Derivation and Internal Validation of a Clinical Prediction Tool for Adult Emergency Department Syncope Patients

Kenneth Kwong

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the MSc degree in Epidemiology

School of Epidemiology, Public Health and Preventive Medicine,
Faculty of Medicine,
University of Ottawa

BRIEF ABSTRACT

Syncope is a common Emergency Department (ED) presentation. An important proportion of syncope patients are at risk of developing serious adverse events (SAEs), such as deaths or arrhythmias following ED disposition. Currently, no clinically-useful decision tool exists to reliably identify high-risk patients.

This study derived a clinical decision tool to identify syncope patients at risk of developing SAEs after ED disposition. This study also examined key methodological considerations involved in deriving decision tools by comparing two different methodological approaches: a traditional and modern approach. The traditional approach led to an eight-variable decision tool that allowed simple clinical interpretation and use. The modern approach, which aims to avoid data-driven methodology and statistical overfitting, was used to derive a ten-variable decision tool. Both decision tools displayed acceptable and comparable performance in internal validation studies (c-statistic 0.87, 95% confidence interval 0.84-0.89). A future external validation study is required to comprehensively compare the methods.

(Word count = 150; 150 allowed)

EXECUTIVE SUMMARY

Introduction: Syncope is a common presentation in the Emergency Department (ED). An important proportion of syncope patients are at risk of developing serious adverse events (SAEs), such as deaths, myocardial infarctions, or cardiac arrhythmias after ED disposition. Despite numerous efforts to derive a decision tool to risk-stratify syncope patients, a reliable and clinically-useful tool has yet to be developed. Many key methodological considerations, such as the handling of missing data, modelling of continuous predictors, and specification of variables included in the tool, are involved in the derivation of decision tools. Recently, there has been an emerging interest in improving the methodology used in decision tool research.

Objectives: The overall objective was to develop a valid and clinically-useful decision tool to identify syncope patients at risk of SAEs after ED disposition. An additional embedded objective was to empirically compare some key methodological considerations involved in deriving decision tools using two different methodological approaches.

Methods: Data from a multi-centre prospective cohort study conducted from 2010 to 2014 was used. Standardized variables from patient history, clinical evaluation, and investigations were collected at the ED visit for syncope. The outcome of interest was defined as SAEs 30-days after ED disposition. The traditional approach employed bivariable screening, categorization of continuous predictors, and stepwise variable selection to generate the multivariable logistic regression prediction model. In comparison, the modern approach modelled continuous variables using non-linear transformations, and pre-specified predictors based on clinical knowledge. Multiple imputation, bootstrap internal validation, and model performance assessments were performed in both approaches.

Results: This study included 4,611 patients, of which 147 (3.5%) had the outcome of interest. The final decision tool derived from the traditional approach consisted of 8 variables: history of heart disease, precipitating factors for vasovagal syncope, MD suspected diagnosis, systolic blood pressure <90 or >180 mmHg, troponin (>99%ile), QRS duration >130 msec, corrected QT interval >480 msec, and abnormal QRS axis (<-30 or >110). The model had acceptable discrimination and calibration (optimism-corrected c-statistic = 0.865, Hosmer-Lemeshow goodness of fit p-value=0.11). In comparison, the final decision tool derived from the modern approach consisted of 10 variables: history of heart disease, MD suspected diagnosis, systolic blood pressure at triage, troponin (>99%ile), corrected QT interval, history of cerebrovascular disease, palpitations, physical exertion, prodrome, and age. The model also had acceptable discrimination and calibration (optimism-corrected c-statistic = 0.867, Hosmer-Lemeshow goodness of fit p-value=0.78). Both decision tools displayed acceptable and comparable performance in internal validation studies.

Conclusions: Two clinical decision tools to risk-stratify ED syncope were successfully developed in this study. Once validated, the decision tools could help physicians reliably assess risk in ED syncope, ultimately having implications in clinical decision-making. In the methodological comparison, the decision tools derived using the two parallel approaches displayed comparable performance. Future external validation studies will be needed to fully assess and compare the performance of these decision tools.

ACKNOWLEDGEMENTS

I would not have been able to complete this thesis without the unwavering support of my supervisors, Venkatesh Thiruganasambandamoorthy and Monica Taljaard. Not only have they provided critical mentorship throughout this thesis project, but have also helped develop my academic career in ways that I previously thought were unimaginable. I would like to thank you both for your ongoing confidence in me.

This project would not have been possible without the members of our research team, including Muhammad Mukarram, Soo-Min Kim, Hina Chaudry, Brittany Mutsaers, Roos Ramaekers, Tauralee Tenn, and Sarah Gaudet. They all worked tirelessly to run the study and collect the data used in the present thesis. Similarly, I would like to thank all the research teams, nurses, and clinicians at each of the sub-sites for their remarkable dedication.

I am forever grateful to my parents, Herman and Elisa, for their endless encouragement to follow my dreams. I would also like to thank my partner, Carmen, for her ongoing support and her encouragement to follow my passion in spite of the burden that it can place on our relationship.

I would like to thank the Canadian Institutes of Health Research, Ontario Innovation Fund, and Physician's Services Incorporated for funding the study. I would also like to kindly thank the School of Epidemiology, Public Health and Preventive Medicine; the University of Ottawa's Faculty of Medicine; and the Canadian Institutes of Health Research for the personal financial support during my tenure as a graduate student.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Syncope

Syncope is a transient loss of consciousness due to temporary global cerebral hypoperfusion[1]. Syncope is characterised by its rapid onset, short duration, and spontaneous complete recovery. The cerebral hypoperfusion that underlies syncope is the result of either decreased cardiac output or decreased peripheral vascular resistance, most often a combination of the two[1].

Syncope can be classified into three categories based on the underlying pathophysiological origin: reflex syncope, syncope due to orthostatic hypotension, or cardiac syncope[1]. Reflex syncope occurs when cardiovascular reflexes, which normally serve to control blood pressure, illicit an inappropriate response to a trigger, leading to vasodilation or bradycardia, more commonly a combination of both. Syncope due to orthostatic hypotension is characterised by an abnormal decrease in systolic blood pressure when standing due to impaired autonomic nervous system function, leading to insufficient vasoconstriction. Cardiac syncope occurs when there is impairment of the cardiac structure or function, ultimately leading to decreased cardiac output. Arrhythmias and structural heart disease are two cardiac abnormalities that can underlie cardiac syncope by altering cardiac output, thereby leading to cerebral hypoperfusion. Although syncope is classified into three distinct categories based on pathophysiological origin, the causes of syncope are often multifactorial and are often difficult to identify during a clinical evaluation[1].

Syncope is a common presentation in the Emergency Department (ED) accounting for 1 to 3% of ED visits, and up to 6% of admissions from the ED[2-5]. Syncope is often a challenging

symptom for ED physicians to deal with. It can be difficult to evaluate the exact cause of syncope as patients have typically recovered from the syncope event prior to their ED visit[6]. In addition, the event details and historical findings may not be reliably reported after the patient's episode of lost consciousness. Clinicians are also challenged by having to consider the wide variety of causes that can cause the syncope episode, and are often limited in their time to perform detailed assessments given the demanding nature of the ED[6]. As a result, confident diagnoses of the underlying cause of syncope may not be made in the ED[6]. The current diagnostic guidelines from the European Society of Cardiology recommend performing a comprehensive evaluation of a patient's clinical history, physical examination, and electrocardiography (ECG) to investigate the underlying cause of syncope[7]. These comprehensive initial evaluations lead to a diagnosis for the underlying cause of syncope in approximately 50% of cases, leaving the other half of syncope patients with an unexplained underlying cause[3-5,8,9].

In many cases, the underlying cause leading to the syncope episode is benign and does not warrant further hospitalization or follow-up. However, some patients have serious and life-threatening underlying conditions such as arrhythmias, myocardial infarction, serious structural heart disease, significant hemorrhage that were the cause of syncope episode; if undetected, these conditions can lead to severe morbidity and even mortality [10-17]. It is estimated that a significant proportion (7 to 23%) of adult ED syncope patients will experience serious outcomes within 7-30 days following ED visit[10-17]. When a diagnosis of the underlying cause of syncope cannot be easily and accurately made in the ED setting, risk stratification becomes necessary to appropriately assess the underlying risk and to safely and effectively manage the syncope patient[1]. The purpose of risk stratification is to identify the individuals at highest risk of

underlying/developing life-threatening events. These patients should be monitored further, admitted to the hospital, or appropriately referred to other specialists, since there are effective treatments available for these serious underlying conditions (e.g., pacemaker/defibrillator insertion for arrhythmias; and coronary angioplasty for patients with acute coronary syndrome)[18]. It is crucial that individuals who will benefit from these treatments are accurately identified in the ED, as this serves as a vital link to accessing these life-saving treatments and resources.

