the same analysis including only the 630 patients with a match size (denominator) of at least 30 patients. The results are reported in

Table 15. Figure 9 represents the scatter plots to visually assess the correlation between the probability of adverse events predicted by logistic regression, ANN, and attribute matching.

Table 15 - ICC and their 95% Confidence Intervals between attribute matching, logistic regression and ANN for the patients with a match size of at least 30 patients

Comparison	Intraclass Correlation	95% CI	
		Lower bound	Upper bound
logistic regression and ANN	0.873	0.853	0.891
logistic regression and attribute matching	0.962	0.958	0.965
ANN and attribute matching	0.917	0.910	0.924

Table legend: CI: Confidence Interval; ANN: Artificial Neural Network.

Figure 9 - Scatter plot of the probability of adverse events predicted by logistic regression and ANN for the patients with a match size of at least 30 patients

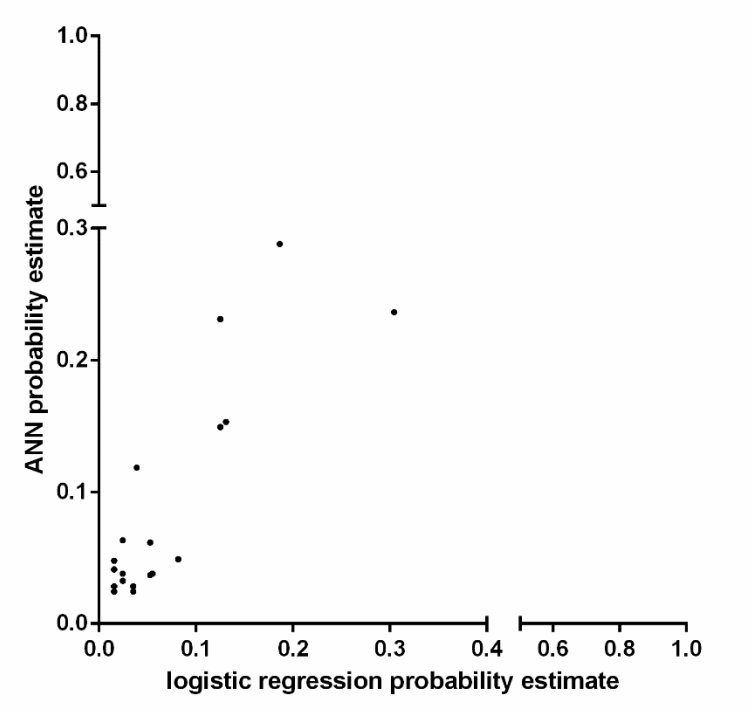


Figure 10 - Scatter plot of the probability of adverse events predicted by logistic regression and attribute matching for the patients with a match size of at least 30 patients

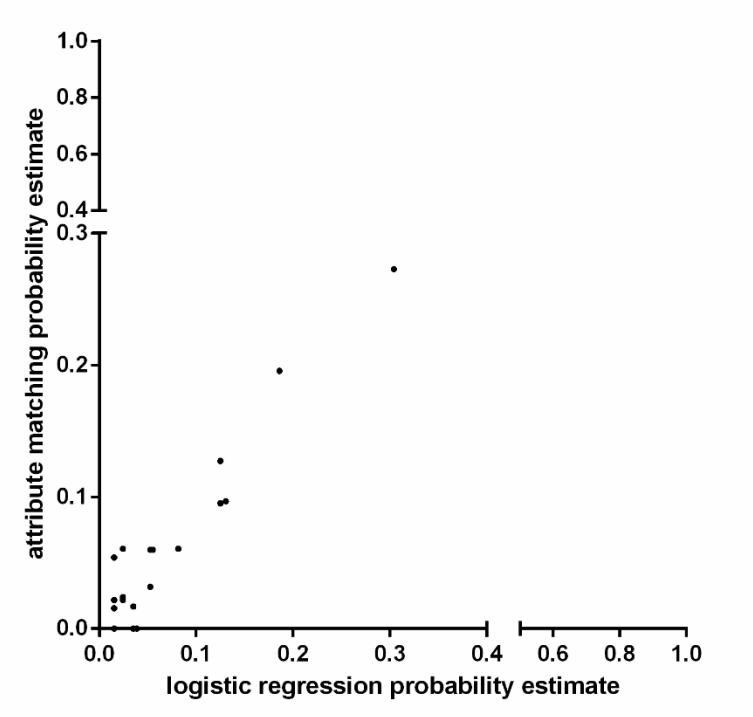
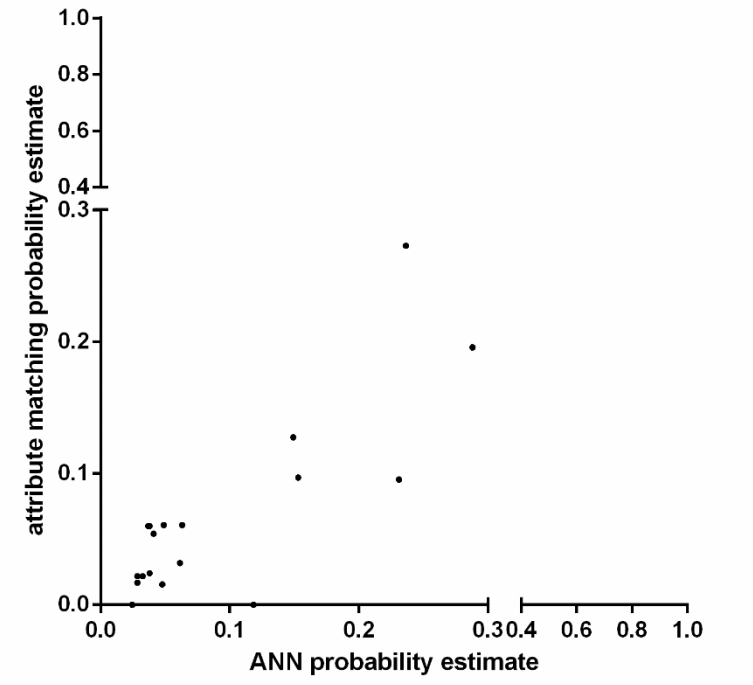


Figure 11 - Scatter plot of the probability of adverse events predicted by ANN and attribute matching for the patients with a match size of at least 30 patients



Prospective validation of the attribute matching

Even if the small number of patients in the derivation dataset does not allow to draw conclusions on the selected predictors or the estimate precision, we applied attribute matching to the patients of the prospective database to understand the possible strengths and weaknesses of this approach.

Among the 354 patients of the prospective database, we identified 179 unique pre-test probability estimates. As expected, only 2 of the 179 (1%) pre-test probability estimates had a match size ≥10 patients and most of them had a match size of 1 patient.

Among the 515 unique combinations of predictors of the derivation database, 363 had no match in the validation dataset, while 33 patients (27 unique combinations of predictors) in the validation database had no match in the derivation one. Therefore, 33 patients (9%) could not be provided with a risk estimate with attribute matching. For the remaining 321 patients, the median estimated probability of serious outcomes was 2% (IQR 0% to 14%). The median probability estimates of logistic regression and ANNs were 8% (IQR 5% to 12%) and 9% (IQR 5% to 25%), respectively.

To assess the correlation between the attribute matching predicted probabilities in the derivation and validation cohort, we calculated the Spearman's rank correlation coefficient. The coefficient was 0.13, thus suggesting a very weak correlation between the probabilities predicted by different combinations of predictors. This can be due to the small number of subjects in both the derivation and validation cohorts and needs to be verified.

The AUC of attribute matching in the prospective cohort was 0.589 (Figure 4). The predictive accuracy according to the three pre-defined cut-offs is reported in Table 16.

Figure 12 - ROC curve for the validation of attribute matching

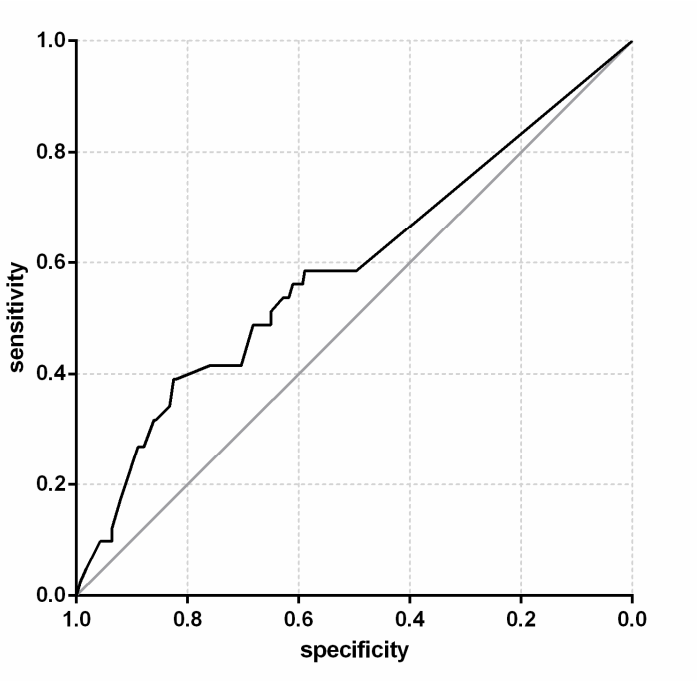


Table 16 - Predictive accuracy of attribute matching according to different cut-offs

Cut-off	FP[#] (n)	FN[#] (n)	Sensitivity[#]	Specificity[#]	PPV[#]	NPV[#]	Patients with a decision[§] (n)
2%	139	17	0.58	0.50	0.15	0.89	158 (45%)
30%	18	36	0.12	0.94	0.22	0.88	23 (7%)
50%	12	37	0.10	0.96	0.25	0.88	16 (5%)

Table legend: n: number; FP: false positives; FN: false negatives; PPV: positive predictive value; NPV: negative predictive value; §: proportion of patients below the 2% or above the 30% and 50% thresholds for discharge and admission on the total number of patients (i.e. 354); #: 321 patients considered in this estimate.

Considering only patients for whom the match size in the derivation cohort was ≥10 patients (the most represented combinations of predictors), we could provide a probability estimate in 157 patients (44%). The predictive accuracy according to the three pre-defined cut-offs is reported in Table 17.

Table 17 - Predictive accuracy of attribute matching according to different cut-offs in 157 patients with at least 10 matched patients in the derivation cohort

Cut-off	FP[#] (n)	FN[#] (n)	Sensitivity[#]	Specificity[#]	PPV[#]	NPV[#]	Patients with a decision[§] (n)
2%	93	3	0.77	0.35	0.10	0.94	54 (15%)
30%	0	12	0.08	1	1	0.92	1 (0.3%)
50%*	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	0

*Table legend: n: number; FP: false positives; FN: false negatives; PPV: positive predictive value; NPV: negative predictive value; *no patient had an estimated probability above 50%; §: proportion of patients below the 2% or above the 30% and 50% thresholds for discharge and admission on the total number of patients (i.e. 354); #: 157 patients considered in this estimate.*

Considering only patients for whom the match size in the derivation cohort was ≥30 patients, we could provide a probability estimate in 52 patients (15%). The predictive accuracy according to the three pre-defined cut-offs is reported in Table 18.