Risk stratification of ED syncope patients also aims to accurately identify patients who are at low risk of developing serious outcomes. These patients may be safely discharged after a brief ED evaluation. Due to the shortage of hospital beds and crowding of EDs in Canada, it is important to develop methods and guidelines to safely discharge these low risk patients in order to optimize the use of healthcare resources[18]. In addition, discharging low-risk patients can have implications in improving patient experience and satisfaction, as these patients will not have to experience unnecessary hospital stays or medical procedures.

1.2 Clinical Decision Tools

Clinical decision tools (also known as clinical prediction models, or clinical prediction tools) are statistical prediction models designed to help guide physicians with diagnostic and therapeutic decisions in the clinical setting. These evidence-based tools are designed to help clinicians with the uncertainty associated with medical decision-making[19]. Clinical decision tools incorporate elements from patient history, physical examination, or simple tests to estimate the risk probability of a certain outcome[19]. Clinicians, patients, and healthcare providers can subsequently use these risk estimates to inform clinical management and treatment decisions.

Clinical prediction tools have the potential to provide more accurate and less variable risk estimates compared to subjective clinical judgement[20-22].

Clinical decision tools should be developed through three distinct and sequential phases (derivation, validation, and implementation) before dissemination; however, in reality, most tools only undergo the derivation phase and never reach the subsequent validation or impact analysis phases[23,24]. The derivation phase refers to the creation of the tool following data collection. In this phase, investigators use a combination of clinical rationale and statistical techniques to identify the predictors most strongly related to the outcome of interest; these predictors are then included in the clinical decision tool[25]. The development of a useful and accurate clinical prediction tool depends on a variety of factors. Some factors relate to the inherent physiology of the disease in question, including the potential for accurate prognosis and the extent of prognostic information provided by the different variables; while other factors relate to the measurement process, such as the capacity to convert the predictor variables into reliable measurements and the capacity to represent these variables in a meaningful model[26].

Prognostic model validation involves testing the derived tool in a new population to assess its performance and reliability[20,23,26]. Many promising clinical decision tools perform well in the derivation phase, but exhibit a substantial reduction in performance during subsequent validation studies[23]. There are two primary reasons attributed to poor validation. First, poor validation may be the result of major differences between the derivation and validation populations[23]. Second, poor validation may also be attributed to deficiencies in the statistical techniques used in the development of the tool[23]. Often times, the derivation of prognostic models leads to the creation of inaccurate models that yield overoptimistic estimates of their true predictive performance risk estimates. In these instances, the derived model may be overfitted to

the derivation dataset, which results in poor generalizability and thus poor validation in other populations[26]. The concept of overfitting is of central importance to the present thesis and is further discussed in Section 1.4.1.2. It is recommended that all clinical decision tools undergo validation before being considered for adoption into clinical practice; however, many clinical decision tools never undergo validation studies[25,27].

Validation can be conducted using temporal validation or external validation on new subjects. Temporal validation refers to evaluating a prognostic model's performance on data from a new population of patients collected from the same sites as the derivation phase at a later time[26]. However, there have been criticisms that temporal validation may provide misleading results since the derivation and validation populations are inherently similar. Thus, a successful temporal validation does not confirm that the prognostic model possesses external validity that enables it to be generalized to other populations[26]. Many prognostic models that present a temporal validation are derived using an inadequately small sample size. It has thus been suggested that pooling the data from the derivation and validation cohort to derive the prognostic model, rather than splitting the data to perform a temporal validation, would be a more effective use of the data by allowing a larger sample size for a more robust derivation[20,26,28]. An external validation on new subjects involves evaluating the derived model on new data collected in a patient population different than the derivation phase. An external validation on new subjects is the most robust form of validation and evaluates the true external validity of the prognostic model; it is thus a critical step in demonstrating the validity of a model before considering its broad application in the clinical setting[26]. Temporal validation and external validation on new subjects are both considered forms of validation. Internal validation (discussed in Section 1.4.5)

is a statistical technique that involves validating the model on the same data used to develop the model, and is not considered a form of external validation.

Implementation, the third phase of clinical decision tool development, assesses the usefulness of the validated tool in the clinical setting[25]. Implementation studies can assess whether the clinical decision tool produces any change in clinician behaviour or a significant improvement in patient management or treatment. Implementation studies can also assess the potential economic benefit associated with the adoption of the clinical decision tool[25]. Implementation is followed by dissemination which involves the widespread adoption of the tool.

1.3 Methodological Standards for Development of Clinical Decision Tools

Methodological standards have been established for the development of Clinical Decision Tools, particularly in the field of Emergency Medicine[19]. These methodological guidelines serve to promote the derivation of robust and clinically-meaningful decision tools. They involve elements of the study design, data collection, and statistical analysis during the development phase of a clinical decision tool. With regards to the study design and data collection components, the outcome or diagnosis to be predicted should be clinically meaningful and well-defined[19]. The outcome measure should be blindly assessed, meaning without knowledge of the values for the predictor variables, to lessen the risk of bias. The predictor variables should be clearly defined, clinically-sensible, and collected in a standardized way[19]. The reliability of predictor variables when measured by different clinicians should be high, providing evidence of a high level of consistency and resulting in a more dependable clinical decision tool. Study subjects should be selected in such a way to be adequately representative of the condition of

interest, and the selection criteria of subjects should use properly defined and appropriate procedures to minimize the risk of selection bias[19]. The statistical techniques for deriving the tools should incorporate sound methods and a sufficiently large sample size is required to support the statistical methods used[19]. The accuracy and performance of the decision tool should be assessed using robust statistical methods. After the tool is derived, the clinical sensibility of the tool should be assessed by the potential end-users, most often clinicians. Clinical sensibility can be evaluated based on the content validity of the tool and potential for applicability in the intended clinical context[19].

1.4 Statistical Methodology for the Development of Prediction Models

Multivariable statistical modelling techniques are usually employed to develop clinical decision tools after the derivation data have been collected. There are many key statistical considerations in the development of a valid and useful clinical decision tool. Some of the major statistical considerations that must be considered include handling of missing data, modelling continuous variables, selecting the variables to be included in the final tool, assessing model performance, and investigating model robustness using internal validation techniques[28-30]. Currently, no consensus exists on the ideal method for developing a clinical prediction model[29].

1.4.1 General Concepts

There are a few general concepts that underlie many of the statistical considerations and methodologies used in prediction research. The main concepts include: the differences between prognostic research and aetiological research, the concept of statistical overfitting, and the trade-off between model complexity and model performance.

1.4.1.1 Differences between Prognostic Research and Aetiological Research

Aetiological research is commonly used in epidemiology and clinical research to investigate whether a specific variable of interest is an independent risk factor for the occurrence of a studied outcome after controlling for other confounders using a multivariable approach. In prognostic research, the goal is to use multiple variables to reliably predict the risk of future outcomes[31]. There are many similarities in the design and analysis employed in prognostic and aetiological research; however, the variables included in a prognostic model do not necessarily elucidate causal associations with the outcomes, rather they are mainly useful in predicting the occurrence of an outcome[31,32]. A model's calibration and discrimination (see definitions in section 1.4.5) are crucially important in prognostic research since its main goal is to accurately estimate risk; these statistical performance measures are less important in aetiological research[31].

1.4.1.2 Overfitting and Shrinkage

One of the most critical elements in clinical prediction modelling is to develop a well-performing model that can provide valid predictions in new subjects. A major threat to the validity of a clinical prediction tool is overfitting, which refers to producing a model that describes the derivation data well but does not generalize to subjects outside the sample[30]. The regression parameters and performance measures calculated from an overfitted derived model are usually overly-optimistic; over-stating the performance that the model will have when applied to a new population[30]. Data-dependent processes used to develop the model will invariably contribute to some level of overfitting in the prognostic model. In particular, the threat of overfitting will be increased when the predictor-outcome relationship is used to drive the model development process; this approach artificially creates an exaggerated association

between the predictor and outcome in the subsequent multivariable model, increasing the risk of type I error[30]. Overfitting is particularly problematic in derivation studies that consider too many candidate prognostic factors relative to the effective sample size[29,33].

In order to minimize the risk of overfitting, it is important to determine the allowable model complexity according to the amount of data available from the derivation dataset. Outcomes in clinical research are commonly dichotomized (e.g., mortality); in these cases, the number of events are defined by the number of observed outcomes — not by the number of individuals enrolled in the study[20]. This value is termed events per variable (EPV) [30,31]. It is suggested that prediction model development studies should include a minimum of 10 EPV to reduce the threat of overfitting; the larger the EPV the lower the threat of overfitting[30,31]. Some researchers argue that the number of candidate predictors corresponds to all variables considered for the prediction model and their corresponding terms for complexity (including non-linear functions, interactions), not just the variables included in the final prediction model[28]. Thus, variables considered formally or informally in bivariable screening (outlined in section 1.4.4), including subjective assessment of graphical plots, tests of linearity, or tests of interactions should be included in the list of candidate variables to calculate the EPV[28]. It is often difficult to account for the level of model uncertainty and overfitting attributed to these procedures[28].

Overfitting is a serious issue in the development of clinical prediction models. In a recent systematic review evaluating the methodology employed to develop prediction models in six large clinical journals, it was found that 66% of published models did not provide sufficient information to accurately evaluate the number of EPV and judge the potential for overfitting[27]. Among the studies that did contain adequate information to evaluate the EPV, it was determined that <10 EPV were used to develop clinical prediction models in 50% of the studies[27]. These

results suggest that many clinical decision models are developed with inadequate sample size and risk the potential for overfitting[27].

Shrinkage is a technique that adjusts the regression parameters and model performance measures in an attempt to account for overfitting[28,30]. Shrinkage deliberately reduces the estimated regression parameters and performance measures to be less optimistic about the model's performance when applied in a new population[28,30]. Shrinkage is quantified using a statistic called "optimism". The optimism is defined as the difference between a model's apparent and true performance, where true performance refers to the underlying population and apparent performance refers to the estimated performance in the study sample[30]. The optimism can be calculated through the application of a heuristic shrinkage factor or a bootstrap resampling method (outlined in section 1.4.5)[28,30].

1.4.1.3 Trade-off between Model Complexity and Model Performance

Clinical applications often emphasize the necessity for practicality and clinical sensibility, leading to the need for a simple and interpretable clinical decision tool[29]. The need for a parsimonious decision tool may be even more important in the ED, as clinicians often have little time to evaluate patients given its demanding environment. However, there is typically a trade-off between model complexity and model performance. In general, models that include more variables and use complex non-linear functions of continuous predictors have better performance and produce more reliable predictions[28,30]. Some researchers argue that it is inappropriate to reduce the complexity of a model by categorizing continuous predictors or removing predictors from the model, since these procedures may have implications in diminishing the model's predictive performance[28]. However, the main goal of developing a

prediction model is to develop a valid model that can be transposable and adopted in clinical practice; thus, it is often necessary to compromise on a parsimonious model that does not sacrifice substantial predictive performance.

1.4.2 Missing Data

Missing data is an unavoidable problem in epidemiological research and needs to be properly addressed. Treating missing data appropriately reduces the threat of bias and improves the validity of the research results[27,28,30]. The extent of missing data is an important consideration in determining the most suitable methodology for handling the missing data[28,30]. There are three primary mechanisms that are used to describe missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR)[27,28,30].

MCAR signifies that missing data elements are completely random and unrelated to any characteristics of the predictor variables or the responses for the subjects[28,30]. In practice, missing data are usually related to outcomes or study characteristics of a patient, making MCAR an unrealistic assumption[29]. MCAR can be ruled out using the observed data by comparing characteristics of subjects with and without missing data[27,28,30]. If important differences exist, MCAR is ruled out; however, if no important differences are found, the MCAR assumption may be possible, although there may still be some unmeasured characteristics for which those with missing data are systematically different than those with complete data[27,28,30].

Under the MAR assumption, the probability that values on any particular variable are missing depends on the observed data (e.g., other variables that were captured in the dataset), but not on any unobserved data. Finally, under MNAR, the probability that values are missing

depend on the unobserved values themselves.[28,30]. It is not possible to distinguish between MAR and MNAR through testing on the observed data[28,30].

Characterizing the mechanisms leading to missing data is critical in deciding the statistical approach used to address the missing data, since each approach relies on assumptions of the missing data mechanism[28,30]. There are three commonly-used procedures employed to deal with missing data: casewise deletion, simple imputation, and multiple imputation.

Case-wise deletion is a procedure that excludes any subject with a missing value for any of the predictor variables or outcome variables in the regression analysis[28,30]. Case-wise deletion is the default procedure used in most statistical programs. Case-wise deletion, though easy to implement, uses data inefficiently as it deletes all subjects with missing values on any relevant variable from the analysis. As a result of this reduction in sample size, the standard errors increase and the power of tests of association decreases[28,30]. Furthermore, when case-wise deletion is applied under MAR or MNAR, the estimated regression coefficients can be biased [28]. Applying case-wise deletion under MCAR will yield unbiased coefficient estimates; however, the issues with increased variability associated with this approach still persist. Thus, case-wise deletion generally addresses missing data sub-optimally.

Simple imputation involves replacing the missing values in the dataset with other plausible values[28,30]. Single imputation with the conditional mean involves replacing missing values with the observed mean or median for other subjects[30]. This approach ignores the information that can be generated from the correlation between predictors, and thus uses the available data sub-optimally. Single regression imputation is another simple imputation technique that addresses this limitation by considering the correlation in the data using a multivariable statistical model[30]. In comparison to case-wise deletion and single imputation with

the conditional mean, single regression imputation provides an improvement under MAR since it takes into account the correlation with other relevant variables to explicitly model the MAR process[30]. Simple imputation uses all the available data that was collected, and thus uses data more efficiently than case-wise deletion. However, all simple imputation methods have a major limitation: they artificially underestimate standard errors since they inherently treat the imputed data as if they were observed data[28,30]. When there are few missing observations in the data (e.g., <5%), applying simple imputation may be adequate since the change in standard error measurements would be negligible[27].

In multiple imputation, multiple plausible imputed data sets are created to reflect the variability and uncertainty associated with predicting missing values[34]. Similar to single regression imputation, information from other correlated variables in the dataset are used to inform the imputation through the specification of a multivariable imputation model. The imputation model aims to approximate the distributional relationships among the variables in the dataset to inform the imputation[28,30]. In the context of a clinical prediction model, the imputation model should include the predictors considered for the prediction model and the outcome variable. In addition, auxiliary variables, which are variables that are not considered for the prediction model but may contribute information about missing values, should be included in the imputation model[30]. It has been suggested that including many variables in the imputation model typically improves the quality of the imputation process[30].

In most cases, multiple variables specified in the imputation model have missing values. Multiple imputation employs data augmentation methods to address this issue and impute the missing data[28,30]. Data augmentation is an iterative process which imputes values for the missing data, then recalculates new estimates for the model parameters based on the recently imputed

values until convergence. Under multiple imputation, the imputation procedure is repeated multiple times to generate *k* completed datasets[28,30]. Rubin and Little have shown that *k* = 3-10 multiple imputations are usually adequate[35]. Each of the *k* datasets is then analyzed using the same statistical method. A pooled estimate for each parameter of interest is then calculated as the mean of the *k* estimates, while the standard error is computed as the square root of the sum of the mean within-imputation variance and a measure of between-imputation variance[34]. In this way, the variation between the *k* estimates are incorporated in the pooled statistical inferences to account for the uncertainty attributed to missing data[28,30].

Multiple imputation is effective in imputing missing data that are MAR, since it explicitly models the MAR mechanism through the use of the imputation model. The MAR assumption is less strict and typically more realistic than the MCAR assumption in clinical and epidemiological research[28,30]. The multiple imputation procedure has also been demonstrated to provide unbiased regression coefficients and robust standard error estimates in data that are MCAR[30]. However, performing multiple imputation on missing data with a MNAR assumption may produce less robust imputations and results[34]. Unfortunately it is not possible to determine whether data are MAR or MNAR using the observed data and there is no simple statistical methodology developed to address missing data under the MNAR, so multiple imputation is likely still the best approach under MNAR [34]. Taken together, multiple imputation is the recommended approach for dealing with missing values on predictor variables, as it helps to preserve power and reduce the risk of bias[27].

In a recent systematic review evaluating the statistical methodology of clinical prediction tools published in 2008 in six high impact general medical journals, it was indicated that missing data was addressed using complete-case analysis in 71% of studies where the primary aim was to

develop a prediction model[27]. In addition, 50% of the studies failed to report or were unclear about the methodology used to handle missing data; many of these studies likely employed complete-case analysis as well[27]. Only 10% of the studies used multiple imputation — the most rigorous strategy for dealing with missing values[27]. Therefore, the methodology used to address missing values has been sub-optimal in the development of clinical prediction models.