Table 18 - Predictive accuracy of attribute matching according to different cut-offs in 52 patients with at least 30 matched patients in the derivation cohort

Cut-off	FP[#] (n)	FN[#] (n)	Sensitivity[#]	Specificity[#]	PPV[#]	NPV[#]	Patients with a decision[§] (n)
2%	27	1	0.68	0.45	0.08	0.97	23 (6%)
30%*	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	0
50%*	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	0

*Table legend: n: number; FP: false positives; FN: false negatives; PPV: positive predictive value; NPV: negative predictive value; *no patient had an estimated probability above 30% and 50%; §: proportion of patients below the 2% or above the 30% and 50% thresholds for discharge and admission on the total number of patients (i.e. 354); #: 52 patients considered in this estimate.*

Some practical examples

To understand the possible strengths and weaknesses of all the existing risk prediction tools, logistic regression, ANN and attribute matching, we randomly selected 10 patients from the prospective database and applied all the above tools to estimate the predicted probability of adverse events. The cases descriptions and predicted probabilities for each patient, together with the ED physician's perceived risk are reported in Table 19.

Table 19 - Example clinical cases with the probabilities predicted by the single tools and clinical judgement

Case	Predicted probabilities
69 years-old man Previous myocardial infarction with preserved left ventricular ejection fraction Postprandial syncope without prodromes while sitting BP 160/80 mmHg, HR 80, peripheral oxygen saturation 98%, respiratory rate 14, normal body temperature ECG: Q waves in the inferior leads	OESIL: 4 (high risk) – 57.1% 1-year all-cause mortality SFSR: 0 (low risk) – 0.8% of 7-day serious adverse events (negative LR 0.06 with a pre-test probability of 11.5%) Canadian: 1 (medium risk) – 3.1% 30-day serious adverse events Neural networks: 48% of 10-day serious adverse events Multivariate analysis: 30% of 10-day serious adverse events Attribute matching: 20/3388, 10/20 events, 50% (95% CI 30-70%) 10-day serious adverse events (2 deaths) ED physician risk assessment: high
47 years-old man Unremarkable past medical history Syncope while sitting with prodromes (lightheadedness) BP 110/70 mmHg, HR 58, peripheral oxygen saturation 100%, respiratory rate 14, normal body temperature ECG: sinus rhythm 54 bpm	OESIL: 0 (low risk) – 0% 1-year all-cause mortality SFSR: 0 (low risk) – 0.8% of 7-day serious adverse events (negative LR 0.06 with a pre-test probability of 11.5%) Canadian: 0 (low risk) – 1.9% 30-day serious adverse events Neural networks: 29% of 10-day serious adverse events Multivariate analysis: 6% of 10-day serious adverse events Attribute matching: 50/3388, 3/50 events, 6% (95% CI 2-16%) of 10-day serious adverse events, 0 deaths

	ED physician risk assessment: intermediate
72 years-old man Affected by arterial hypertension Syncope without prodromes while standing BP 110/80 mmHg, HR 62, peripheral oxygen saturation 96%, respiratory rate 12, normal body temperature ECG: sinus rhythm 54 bpm	OESIL: 2 (high risk) – 19.6% 1-year all-cause mortality SFSR: 0 (low risk) – 0.8% of 7-day serious adverse events (negative LR 0.06 with a pre-test probability of 11.5%) Canadian: 0 (low risk) – 1.9% 30-day serious adverse events Neural networks: 24% of 10-day serious adverse events Multivariate analysis: 8% of 10-day serious adverse events Attribute matching: 14/3388, 2/14 events, 14% (95% CI 4-40%) of 10-day serious adverse events, 0 deaths ED physician risk assessment: intermediate
21 years-old man Unremarkable past medical history Syncope without prodromes while standing with trauma BP 100/60 mmHg, HR 90, peripheral oxygen saturation 100%, respiratory rate 16, normal body temperature ECG: sinus rhythm 65 bpm	OESIL: 1 (low risk) – 0.8% 1-year all-cause mortality SFSR: 0 (low risk) – 0.8% of 7-day serious adverse events (negative LR 0.06 with a pre-test probability of 11.5%) Canadian: 0 (low risk) – 1.9% 30-day serious adverse events Neural networks: 5% of 10-day serious adverse events Multivariate analysis: 2% of 10-day serious adverse events Attribute matching: 11/3388, 0/11 events, 0% (95% CI 0-26%) of 10-day serious adverse events, 0 deaths ED physician risk assessment: intermediate
47 years-old woman Unremarkable past medical history Syncope with prodromes while sitting BP 120/75 mmHg, HR 92, peripheral oxygen saturation 98% ECG: sinus rhythm 62 bpm	OESIL: 0 (low risk) – 0% 1-year all-cause mortality SFSR: 0 (low risk) – 0.8% of 7-day serious adverse events (negative LR 0.06 with a pre-test probability of 11.5%) Canadian: 0 (low risk) – 1.9% 30-day serious adverse events Neural networks: 2% of 10-day serious adverse events Multivariate analysis: 4% of 10-day serious adverse events Attribute matching: 59/3388, 1/59 events, 1.7% (95% CI 0.3-9%) of 10-day serious adverse events, 0 deaths ED physician risk assessment: intermediate
32 years-old man Unremarkable past medical history Syncope without prodromes while standing preceded by a painful stimulus BP 95/65 mmHg, HR 48, peripheral oxygen saturation 99%, respiratory rate 16 ECG: sinus rhythm 43 bpm with signs of vagal activity	OESIL: 1 (low risk) – 0.8% 1-year all-cause mortality SFSR: 0 (low risk) – 0.8% of 7-day serious adverse events (negative LR 0.06 with a pre-test probability of 11.5%) Canadian: -1 (low risk) – 1.2% 30-day serious adverse events Neural networks: 10% of 10-day serious adverse events Multivariate analysis: 6% of 10-day serious adverse events Attribute matching: 9/3388, 0/9 events, 0% (95% CI 0-30%) of 10-day serious adverse events, 0 deaths ED physician risk assessment: low
79 years-old man Unremarkable past medical history Syncope without prodromes while standing preceded by shortness of breath BP 125/80 mmHg, HR 80, peripheral oxygen saturation 100%, respiratory rate 20, normal body temperature	OESIL: 2 (high risk) – 19.6% 1-year all-cause mortality SFSR: 1 (high risk) – 24.7% of 7-day serious adverse events (positive LR 2.53 with a pre-test probability of 11.5%) Canadian: 0 (low risk) – 1.9% 30-day serious adverse events Neural networks: 25% of 10-day serious adverse events Multivariate analysis: 8% of 10-day serious adverse events