1.4.3 Treatment of Continuous Variables

1.4.3.1 Categorization

Continuous variables are often categorized in the derivation of clinical prediction models. It is believed that the categorization of continuous variables facilitates the ease of interpretation and clinical usefulness of the derived prediction model[36]. Clinical decision-making often requires two discrete classes (e.g., normal/ abnormal, treat/ do not treat), so categorization in the form of dichotomization likely reflects this process. Many successful clinical prediction tools have been successfully derived and externally validated using this approach[10,37-40].

Categorization of a continuous predictor has been demonstrated to have substantial disadvantages, in particular, providing a poor representation of the underlying risk association [30,36,41]. Categorization imposes a marked change in risk at a specific threshold; this assumption implies that individuals close to but on opposite sides of the threshold are characterized as having very different risks, which is typically an unrealistic assumption[30,36]. In addition, categorization assumes a constant risk below or above the specified threshold, which may be unrealistic as predictor-response relationships in continuous variables often demonstrate gradual dose-dependent response relationships[30]. By ignoring this dose-dependent predictor-response relationship that often exists in continuous predictors, the use of categorization can reduce the

power to detect real predictor-response relationships[36]. It has been demonstrated that dichotomizing continuous predictors at the median leads to a loss of efficiency by 35 percent — a serious loss of data and statistical power[36,42]. As a result, previous studies have found that the performance of multivariable regression models becomes worse with categorization[36,41].

Despite its disadvantages, categorization is still frequently used in practice. This requires choosing relevant cut-points prior to analysis. The approaches used to choose a cut-point are often arbitrary and heterogeneous across different studies[36]. The cut-point is often defined using a data-driven approach that involves the predictor-outcome relationship. The approach involves trying possible cut-points and choosing a value that maximizes sensitivity or minimizes the p-value in the bivariable relationship[36]. In essence, the process artificially exaggerates the association between the predictor and outcome, leading to an increase risk of Type I error[36,41]. The problem can be further exacerbated when the categorized continuous variable is included in a multivariable analysis: the categorized predictor can artificially attenuate the predictor-outcome relationships of other true variables included in the multivariable model, leading to unreliable and spurious results[36,41].

1.4.3.2 Parametric modelling

Continuous predictors can be modelled as linear or non-linear relationships in prediction models. When the predictor is modelled using a linear relationship, its effect is assumed to be constant across its entire range [30]. Imposing a linear relationship leads to a simple analysis and interpretation and thus, is a frequent choice in practice. However, it may be unwise to arbitrarily assume linearity since the misspecification of a linear relationship can lead to inaccurate predictions in new patients[29].

Predictors can be modelled more flexibly using non-linear functions. The most common approaches to non-linear modelling of continuous predictors are polynomials, fractional polynomials, and spline functions[30]. Polynomials involve adding a polynomial term as an extension to a model with a linear term, such as square, cubic, square root, or logarithmic terms [30]. Polynomial functions are easy to implement; however they are restricted in the shapes they can assume[30]. Fractional polynomials are an extension of ordinary polynomials that include non-positive and fractional powers from the set of -2, -1, -0.5, 0, 0.5, 1, 2, 3[30]. Fractional polynomials can be modelled by combining multiple polynomial terms to describe the non-linear relationships. Although fractional polynomials are useful in flexibly modelling continuous variables, they are susceptible to distortion caused by values at the tails of the distribution[30].

Restricted cubic splines functions are powerful tools that can flexibly characterize complex dose-response association between a continuous exposure and outcome[43,44]. A spline regression is defined by piecewise polynomial functions that model adjacent intervals of continuous data[44]. The junction between each interval is called a "knot", and typically a small number of knots (3 to 8) are adequate[44]. Each interval is modelled using cubic spline functions (polynomial function with a degree of 3), that maintain continuity and smoothness at each knot[44]. The locations of the knots do not necessarily have to be pre-specified — the marginal distribution of the continuous variable can be used to define the knot location without increasing the risk of bias[28]. Restricted cubic splines are also constrained to linearity before the first knot and after the last knot, which prevents abnormal deviations that may arise in open-ended polynomial functions[44].

One additional advantage of restricted cubic splines is their ability to be modelled parsimoniously with few required coefficient estimates. Restricted cubic spline regressions with

K knots require K-1 coefficient estimates, which is generally less than the number required to model polynomial and fractional polynomial functions[44]. Therefore, restricted cubic splines can model continuous data flexibly without substantially increasing the risk of statistical overfitting[28]. Some disadvantages of restricted cubic splines are related to the flexibility of spline regressions, including the occurrence of irregularities in the fitted curves that may be unrealistic and open to interpretation, and the inability to describe the spline functions in simple terms[29,45,46]. Nevertheless, spline regressions often provide a less biased alternative to standard linear or curvilinear analytical techniques used in modelling continuous variables, and are thus the recommended approach to handling continuous variables in clinical prediction modelling[43,44].

Current evidence suggests that continuous variables are handled sub-optimally in the derivation of clinical prediction models[27]. When evaluating the methodology employed in the development of clinical prediction models, continuous variables were categorized in 47% of studies[27]. Only 19% of studies used non-linear transformations to model the continuous predictor variable[27]. Thus, the development of clinical prediction tools often employs sub-optimal methodology in the treatment and modelling of continuous variables.

1.4.4 Model Specification

Selecting variables to be included in the model is often considered the most difficult step in the development of a prediction model, and there is no consensus on the best approach[29]. A combination of statistical methods and subject-matter expertise can be used to identify the predictors included in the model.

Bivariable screening is an approach that uses the predictor-outcome relationship to select variables for inclusion in the multivariable model. In this approach, variables with a significant

statistical association, defined by a p-value below an arbitrary threshold (usually 0.05), are included as candidates in the multivariable model[47]. Bivariable screening can be a quick and easy tool to reduce the number of predictors included in the multivariable model[47]. However, bivariable screening may be an inappropriate practice because it cannot properly control for confounding as its approach inherently assesses one variable while ignoring all others[47]. When confounding is inadequately controlled, the effects of a covariate on the outcome may be biased or distorted. Thus, the use of bivariable screening to select predictors into a multivariable model may lead to the rejection or inclusion of inappropriate variables, leading to an inaccurate prediction model[47]. In addition, bivariable screening increases the risk of biased results and overfitting since it explicitly uses the predictor-outcome relationship to guide the model-fitting procedure[27].

Clinical prediction tools often consider many candidate variables; to derive more parsimonious models for use in clinical settings, stepwise variable selection techniques are often employed [48]. Stepwise variable selection involves a sequential model building process that adds or drops one predictor at a time depending on a pre-specified criterion, the most common criterion being the p-value of a predictor[48]. There are three main stepwise variable selection algorithms that are widely used in practice: backward elimination, forward selection, and stepwise selection[48]. In backward elimination for example, using a p-value of 0.05 as the criterion, the model begins with all the potential predictors and the predictor with the highest p-value is removed from the model. The model is then re-estimated and the process continues until all remaining predictors in the model have a p-value <0.05 [48]. In contrast, forward selection starts with no predictors in the model and the potential predictor with the lowest p-value is added and remains in the model. This process continues until no additional predictors meet the entrance

criteria, for example, having a p-value <0.05. The stepwise selection procedure uses the same procedure as forward selection except a predictor in the model can be dropped if its p-value exceeds a certain significance threshold (e.g., p-value >0.05) after the model is re-estimated with the addition of a predictor[48].

Some studies have examined the stability of stepwise variable selection techniques when applied to small datasets[30,49]. Stepwise variable selection procedures have been demonstrated to lead to inconsistent results[30,49]. For example, researchers used bootstrap samples (samples created by random draws with replacement from the original dataset) to demonstrate that different predictors can be selected in each bootstrap sample[30,49]. Thus, prediction models derived from stepwise variable selection procedures may not be reproducible and may be sensitive to random fluctuations in the data[49]. Sample size is one crucial factor that determines the stability of the variable selection procedure; larger samples are expected to have more stability and increased power to identify true predictors[30].

Stepwise variable selection can also lead to biased estimates of regression coefficients and unreliable standard error estimates[30,50]. When predictors are removed from the full model, the estimated effect associated with that predictor can be redistributed to other predictors that remain in the model. As a result, the distributions of the coefficients for the remaining variables are exaggerated and biased away from the null[30,50]. This phenomenon also creates an erroneous estimate of variability, leading to underestimated standard errors and exaggerated p-values for the regression coefficients[30]. Biased parameter estimates and unreliable standard error estimates can be an especially great concern when the stepwise variable selection is applied to a small dataset with a stringent selection criterion; leading to an increased risk of Type I error and the identification of false predictor-outcome associations[30,50].