ECG: sinus rhythm 56 bpm	Attribute matching: 26/3388, 0/26 events, 0% (95% CI 0-13) of 10-day serious adverse events, 0 deaths ED physician risk assessment: high
75 years-old woman Affected by atrial fibrillation Syncope while standing with trauma preceded by shortness of breath, lightheadedness, nausea and vomiting BP 70/40 mmHg, HR 74, peripheral oxygen saturation 93%, respiratory rate 20, body temperature 37.7°C ECG: sinus rhythm 84 bpm, negative T waives in V3-V5	OESIL: 2 (high risk) – 19.6% 1-year all-cause mortality SFSR: 3 (high risk) – 24.7% of 7-day serious adverse events (positive LR 2.53 with a pre-test probability of 11.5%) Canadian: 3 (medium risk) – 8.1% 30-day serious adverse events Neural networks: 14% of 10-day serious adverse events Multivariate analysis: 21% of 10-day serious adverse events Attribute matching: 1/3388, 0/1 events, 0% (95% CI 0-79) of 10-day serious adverse events, 0 deaths ED physician risk assessment: high
66 years-old woman Previous syncope the year before Syncope while standing with trauma preceded by chest pain, shortness of breath and lightheadedness BP 120/70 mmHg, HR 88, peripheral oxygen saturation 97%, respiratory rate 23, normal body temperature ECG: sinus rhythm 82 bpm, V1 and V2 leads compatible with Brugada pattern	OESIL: 2 (high risk) – 19.6% 1-year all-cause mortality SFSR: 1 (high risk) – 24.7% of 7-day serious adverse events (positive LR 2.53 with a pre-test probability of 11.5%) Canadian: 0 (low risk) – 1.9% 30-day serious adverse events Neural networks: 33% of 10-day serious adverse events Multivariate analysis: 21% of 10-day serious adverse events Attribute matching: 0/3388, 0 events, 0% of 10-day serious adverse events, 0 deaths ED physician risk assessment: high
90 years-old man Previous syncope the year before. Known ischemic cardiomyopathy with reduce ejection fraction, congestive heart failure, pulmonary hypertension, valvular heart disease and arterial hypertension Syncope while sitting with trauma preceded by lightheadedness BP 120/70 mmHg, HR 80, peripheral oxygen saturation 91%, respiratory rate 24, normal body temperature ECG: sinus rhythm 84 bpm, negative T waives in the anterior and lateral leads	OESIL: 3 (high risk) – 34.7% 1-year all-cause mortality SFSR: 1 (high risk) – 24.7% of 7-day serious adverse events (positive LR 2.53 with a pre-test probability of 11.5%) Canadian: 1 (medium risk) – 3.1% 30-day serious adverse events Neural networks: 10% of 10-day serious adverse events Multivariate analysis: 30% of 10-day serious adverse events Attribute matching: 27/3388, 6/27 events, 22% (95% CI 11-41%) of 10-day serious adverse events, 1 death ED physician risk assessment: high

Table legend: BP: blood pressure; HR: heart rate; ECG: electrocardiogram; ED: Emergency Department; CI: Confidence Interval; OESIL: Osservatorio Epidemiologico sulla Sincope nel Lazio risk score; SFSR: San Francisco Syncope Rule.

The comparison of the risk predicted by different clinical decision tools, logistic regression, ANN, and attribute matching shows that there is a high heterogeneity within the same patient (Table 20).

Table 20 - Predicted probabilities according to different prediction tools, ANN, attribute matching and clinical judgement in the 10 example patients

Case n	OESIL		SFSR		Canadian Syncope Risk Score		ANN 10-day SAE (%)	Regression 10-day SAE (%)	Attribute matching		ED physician
	score	1-year mortality (%)	score	7-day SAE (%)	score	30-day SAE (%)			patients at risk*	10-day SAE (%)	
1	4 (high risk)	57.1	0 (low risk)	0.8	1 (medium risk)	3.1	48	30	20	50	High risk
2	0 (low risk)	0	0 (low risk)	0.8	0 (low risk)	1.9	25	6	50	6	Intermediate risk
3	2 (high risk)	19.6	0 (low risk)	0.8	0 (low risk)	1.9	24	8	14	14	Intermediate risk
4	1 (low risk)	0.8	0 (low risk)	0.8	0 (low risk)	1.9	5	2	11	0	Intermediate risk
5	0 (low risk)	0	0 (low risk)	0.8	0 (low risk)	1.9	2	4	59	1.7	Intermediate risk
6	1 (low risk)	0.8	0 (low risk)	0.8	-1 (low risk)	1.2	10	6	9	0	Low risk
7	2 (high risk)	19.6	1 (high risk)	24.7	0 (low risk)	1.9	25	8	26	0	High risk
8	2 (high risk)	19.6	3 (high risk)	24.7	3 (medium risk)	8.1	14	2	1	0	High risk
9	2 (high risk)	19.6	1 (high risk)	24.7	0 (low risk)	1.9	33	21	0	n.a.	High risk
10	3 (high risk)	34.7	1 (high risk)	24.7	1 (medium risk)	3.1	10	30	27	22	High risk

*Table legend: OESIL: Osservatorio Epidemiologico sulla Sincope nel Lazio risk score; SFSR: San Francisco Syncope Rule; ANN: Artificial Neural Network; ED: Emergency Department; SAE: serious adverse events; *: number of patients with the same combination of risk factors; n.a.: non applicable.*

DISCUSSION

Risk stratification and appropriate disposition are challenging in the management of syncope patients in the ED. Syncope is a common symptom for a large number of conditions spanning from benign to life-threatening diseases, and its prognosis is highly heterogeneous [1,52].