An alternative approach to the stepwise variable selection approach is to pre-specify the variables to be included in the full model using solely subject-matter expertise. The complexity of the model is restricted by the amount of available data used to derive the decision tool; smaller samples enable fewer predictors to be included in the clinical prediction model[28]. This approach is claimed to reduce the risk of overfitting and provides correct standard errors and p-values for the parameter estimates[28]. However, a major drawback to this approach is that pre-specifying the full prediction model may be difficult in advance of the study[29]. In well-developed areas of research, such as cardiovascular health, identifying risk factors to be included in a clinical decision tool in advance may be possible; however, this is unlikely the case in more underdeveloped areas of research[51].

A review of published clinical prediction rules showed that bivariable screening is used to reduce the number of candidate predictors in 15% of clinical prediction model studies[27]. Once variables are selected for a multivariable model, stepwise variable selection was performed in 22% of studies to derive a more parsimonious model, while 42% of studies retained all the predictors in the model regardless of statistical significance[27]. The method used to select predictors in the multivariable analysis was not described in 29% of studies[27]. These results highlight the heterogeneous approach used to select variables to be included in a prediction model. Although bivariable screening and stepwise variable selection are often ill-advised based on the aforementioned statistical considerations, it may be very difficult to meaningfully evaluate the most important candidate predictors to be included in a model a priori based on subject-matter expertise, especially in underdeveloped areas of research. Thus, bivariable screening and stepwise variable selection may be useful in pruning the number of candidate predictors in some situations despite their documented statistical limitations.

1.4.5 Model Performance Measures and Internal validation

1.4.5.1 Model Performance Measures

Model performance measures are critical in evaluating a clinical prediction model. Overall performance measures, such as the Nagelkerke's R^2 and the Brier Score measure the difference between the predicted outcome (as measured by the model) and the observed outcome. Models with smaller distances between the predicted and observed outcome have better performance[30].

Model discrimination and calibration are key performance indicators that dictate the accuracy and quality of prediction model risk estimates. Model discrimination refers to a model's ability to distinguish between participants with or without the outcome. The most commonly used measure of discrimination is the concordance statistic (c-statistic), which is equal to the area under the receiver operating characteristic (ROC) curve for binary outcomes. ROC curves plot the sensitivity against the false positive rate throughout varying discrimination thresholds[30]. Model calibration refers to the agreement between the predicted and observed risk of outcome; some tools and measures to assess calibration include calibration plots, which graphically compare the observed probabilities with the model-derived expected probabilities, and the Hosmer-Lemeshow goodness of fit statistic[27].

Current guidelines recommend that these performance measures should be measured and reported in the development and validation of clinical prediction models[29,31]. However, a review of published clinical prediction models showed that only 13% of the 15 studies examined listed overall performance measures using the R^2 statistic, and only 27% of these studies reported elements of calibration using either a calibration plot or a Hosmer-Lemeshow goodness of fit

statistic[27]. In this review, model discrimination was fairly widely assessed and reported using the c-statistic in 80% of the 15 studies examined[27].

1.4.5.2 Internal Validation

Model validation aims to determine whether the predictions generated from the derived model are likely to be valid and reproducible when applied to a new population not used to develop the model[28,30]. The internal validation procedure works by generating multiple validation datasets to test the prediction model; each validation dataset is a variation of the original dataset, deliberately introducing an element of variability[30]. Internal validation methods are helpful in identifying issues in the model-fitting process and present more realistic performance measures than those reported by estimating the model using solely the original derivation data[52]. However, it is important to note that the full validity of a model cannot be assessed using only internal validation procedures since the model is being tested in the same population used for derivation. Therefore, internal validation procedures typically present overly-optimistic assessments of the model's true performance[28]. Nonetheless, internal validation is still a crucial component in identifying statistical issues in the model fitting process and assessing the preliminary validity of a clinical prediction model. Some methods of internal validation include data-splitting, cross-validation, jackknifing, and bootstrapping[30].

The bootstrap method is the recommended approach for internally validating prediction models[28]. In bootstrap internal validation, bootstrap samples are created by randomly drawing subjects with replacement from the original sample[30]. Some subjects may not be included in the bootstrap sample while others may be included multiple times, which deliberately introduces an element of randomness and variability in the process. The bootstrap samples are the same size as

the original sample, allowing the precision of estimates in each bootstrap sample to be consistent with the original sample[30].

Bootstrap resampling is a useful non-parametric technique that can be used to provide an estimate of the empirical distribution of summary measures from a sample population[30]. In particular, the bootstrap can be used to calculate the standard error and thus confidence intervals for any measure of a statistical model, such as a performance measure (e.g., the c-statistic in logistic regression) or regression coefficients[30].

One particularly useful application of the bootstrap method is in estimating the degree of optimism attributed to overfitting in the prediction model[30]. In this approach, the prognostic model is re-derived in each bootstrap sample and applied to the original sample. The difference in accuracy index between the bootstrap sample and the original sample is an estimate of the optimism. The optimism-corrected fit statistics are then calculated by subtracting the optimism value from the model's apparent accuracy index[30]. Bootstrapping can be used to estimate the optimism on any model fit index, and can also be used to introduce shrinkage in the regression coefficients to account for overfitting[30] (see section 1.4.1.2 for a discussion of overfitting and shrinkage).

It is recommended that internal validation procedures be used in the derivation of clinical prediction tools to assess preliminary robustness and validity. Despite this recommendation, only 33% of the 15 studies reviewed by Bouwemeester *et al.* reported using this procedure in the development of clinical prediction models [27,28].

1.5 Improving Methodology in Prediction Modelling Research

Despite their potential for improving and standardizing patient care, many clinical decision tools are never validated or ultimately adopted into clinical practice[27,29]. Some of the lack of success in the adoption of clinical prediction tools has been attributed to the use of sub-optimal methodology and lack of transparent reporting used in prediction research[33]. The Prognosis research strategy (PROGRESS) Partnership is an international collaboration largely devoted to developing methods, recommendations, and guidelines for improving clinical prediction model reporting and methodology[29]. The PROGRESS Partnership strongly emphasizes many elements of statistical methodology, such as the need to have an adequate sample size and the collection of high-quality data when deriving a clinical prediction model. With respect to the statistical methodology, the PROGRESS Partnership addresses many of the aforementioned statistical considerations (section 1.4.2 to 1.4.5); such as dealing with missing data, modelling continuous prognostic factors appropriately, considering allowable model complexity, and verifying model assumptions[29]. In particular, the PROGRESS Partnership advocates the need for a critical evaluation of missing data, clearly quantifying the degree of missing data and the potential influence the missing data will have on the overall results[29]. The group also discourages the use of categorization in favour of using non-linear transformations to model continuous predictors[29]. The PROGRESS Partnership is also a strong proponent for the thorough evaluation and reporting of a prognostic model's performance with respect to its statistical discrimination and calibration[29].

The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) initiative recently developed a set of standardized recommendations, called the TRIPOD statement, for prediction research[53]. The TRIPOD initiative advocates for full

and clear reporting so that readers can objectively judge the strengths and weaknesses of the prediction study, adequately assess the risk of bias, and determine the potential usefulness of the decision tool[53]. The TRIPOD statement explicitly identifies a checklist of 22 items that are deemed essential in the reporting of clinical prediction model studies[53]. This checklist is intended to be a guideline to support authors in describing their prediction model studies and provides a framework for readers to critically assess these studies.

The TRIPOD statement contains elements of statistical methodology that should be addressed when reporting clinical prediction model studies; many of these elements pertain to the statistical considerations used in developing prediction models outlined above (section 1.4.2 to 1.4.5). The TRIPOD statement advocates the transparent reporting with regards to the justification of the sample size, the approach used to handle missing data, the handling/coding of analysis variables, the model-building process used to identify the predictors included in the final model, the method of internal validation, and measures of model performance[53].

1.6 Existing Clinical Decision Tools for Syncope

Currently, there exists nine (9) original published studies that aimed to develop risk-stratification tools for ED syncope patients[8,10,11,16,17,54-58]. These studies include San Francisco Syncope Rule (SFSR), Boston Syncope Criteria, Osservatorio Epidemiologico sulla sincope nel Lazio (OESIL), Short-Term Prognosis of Syncope (StePS), Evaluation of Guidelines in Syncope Study (EGSYS), Syncope Risk Score, Risk Stratification of Syncope (ROSE) Rule, a study conducted by Martin *et al*., and a study conducted by Sarasin *et al.* [8,10,11,16,17,54-58]. There is much heterogeneity among these studies, as they differ in timeframe of outcome assessment, definition of outcomes, and specific study populations. A description of the studies is presented in **Table 1**.

We performed a detailed evaluation of these prediction tools, with an emphasis on evaluating the conduct and reporting of the methodology employed. We followed the approach taken in other studies to evaluate the methodology and reporting of subject-specific clinical prediction models[59-61]. A description of the methodology employed in the development of each syncope clinical prediction tool is presented in **Table 2**, and a summary of the methodology is provided in **Table 3**.

Table 1. Summary of pre-existing clinical prediction tools for syncope

Study	Year of study	Country	Outcome	Risk predictors in final model	Model Performance (95% CI)
SFSR[10]	2000	United States	Death, myocardial infarction, arrhythmia, pulmonary embolism, stroke, subarachnoid hemorrhage, significant hemorrhage, or any condition causing a return ED visit and hospitalization for a related event following 7 days of their ED visit	1.Abnormal ECG 2. Complaint of shortness of breath 3. Hematocrit <30% 4. Systolic BP < 90 mmHg 5. Congestive heart failure	Sensitivity 96% (92% to 100%); Specificity 62% (58% to 66%)
OESIL[55]	1997	Italy	Total Mortality following 1 year after ED visit	1.Age >65 years 2. History of cardiovascular disease 3. Lack of prodrome 4. Abnormal ECG	c-statistic: 0.897
STePS[56] (short-term)	2004	Italy	Death and need for major therapeutic procedures within 10 days after ED visit. Major therapeutic procedures include: cardiopulmonary resuscitation, pacemaker or implantable cardioverter-defibrillator insertion, intensive care unit admittance, acute antiarrhythmic therapy	1.Age>65 2. Male sex 3. Structural heart disease 4. Heart failure 5. Trauma 6. No symptoms of impending syncope 7. Abnormal ECG	None reported
STePS[56] (long-term)	2004	Italy	Death and need for major therapeutic procedures within 11 days to 1 year of ED visit. Major therapeutic procedures are the same as those listed for STePS (short-term)	1.Age >65 years 2. History of neoplasm 3. Cerebrovascular disease 4. Structural heart disease 5. ventricular arrhythmia	None reported
ROSE[16]	2007	United Kingdom	Death and serious outcome at 1 month after ED visit. Serious outcomes include: acute myocardial infarction; life-threatening arrhythmia, ventricular tachycardia, ventricular pause, ventricular standstill; pacemaker or cardiac defibrillator;	1.Brain natriuretic peptide >300 pg/mL 2. Stool positive for occult blood 3. Oxygen saturation <94% 4. Hemoglobin <90 g/L	c-statistic: 0.83; Sensitivity 92.5%; Specificity 73.8%

Study	Year of study	Country	Outcome	Risk predictors in final model	Model Performance (95% CI)
			pulmonary embolus, cerebrovascular accidence, hemorrhage, acute surgery	5. Chest pain at time of syncope 6. Bradycardia	
Sarasin *et al.*[8]	1998	Switzerland	Arrhythmia following ED visit	1.Abnormal ECG 2. Congestive heart failure 3. Age >65	c-statistic*: 0.84 (0.77 to 0.91)
EGSYS[57]	2004	Italy	Diagnosis of Cardiac syncope based on European Society of Cardiology guidelines or mortality within 24 months following ED visit	1.Abnormal ECG and/or heart disease 2. Palpitations before syncope 3. syncope during effort 4. Syncope while supine 5. Autonomic prodrome 6. Predisposition to vasovagal syncope	Sensitivity: 95%; Specificity: 61%
Martin *et al.*[54]	1981	United States	Mortality or arrhythmias within 1 year following ED visit	1.abnormal ECG 2. History of ventricular arrhythmia 3.History of congestive heart failure 4. Age greater than 45 years	c-statistic: 0.77 to 0.83
Boston Syncope Criteria[17]	2003	United States	Critical intervention or adverse outcome within 30 days after ED visit. Critical interventions include: implantable cardiac defibrillator placement, percutaneous coronary intervention, cardiopulmonary resuscitation, or correction of carotid stenosis. Adverse outcomes included death, pulmonary embolus, stroke, sepsis, ventricular dysrhythmia, atrial dysrhythmia, hemorrhage, myocardial infarction, or cardiac arrest.	1.Signs of acute coronary syndrome 2. Signs of conduction disease 3. Worrisome cardiac history 4. Valvular heart disease 5. Family history of sudden death 6. Persistent abnormal vital signs in ED 7. Volume depletion 8. Primary central nervous system event	Sensitivity: 97% (93% to 100%); Specificity: 62% (56% to 69%)
Syncope	2002	United	Serious clinical event occurring within 30 days	1. Age greater than 90 years	c-statistic: 0.61

Study	Year of study	Country	Outcome	Risk predictors in final model	Model Performance (95% CI)
Risk Score		States	following ED visit. Serious events include death, arrhythmias, myocardial infarction, diagnosis of structural heart disease, pulmonary embolism, aortic dissection, stroke/transient ischemic attack, subarachnoid hemorrhage, hemorrhage, or anemia requiring blood transfusion	2. Male sex 3. History of arrhythmia 4. Triage systolic blood pressure >160 5. Abnormal ECG 6. Abnormal Troponin I 7. Near-syncope	(0.57 to 0.65)

*Performance measure adjusted for optimism

Table 2. Methodology used to derive pre-existing clinical prediction tools in syncope

	SFSR	OESIL	STePS	ROSE	Sarasin *et al.*	EGSYS	Martin *et al.*	Boston Syncope Criteria	Syncope Risk Score
Outcome clearly defined	✓	✓	✓	✓	✓	✓	✓	✓	✓
Predictor clearly defined	✓	✓		✓	✓	✓	✓	✓	✓
Study participants clearly described	✓	✓	✓	✓	✓	✓	✓	✓	✓
Inter-observer reliability of predictors assessed	✓								✓
Statistical analyses described	✓	✓	✓	✓	✓	✓	✓		✓
Sample Size									
Adequate information to assess Events per Variable	✓	✓	✓	✓		✓			
Events per Variable	1.58	3.44	3.15	0.58		0.44			
Events Variables Considered for Prediction Model	79 50	31 9	41 13	40 69		23 52			
Overfitting mentioned or discussed					✓				
Missing Data									
Single Conditional-mean imputation				✓					✓
Not Reported	✓	✓	✓		✓	✓	✓		
Modelling Continuous Predictors									
Categorization	✓	✓	✓	✓	✓	✓	✓		✓
Model Specification									
Statistical Model Used Recursive Partitioning Cox Regression Logistic Regression	 ✓	 ✓	 ✓	 ✓	 ✓	 ✓	 ✓ ✓	 	 ✓
Variables Pre-specified								✓	

	SFSR	OESIL	STePS	ROSE	Sarasin *et al.*	EGSYS	Martin *et al.*	Boston Syncope Criteria	Syncope Risk Score
Bivariable Screening	✓	✓	✓	✓	✓	✓	✓		
Stepwise Variable Selection		✓	✓			✓			✓
Model Performance									
c-statistic		✓		✓		✓	✓		✓
Optimism-corrected c-statistic					✓				
Calibration Assessed									✓
Sensitivity & specificity	✓			✓		✓		✓	
Internal Validation					✓				✓
External Validation									
Temporal Validation		✓		✓		✓	✓		
External Validation on new patients	✓				✓				

Table 3. Summary of methodology used to derive pre-existing clinical prediction tools in Syncope

Methodological Characteristic	**% (n)**
Outcome Clearly Defined	100 (9)
Predictor Clearly Defined	89 (8)
Study Participants clearly described	100 (9)
Inter-observer reliability of predictors assessed	22 (2)
Statistical analyses described*	100 (8)
Sample Size	
Adequate information to assess number of events per variable	63 (5)
<10 events per variable**	100 (5)
Overfitting mentioned or discussed	13 (1)
Missing Data	
Single Conditional-mean imputation	25 (2)
Not Reported	75 (6)
Modelling Continuous Predictors	
Categorization	100 (8)
Model Specification	
Statistical Model***	
Logistic Regression	63 (5)
Recursive Partitioning	25 (2)
Cox Regression	25 (2)
Variables Pres-specified	0 (0)
Bivariable Screening	88 (7)
Stepwise Variable Selection	50 (4)
Model Performance	
c-statistic	75 (6)
Optimism-corrected c-statistic	13 (1)
Calibration Assessed	13 (1)
Sensitivity & specificity	50 (4)
Internal Validation	25 (2)
External Validation	
Temporal Validation	50 (4)
External Validation on new patients	25 (2)

*Boston Syncope Criteria excluded from statistical description since no formal statistics employed, leaving 8 studies in denominator

**Only 5 studies contained sufficient information to evaluate number of events per variable

***Some studies used a combination of statistical models to develop the prediction model

1.6.1 Overview of Studies

The nine studies sought to derive clinical prediction tools to evaluate the risk of serious outcomes among ED syncope patients. Five studies evaluated the risk of short-term outcomes, which were outcomes that occurred in the subsequent 30 days following initial presentation in the ED[8,10,16,17,58]. Three studies evaluated the long-term outcomes, which were outcomes that occurred within one year or longer after initial ED visit[54,55,57]. One study evaluated both a short-term and long-term outcome[56].