Recent evidence showed that clinical prediction tools have failed in identifying the risk associated with the conditions underlying syncope [16,39,40]. This could be due to different causes. First, the heterogeneity in syncope etiologies makes it difficult to find a single tool to identify all of them. Moreover, as syncope-related serious adverse events are relatively uncommon, the traditionally derived risk stratification tools may be overfit. For example, to be reliable in the short-term prediction of adverse events, multivariate analysis would need a high number of outcomes (about 10 every predictor identified). To overcome this problem, some of the available clinical prediction tools have considered a long-lasting time scale, i.e. 1 year, instead of a more useful but shorter one, for example 7–30 days. The use of a 1 year time scale gave the possibility to obtain a larger number of unfavorable events making the multivariate analysis approach robust enough [41,53]. Otherwise, the required sample size would be very large, and no currently available study has the required statistical power. Conversely, other studies considering short-term adverse events used a recursive partitioning model to derive a 'yes/no' rule. However, continuous risk estimates may be more clinically relevant for making disposition (i.e. admit or discharge) decisions.

Artificial neural networks could have overcome the limitations of the traditionally derived clinical prediction tools given their capability to approximate any type of functions, especially the non-linear ones. Indeed, the possibility to set more layers linked together with different functions results in both the possibility to generalize results and to improve the effectiveness of input data analysis.

To assess the possible strengths and weaknesses and to compare the different statistical methodologies to derive prediction tools, we decided to derive both a multivariate logistic regression model and an ANN on the same large retrospective database and to prospectively validate them in a new dataset of 354 patients. Our data show that both the logistic model and ANN have a poor predictive accuracy in identifying serious adverse events in patients presenting with syncope to the ED and its use should be discouraged for both admission and discharge disposition. Moreover, the 10 example patients show how heterogeneously the risk of adverse events is estimated by the different tools. This means that, even if their accuracy is fair in the mean patient, we do not know what tool performs better in the single patient. Therefore, when assessing a specific patient, we know that some of them are probably correctly estimating the risk, but we do not know which ones.

This report introduces a method to estimate of the probability of serious adverse based upon computer assisted, database-derived, attribute matching. The system operates by allowing the clinician to

input a predefined set of clinical attributes for a subject for whom the probability of a serious outcome is desired. When executed, a computer program queries a large patient database, and returns only the patients who share the identical attribute profile as the new patient being evaluated. The proportion of these attribute-matched subjects who had a clinical outcome of interest comprises the point estimate of the probability. This process is similar to the definition of pre-test probability by an expert clinician, which, having seen many patients who had similar clinical characteristics as the patient under consideration, could provide an estimate of the probability of something bad happening.

We recognize that as the number and complexity of the input attributes increases, this will create a more specific and potentially more accurate clinical profile, but at a cost of reduced match size if the reference database remains the same size. In theory, the ideal attribute matching system would allow a very detailed clinical profile to be matched against a very large reference database. In the present work, we used a ten-attribute profile and a 3388-patient database. Only 1052 (31%) of them had a combination of predictors with a match size of 30 or more. Therefore, our data do not allow to offer a clinically useful prediction tool at this stage, but this method seems promising, as it has some advantages as compared to model-derived clinical decision tools. Indeed, the successful use of a model to predict the probability of a serious outcome requires that the results are reproduced in an external validation so that both the external validity and robustness of the model are verified. Moreover, models require that the predictors are assigned a weight that allow to estimate the risk of adverse events in every patient, also in those that had no matching subject in the derivation database (for example for patients that have a rare condition). Attribute matching differs from scoring systems derived from logistic regression or ANN, which use predictor variables expressed by an individual patient under consideration to guide that patient into a predefined category that predicts a probability. This outcome probability is estimated from knowledge (i.e., the magnitude of importance of predictor variables) manifested by the patients that were used to construct the logit equation or ANN. On the other hand, attribute matching works in reverse fashion. Instead of placing the patient under consideration into a category, the computer program finds the patients from a reference database who ‘‘look like’’ the patient insofar as they are identical on the binary predictor variables. Therefore, the risk of patients with an uncommon combination of predictors, might not be able at all to find a match in the derivation dataset. However, being aware that the patient’s estimated probability might be based on very limited evidence, will allow both the clinician and the patient to take a decision conscious that it might be based on uncertainty, rather than deciding on the false confidence provided by models.

Attribute matching has some other advantages: 1) The possibility to have as output not only the probability of a composite serious outcome, but a detailed risk profile based on the probability of different outcomes, will allow a more personalized decision making. Also, the possibility to make the risk profile

explicit could allow to share the decision with the patient. 2) The need for a large dataset should make external validity not a problem, because the collection of patients from multiple contexts is essential to reach the required dataset size. 3) As there is no need for calculations and model creations, patients could be always added to the dataset thus increasing the probability estimate precision. 4) The flexibility of attribute matching would allow to consider different predictors in different patients, thus allowing a more precise estimate.

Study limitations

Some limitations of this study should be acknowledged. Besides the above peculiar characteristics of syncope, that make it difficult to derive accurate prediction tools, the choice of appropriate predictors and the correct identification of outcomes is crucial. The database we used for derivation was collected for different purposes. Moreover, the included studies differed as to endpoints and clinical variables considered. In addition, there is heterogeneity in health system organizations, data collection forms, and ECG interpretation. However, as a large syncope database does not exist, we did our best to collect a large number of patients and to make the data as homogeneous as possible so that they could be analyzed together.

Moreover, the optimal database size and number of attributes that can be used to create valuable probability estimates remains uncertain. It remains unknown how many patients in the reference database must be returned for any given profile to provide a reliable estimate. Indeed, despite the large size of the reference database, several profiles returned zero patients, suggesting that patients with these patterns will be rarely encountered, but otherwise providing little inference into the probability of the outcome with this uncommon presentation. We could argue that increasing the match size by 10-times (thus enrolling about 34000 patients) will provide a precise estimate for the combinations of predictors that already had matching patients in the derivation dataset. However, the number of subjects to enroll to observe patients with all the possible unique pre-test probability estimates cannot be calculated, as some of them could be extremely rare or even unreal.