The Boston Syncope Criteria was an anomalous study as it did not truly develop a clinical decision tool using standard methodology and statistical techniques; rather, it was a consensus-based algorithm developed using recommendations from societal guidelines and clinical literature[17]. The Boston Syncope Criteria was thus included in the assessment of study design (section 1.6.2) but was excluded from the subsequent statistical methodological evaluation, leaving 8 studies for methodological evaluation.

1.6.2 Sample Size and Overfitting

The number of events per variable (EPV), defined as the number of serious outcomes observed in the study divided by the number of candidate predictors analyzed, could not be calculated in 38% (n=3) of the syncope prediction model studies. In another 3 studies, the information was not explicit and certain assumptions needed to be made in order to estimate the EPV. Nevertheless, from the 5 of 8 studies where the EPV could be calculated, the values ranged from 0.4 to 3.4; median of 1.6. Thus, all 5 studies examined were markedly below the recommended guideline of a minimal of 10 EPV, increasing the risk of statistical overfitting[30,31].

From the eight syncope prediction model studies analyzed, only one article explicitly addressed the issue of overfitting in the model development process[8].

1.6.3 Missing data

Descriptions of missing data or the methods used to address missing data were not reported in 75% (n=6) of the studies. Complete-case analysis was likely used in these studies, potentially leading to biased results and inaccurate estimates[27].

Two studies made explicit mention of missing data and the methods that were used to address these issues. In the ROSE study, the degree of missingness in the predictor variables was not outlined, but the methods used to treat missing data were reported. Missing values for categorical variables in the ROSE study were designated as a separate category, which has been documented to lead to biased results[27]. Therefore, we believe that missing data was inadequately handled in the ROSE study since there was no mention of the degree of missingness, and the methodology employed to address missing data was sub-optimal. In the Syncope Risk Scale study, the researchers explicitly reported a low degree of missingness and described the imputation techniques used to address the missing data[58]. We thus believe that missing data was appropriately handled in the Syncope Risk Scale study.

Overall, missing data was handled sub-optimally in 88% (n=7) of the syncope prediction model studies. Most studies failed to report the degree of missingness in the predictor variables, neglected to describe how missing data was handled, or used inappropriate techniques to address the missing data.

1.6.4 Modelling Continuous Variables and Model Specification

All 8 studies were developed by categorizing continuous predictors in the prediction models. Treating continuous variables as categorical is ill-advised, even though it is a common practice in the derivation of prediction models[27,29,36].

Logistic regression was the most common statistical model used 63% (n=5); recursive partitioning was used in 25% (n=2) and Cox regression was used in 25% (n=2) of studies. Stepwise variable selection is compatible for use in logistic and Cox regression, but cannot be employed in recursive partitioning. From the six studies that used either logistic or Cox regression, 67% (n=4) studies performed a stepwise variable selection procedure. From the studies where the number of candidate variables considered could be ascertained (n=5), the median number of candidate predictors considered was 50. Bivariable screening was used to reduce the often large number of predictors considered for the multivariable model in 88% (n=7) of syncope prediction tools.

Overall, the selection of predictors to be included in the final syncope prediction models was largely data-driven. It appears that in most of the studies, many candidate predictors were considered so the data-driven methodology likely reflects the necessity to reduce the number of predictors. It may be difficult to pre-specify a parsimonious list of predictors given the heterogeneity of serious outcomes that occur in syncope patients and the fact that syncope is a widely under-studied condition. Although we acknowledge that the Boston Syncope Criteria was not a formal decision tool developed using accepted methodological standards, it was the only study that attempted to use clinical literature and knowledge to pre-specify a risk-assessment tool[17]. However, the final algorithm presented as the Boston Syncope Criteria included 8

variables consisting of 28 different fields; the large list of predictors highlights some of the difficulties in pre-specifying a useful and parsimonious tool to risk-stratify syncope patients[17].

1.6.5 Model Performance

The c-statistic was commonly used to report discrimination in 75% (n=6) of studies. Only one study addressed elements of model calibration by reporting the Hosmer-Lemeshow goodness of fit statistic[58]. No overall performance measures, such as the R^2 or Brier Score, were reported in any of the existing syncope prediction model studies.

The model performance was only partially assessed and reported in studies describing the development of a syncope clinical prediction tool. Overall performance measures such as the R^2 or Brier Score could be calculated and reported for readers to gauge the overall statistical performance of the prediction models. Calibration measures are important in prediction research to judge whether the predicted probabilities correspond with the observed probabilities of the outcome under study, yet model calibration was not addressed in 88% (n=7) of the studies examined[29,31]. Although one study did report elements of calibration using the Hosmer-Lemeshow goodness of fit statistic, this test has been demonstrated to provide unreliable results that are sensitive to fluctuations in sample size[62]. Future studies could be enhanced by providing a more comprehensive assessment and reporting of model calibration, such as the inclusion of a calibration plot and an assessment of the calibration intercept and slope[27].

1.6.6 Internal and External validation

Internal validation was performed in 25% (n=2) of the syncope prediction model development studies. External validation was performed in 75% (n=6) of the studies; however,

the majority (n=4) of the external validation methods involved temporal validation. Often times, the model performance measures decreased noticeably during model validation, suggesting that statistical overfitting may have influenced the model development. For example, in the study conducted by Sarasin *et al.*, the c-statistic decreased from 0.88 in the development phase to 0.84 during the internal validation phase; the c-statistic decreased even further to 0.75 in the external validation phase[8].

Internal validation was performed in a small proportion (25%, n=2) of the syncope prediction model studies. Future studies should include internal validation techniques to assess the robustness of the developed prediction model, and potentially identify any statistical issues that may have arisen during the model-fitting process[52]. A surprisingly large proportion of studies included an external validation component (75%, n=6), which is a very promising trend. Future studies should extend beyond temporal validation and focus on external validation studies in new populations to fully assess the generalizability of the prediction models[52].

1.6.7 Summary

We reviewed 9 published studies that report the development of a clinical prediction tool for risk stratifying syncope patients at risk of developing serious adverse outcomes. The statistical methodology employed in the development of syncope clinical decision tools was generally sub-optimal; some of the inadequacies include insufficient sample size, inappropriate treatment of missing data, and categorization of continuous variables. The model building strategies were largely data-driven and marked by bivariable screening and stepwise variable selection. These variable selection methods have their documented shortcomings; however, it may be difficult to accurately pre-specify the models to realistically avoid these limitations[30,47,49].

The reporting of model performance was deficient especially when addressing calibration, and internal validation procedures were under-utilized.

1.7 Thesis Objectives

There is a need for a robust and valid clinical decision tool for risk stratifying syncope patients at risk of developing severe adverse outcomes. Such a tool will allow clinicians to provide a more concrete basis for making disposition decisions, admitting/arranging appropriate follow-up for those at high risk and quick disposition of those at low risk of developing serious outcomes. Developing a robust clinical prediction model to risk-stratify patients has been identified as a priority for ED syncope research[63]. To date, no robust clinical decision tool meeting all methodological criteria has been developed and validated for this purpose. As outlined in our review assessing the current methodology employed in developing syncope prediction tools, the current tools possess some key methodological inadequacies that may limit their widespread reliability and applicability. Given the identified gaps in the literature, the main objectives of this thesis are:

1. To develop a robust clinical decision tool for identification of ED syncope patients at risk of serious outcomes after ED disposition;
2. To empirically examine the implications of key decisions in the statistical modeling process; in particular, to compare the statistical validity and usefulness of the tool derived under two alternative approaches that differ with respect to coding of predictors and selection of candidate predictor variables;
3. To assess the internally validated model performance of the tool using bootstrapping and various statistical measures, including discrimination and calibration.