Finally, the use of attribute matching might be perceived as rudimental as compared to mathematical models. However, the current possibility to collect large datasets through international collaborations, big data and internet might overtake the need for models, that have been created to overcome the problem of making an estimate with limited observations.

CONCLUSIONS AND FUTURE PERSPECTIVES

As the model-derived clinical decision tools have failed in the accurate identifications of patients that will experience serious outcomes, the attribute matching could provide an alternative solution to the risk stratification of patients with syncope. Indeed, the prediction tools based on models, such as logistic regression and neural networks provide a risk estimate in every case, also in patients whose clinical characteristics are different from each patient's characteristics in the derivation cohort. Conversely, the attribute matching would need tremendously large datasets to provide accurate risk estimates, especially in patients with an uncommon combination of predictors, and might not be able at all to find a match in the derivation dataset. However, being aware of the patient's estimated probability based on previous experience will allow both the clinician and the patient to take a decision conscious that it might be based on uncertainty, rather than deciding on the false confidence provided by models. This is crucial in the perspective of a modern medicine increasingly based on personalized medicine and shared decision making.

Future studies should focus on international collaborations to build large prospective datasets. However, even if some steps towards the definition of standardized inclusion criteria, data collection and outcome assessment have already been made [42,53], the low inter-rater agreement in the assessment of both predictors and outcome measures could undermine the predictive accuracy of CDTs [20]. After facing these problems, efforts should be made to assess if attribute matching adds any value to both the CDTs based on statistical models and the implicit estimate of probability from clinicians with variable experience. Moreover, the introduction of new and more complex input attributes (for example the specific ECG abnormality) and the possibility to provide as output a detailed risk profile (i.e. the risk of different adverse events) rather than the probability of a composite outcome will create a more specific and potentially more accurate clinical profile.

REFERENCES

[1] A. Moya, R. Sutton, F. Ammirati, J.J. Blanc, M. Brignole, J.B. Dahm, J.C. Deharo, J. Gajek, K. Gjesdal, A. Krahn, M. Massin, M. Pepi, T. Pezawas, G.R. Ruiz, F. Sarasin, A. Ungar, J.G. van Dijk, E.P. Walma, W. Wieling, Guidelines for the diagnosis and management of syncope (version 2009), Eur Hear. J. 30 (2009) 2631–2671.

[2] L. Furlan, G. Costantino, M. Solbiati, P. Alboni, Definition and classification of transient loss of consciousness, 2015. doi:10.1007/978-3-319-09102-0_3.

[3] W.N. Kapoor, M. Karpf, S. Wieand, J.R. Peterson, G.S. Levey, A prospective evaluation and follow-up of patients with syncope, N.Engl.J Med. 309 (1983) 197–204.

[4] B.C. Sun, Quality-of-life, health service use, and costs associated with syncope, Prog Cardiovasc Dis. 55 (2013) 370–375.

[5] M.S. Rose, M.L. Koshman, S. Spreng, R. Sheldon, The relationship between health-related quality of life and frequency of spells in patients with syncope., J. Clin. Epidemiol. 53 (2000) 1209–16.

[6] E.S. Soteriades, J.C. Evans, M.G. Larson, M.H. Chen, L. Chen, E.J. Benjamin, D. Levy, Incidence and prognosis of syncope., N. Engl. J. Med. 347 (2002) 878–85. doi:10.1056/NEJMoa012407.

[7] K.S. Ganzeboom, N. Colman, J.B. Reitsma, W.K. Shen, W. Wieling, Prevalence and triggers of syncope in medical students., Am. J. Cardiol. 91 (2003) 1006–8, A8.

[8] A. Serletis, S. Rose, A.G. Sheldon, R.S. Sheldon, Vasovagal syncope in medical students and their first-degree relatives, Eur Hear. J. 27 (2006) 1965–1970.

[9] K.S. Ganzeboom, G. Mairuhu, J.B. Reitsma, M. Linzer, W. Wieling, D.N. van, Lifetime cumulative incidence of syncope in the general population: a study of 549 Dutch subjects aged 35-60 years, J Cardiovasc. 17 (2006) 1172–1176.

[10] L.Y. Chen, B.J. Gersh, D.O. Hodge, W. Wieling, S.C. Hammill, W.K. Shen, Prevalence and clinical outcomes of patients with multiple potential causes of syncope, Mayo Clin Proc. 78 (2003) 414–420.

[11] F. D'Ascenzo, G. Biondi-Zoccai, M.J. Reed, G.Z. Gabayan, M. Suzuki, G. Costantino, R. Furlan, A. Del Rosso, F.P. Sarasin, B.C. Sun, M.G. Modena, F. Gaita, Incidence, etiology and predictors of adverse outcomes in 43,315 patients presenting to the Emergency Department with syncope: an international meta-analysis., Int. J. Cardiol. 167 (2013) 57–62. doi:10.1016/j.ijcard.2011.11.083.

[12] F. Colivicchi, F. Ammirati, D. Melina, V. Guido, G. Imperoli, M. Santini, Development and prospective validation of a risk stratification system for patients with syncope in the emergency department: the OESIL risk score, Eur.Heart J. 24 (2003) 811–819.

[13] M.J. Reed, D.E. Newby, A.J. Coull, R.J. Prescott, K.G. Jacques, A.J. Gray, The ROSE (risk stratification of syncope in the emergency department) study, J Am.Coll.Cardiol. 55 (2010) 713–721.

[14] J. V Quinn, I.G. Stiell, D.A. McDermott, K.L. Sellers, M.A. Kohn, G.A. Wells, Derivation of the San Francisco Syncope Rule to predict patients with short-term serious outcomes, Ann Emerg.Med. 43 (2004) 224–232.

[15] A. Birnbaum, D. Esses, P. Bijur, A. Wollowitz, E.J. Gallagher, Failure to validate the San Francisco Syncope Rule in an independent emergency department population., Ann. Emerg. Med. 52 (2008) 151–9. doi:10.1016/j.annemergmed.2007.12.007.

[16] L.A. Serrano, E.P. Hess, M.F. Bellolio, M.H. Murad, V.M. Montori, P.J. Erwin, W.W. Decker, Accuracy and quality of clinical decision rules for syncope in the emergency department: a systematic review and meta-analysis, Ann.Emerg.Med. 56 (2010) 362–373.

[17] V. Thiruganasambandamoorthy, E.P. Hess, E. Turko, J.J. Perry, G.A. Wells, I.G. Stiell, Outcomes in Canadian emergency department syncope patients--are we doing a good job?, J Emerg.Med. 44 (2013) 321–328.

[18] A. Del Rosso, A. Ungar, R. Maggi, F. Giada, N.R. Petix, S.T. De, C. Menozzi, M. Brignole, Clinical predictors of cardiac syncope at initial evaluation in patients referred urgently to a general hospital: the EGSYS score, Heart. 94 (2008) 1620–1626.

[19] S.A. Grossman, J. Bar, C. Fischer, L.A. Lipsitz, L. Mottley, K. Sands, P. Zimetbaum, N.I. Shapiro, Reducing admissions utilizing the Boston Syncope Criteria, J Emerg Med. 42 (2012) 345–352.

[20] V. Thiruganasambandamoorthy, K. Kwong, G.A. Wells, M.L.A. Sivilotti, M. Mukarram, B.H. Rowe, E. Lang, J.J. Perry, R. Sheldon, I.G. Stiell, M. Taljaard, Development of the Canadian Syncope Risk Score to predict serious adverse events after emergency department assessment of syncope, Can. Med. Assoc. J. 188 (2016) E289–E298. doi:10.1503/cmaj.151469.

[21] T.G. McGinn, G.H. Guyatt, P.C. Wyer, C.D. Naylor, I.G. Stiell, W.S. Richardson, for the E.-B.M.W. for the Evidence-Based Medicine Working Group, Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group., JAMA. 284 (2000) 79–84. doi:10.1001/jama.284.1.79.

[22] G. Guyatt, S. Walter, H. Shannon, D. Cook, R. Jaeschke, N. Heddle, Basic statistics for clinicians: 4. Correlation and regression., CMAJ. 152 (1995) 497–504. http://www.ncbi.nlm.nih.gov/pubmed/7859197 (accessed October 17, 2017).

[23] T.E. Rudy, J.A. Kubinski, J.R. Boston, Multivariate analysis and repeated measurements: A primer, J. Crit. Care. 7 (1992) 30–41. doi:10.1016/0883-9441(92)90006-S.

[24] E.F. Cook, L. Goldman, Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis., J. Chronic Dis. 37 (1984) 721–31. http://www.ncbi.nlm.nih.gov/pubmed/6501544 (accessed October 17, 2017).

[25] W.. Baxt, Application of artificial neural networks to clinical medicine, Lancet. 346 (1995) 1135–1138. doi:10.1016/S0140-6736(95)91804-3.

[26] M.R. Chernick, Bootstrap Methods, Volume 619, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2007. doi:10.1002/9780470192573.

[27] E.W. Steyerberg, Statistical Models for Prediction, in: Clin. Predict. Model. A Pract. Approach to Dev. Validation, Updat., Springer New York, New York, NY, 2009: pp. 53–82. doi:10.1007/978-0-387-77244-8_4.

[28] D.R. Cox, Regression Models and Life-Tables, J. R. Stat. Soc. Ser. B. 34 (1972) 187–220. http://www.jstor.org/stable/2985181 (accessed October 17, 2017).

[29] E.W. Steyerberg, Clinical Prediction Models, 1st ed., Springer New York, New York, NY, 2009. doi:10.1007/978-0-387-77244-8.

[30] C. Strobl, J. Malley, G. Tutz, An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests., Psychol. Methods. 14 (2009) 323–348. doi:10.1037/a0016973.

[31] L. Breiman, J. Friedman, C.J. Stone, R. Olshen, Classification and regression trees, Chapman & Hall, 1984.

[32] L. Breiman, Random forests, (2001).

[33] W. Hannover, M. Richard, N.B. Hansen, Z. Martinovich, H. Kordy, A Classification Tree Model for Decision-Making in Clinical Practice: An Application Based on the Data of the German Multicenter

Study on Eating Disorders, Project TR-EAT, Psychother. Res. 12 (2002) 445–461. doi:10.1080/713664470.

[34] J. V Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes., J. Clin. Epidemiol. 49 (1996) 1225–31. http://www.ncbi.nlm.nih.gov/pubmed/8892489 (accessed October 17, 2017).

[35] S.S. Cross, R.F. Harrison, R.L. Kennedy, Introduction to neural networks, Lancet. 346 (1995) 1075–1079.

[36] G. Falavigna, Financial ratings with scarce information: A neural network approach, Expert Syst. Appl. 39 (2012) 1784–1792. isi:000298027300021.

[37] F. Bektas, C. Eken, S. Soyuncu, I. Kilicaslan, Y. Cete, Artificial neural network in predicting craniocervical junction injury: an alternative approach to trauma patients, Eur J Emerg Med. 15 (2008) 318–323.

[38] R.F. Harrison, R.L. Kennedy, Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation, Ann Emerg Med. 46 (2005) 431–439.

[39] R.T. Saccilotto, C.H. Nickel, H.C. Bucher, E.W. Steyerberg, R. Bingisser, M.T. Koller, San Francisco Syncope Rule to predict short-term serious outcomes: a systematic review, CMAJ. 183 (2011) E1116–E1126.

[40] G. Costantino, G. Casazza, M. Reed, I. Bossi, B. Sun, A. Del Rosso, A. Ungar, S. Grossman, F. D'Ascenzo, J. Quinn, D. McDermott, R. Sheldon, R. Furlan, Syncope risk stratification tools vs clinical judgment: an individual patient data meta-analysis., Am. J. Med. 127 (2014) 1126.e13-25. doi:10.1016/j.amjmed.2014.05.022.

[41] G. Costantino, B.C. Sun, F. Barbic, I. Bossi, G. Casazza, F. Dipaola, D. McDermott, J. Quinn, M.J. Reed, R.S. Sheldon, M. Solbiati, V. Thiruganasambandamoorthy, D. Beach, N. Bodemer, M. Brignole, I. Casagranda, A. Del Rosso, P. Duca, G. Falavigna, S.A. Grossman, R. Ippoliti, A.D. Krahn, N. Montano, C.A. Morillo, B. Olshansky, S.R. Raj, M.H. Ruwald, F.P. Sarasin, W.-K. Shen, I. Stiell, A. Ungar, J. Gert van Dijk, N. van Dijk, W. Wieling, R. Furlan, Syncope clinical management in the emergency department: a consensus from the first international workshop on syncope risk stratification in the emergency department., Eur. Heart J. 37 (2016) 1493–8. doi:10.1093/eurheartj/ehv378.

[42] B.C. Sun, V. Thiruganasambandamoorthy, J.D. Cruz, Standardized reporting guidelines for emergency department syncope risk-stratification research, Acad Emerg Med. 19 (2012) 694–702.

[43] S.A. Grossman, C. Fischer, L.A. Lipsitz, L. Mottley, K. Sands, S. Thompson, P. Zimetbaum, N.I. Shapiro, Predicting adverse outcomes in syncope, J Emerg Med. 33 (2007) 233–239.

[44] A. Ungar, A. Del Rosso, F. Giada, A. Bartoletti, R. Furlan, F. Quartieri, A. Lagi, A. Morrione, C. Mussi, M. Lunati, G. De Marchi, T. De Santo, N. Marchionni, M. Brignole, Early and late outcome of treated patients referred for syncope to emergency department: the EGSYS 2 follow-up study, Eur.Heart J. 31 (2010) 2021–2026.

[45] G. Costantino, F. Perego, F. Dipaola, M. Borella, A. Galli, G. Cantoni, S. Dell'Orto, S. Dassi, N. Filardo, P.G. Duca, N. Montano, R. Furlan, Short- and long-term prognosis of syncope, risk factors, and role of hospital admission: results from the STePS (Short-Term Prognosis of Syncope) study., J. Am. Coll. Cardiol. 51 (2008) 276–83. doi:10.1016/j.jacc.2007.08.059.

[46] B.C. Sun, C.M. Mangione, G. Merchant, T. Weiss, G.Z. Shlamovitz, G. Zargaraff, S. Shiraga, J.R. Hoffman, W.R. Mower, External validation of the San Francisco Syncope Rule., Ann. Emerg. Med. 49 (2007) 420–7, 427–4. doi:10.1016/j.annemergmed.2006.11.012.

[47] G. Costantino, G. Falavigna, M. Solbiati, I. Casagranda, B.C. Sun, S.A. Grossman, J. V Quinn, M.J.

Reed, A. Ungar, N. Montano, R. Furlan, R. Ippoliti, Neural networks as a tool to predict syncope risk in the Emergency Department., Europace. (2016). doi:10.1093/europace/euw336.

[48] J.A. Kline, D.M. Courtney, M.P. Than, K. Hogg, C.D. Miller, C.L. Johnson, H.A. Smithline, Accuracy of very low pretest probability estimates for pulmonary embolism using the method of attribute matching compared with the wells score, Acad. Emerg. Med. 17 (2010) 133–141. doi:10.1111/j.1553-2712.2009.00648.x.

[49] J.A. Kline, W.B. Stubblefield, Clinician gestalt estimate of pretest probability for acute coronary syndrome and pulmonary embolism in patients with chest pain and dyspnea, Ann. Emerg. Med. 63 (2014) 275–280. doi:10.1016/j.annemergmed.2013.08.023.

[50] J.A. Kline, C.L. Johnson, C. V Pollack, D.B. Diercks, J.E. Hollander, C.D. Newgard, J.L. Garvey, Pretest probability assessment derived from attribute matching, BMC Med. Inform. Decis. Mak. 5 (2005) 26. doi:10.1186/1472-6947-5-26.

[51] P.E. Shrout, J.L. Fleiss, Intraclass correlations: Uses in assessing rater reliability., Psychol. Bull. 86 (1979) 420–428. doi:10.1037/0033-2909.86.2.420.

[52] W.-K. Shen, R.S. Sheldon, D.G. Benditt, M.I. Cohen, D.E. Forman, Z.D. Goldberger, B.P. Grubb, M.H. Hamdan, A.D. Krahn, M.S. Link, B. Olshansky, S.R. Raj, R.K. Sandhu, D. Sorajja, B.C. Sun, C.W. Yancy, 2017 ACC/AHA/HRS Guideline for the Evaluation and Management of Patients With Syncope: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines, and the Heart Rhythm Society, Circulation. (2017) CIR.0000000000000499. doi:10.1161/CIR.0000000000000499.

[53] B.C. Sun, G. Costantino, F. Barbic, I. Bossi, G. Casazza, F. Dipaola, D. McDermott, J. Quinn, M. Reed, R.S. Sheldon, M. Solbiati, V. Thiruganasambandamoorthy, A.D. Krahn, D. Beach, N. Bodemer, M. Brignole, I. Casagranda, P. Duca, G. Falavigna, R. Ippoliti, N. Montano, B. Olshansky, S.R. Raj, M.H. Ruwald, W.-K. Shen, I. Stiell, A. Ungar, J.G. van Dijk, N. van Dijk, W. Wieling, R. Furlan, Priorities for Emergency Department Syncope Research, Ann. Emerg. Med. 64 (2014) 649–655.e2. doi:10.1016/j.annemergmed.2014.04.014.