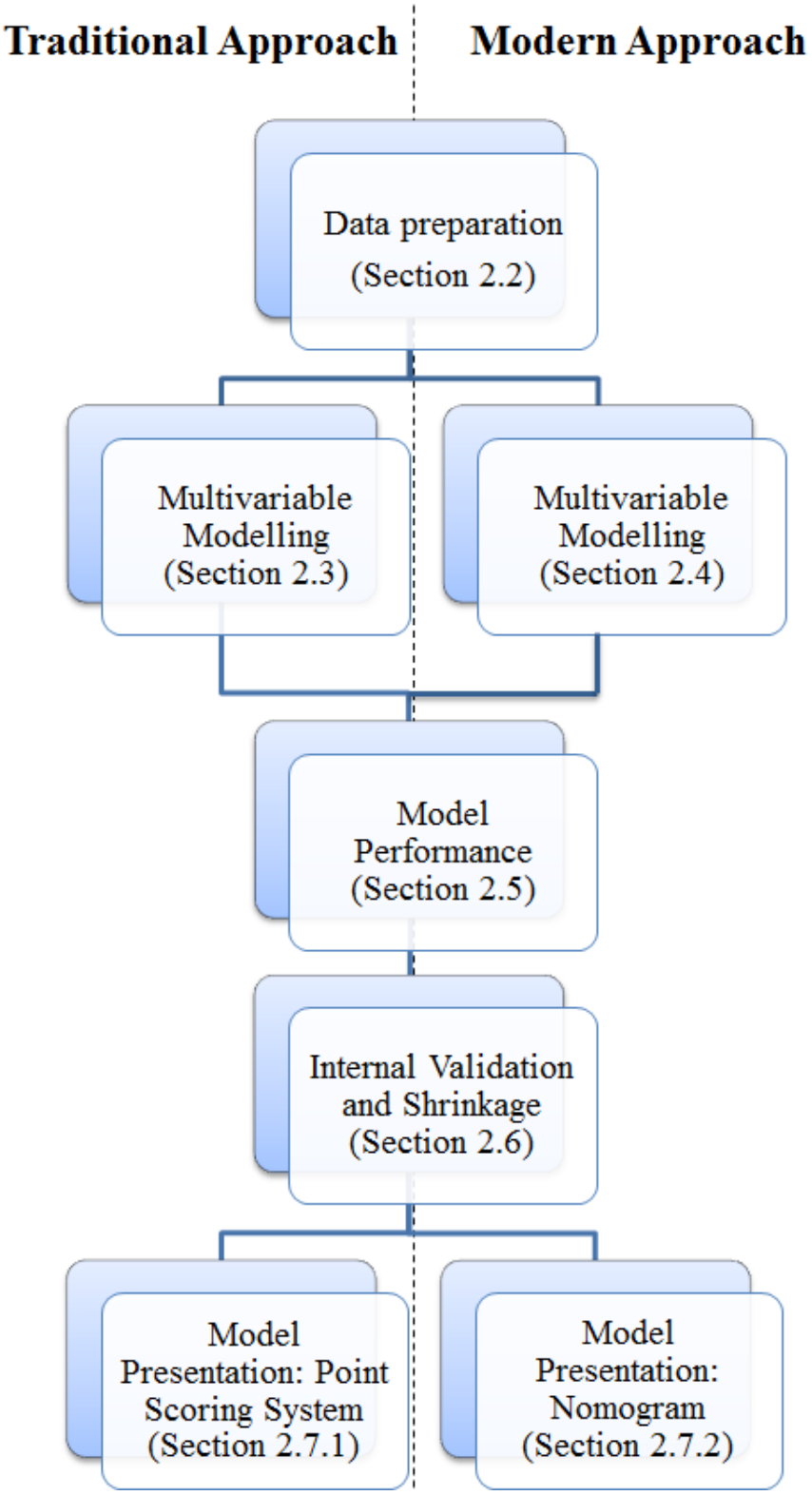
CHAPTER 2: METHODS

This thesis was conducted in the context of the Risk Stratification of Adult ED Syncope Study – Phase I study[64]. The main objective of this study was to develop a clinical prediction tool to accurately predict those at risk for serious outcomes after ED disposition among adult syncope patients.

The study protocol was conceived and developed by the principal investigator (Dr. Venkatesh Thiruganasambandamoorthy) and his team. The study initiation and data collection was completed by the research team before the commencement of the present thesis. The main objective of the overall study, namely the derivation of the clinical prediction tool, was addressed in this thesis. This thesis made original contributions to the overall study through the development and completion of a comprehensive data analysis plan used to derive the clinical prediction tool.

As outlined in the thesis objectives, we sought to empirically examine some of the key decisions in the statistical modelling process used in prediction model derivation. In order to investigate this objective, we derived the clinical prediction model using two different methodologic approaches, herein termed the "traditional approach" (section 2.3) and the "modern approach" (section 2.4). The specific philosophies and key methodological characteristics of these two approaches are listed in their respective methods sections. As a result of this comparison of statistical methodologies, some steps outlined in the methods section (Chapter 2) were applicable to both the "traditional" and "modern" approach. A schematic outlining the common and separate elements for the traditional and modern approach is presented in **Figure 1**.

Figure 1. Schematic diagram outlining methodology used to develop the syncope prediction model using the traditional and modern methodological approaches



Data preparation (Section 2.2) was done to create a clean analytic dataset to be used in both methodological approaches. The traditional approach (Section 2.3) was performed on the cleaned dataset to derive one prediction model, herein termed the "traditional model". Afterwards, the modern approach (Section 2.4) was performed on the same dataset to derive an alternative prediction model, herein termed the "modern model". The traditional and modern models each underwent the same assessment of model performance measures (Section 2.5), and internal validation and shrinkage (Section 2.6). The traditional and modern models were presented as a point scoring system (Section 2.7.1) and a nomogram (Section 2.7.2) respectively.

2.1 Study Overview

A brief overview of the study design and data collection is presented here to provide context for the data analysis; a full description can be found in a separate publication[64].
The study was a prospective observational cohort study in six Canadian EDs: The Ottawa Hospital (Civic and General Campuses), Kingston General Hospital, Hotel Dieu Hospital in Kingston, Foothills Medical Centre in Calgary, and the University of Alberta Hospital in Edmonton. Patient enrolment started at The Ottawa Hospital (Civic and General Campuses) in October 2010, while the other four study sites enrolled patients starting in January 2012. All sites enrolled patients until March 1, 2014. Patients presenting to the Emergency Departments (EDs) within 24 hours of the event, who were aged 16 or over, and who fulfilled the definition of syncope, namely sudden, transient loss of consciousness followed by spontaneous complete recovery were eligible for this study. Patients who were previously enrolled in this study, experienced prolonged loss of consciousness for longer than 5 minutes, had mental status changes from baseline, had significant trauma requiring admission, experienced loss of

consciousness due to alcohol intoxication or illicit drug abuse, or had obvious seizure or head trauma were ineligible for the study. The criterion for selecting study patients was consistent with established methodological guidelines for emergency department syncope risk-stratification research[65].

Patient assessments were completed by emergency medicine physicians or residents, and data were collected using standardized case report forms. The variables collected in this study were based on relevant clinical literature and input from subject-matter experts. The case report form consisted of sections covering patients' medical history, their pre-hospital care, and physical examinations, investigations and diagnoses in the ED. The section on patient history included demographical information, details of the syncope event, the patient's past medical history including family history of relevant comorbidities and a list of current medications. Pre-hospital care included whether the patient arrived by ambulance and details from any paramedic findings. Patient vital signs during triage, Glasgow Coma Scale score, and findings from the examination in the ED were collected in the physical examination section, while laboratory values, radiological investigations, electrocardiogram (ECG) characteristics, and other cardiac monitoring results were collected in the investigations section. In a final section, the ED physicians indicated their judgement of the diagnosis, their certainty of this diagnosis, and their estimated probability of a serious outcome, as well as any post-ED care decisions.

The main outcome of interest in this study (referred to as serious outcome, serious adverse event, outcome of interest, or event of interest) was a composite of any serious event, including death, arrhythmia, myocardial infarction, serious structural heart disease, aortic dissection, pulmonary embolism, severe pulmonary hypertension, subarachnoid hemorrhage, significant hemorrhage, or procedural interventions for treatment of syncope (e.g., pacemaker

and/ or defibrillator insertion, cardioversion for arrhythmias, surgery for valvular heart disease). Only events occurring after ED disposition and within 30 days following ED visit were considered. Serious outcomes identified in the ED during the evaluation do not require risk-stratification, since these patients are in the position to receive immediate and appropriate treatment by the attending healthcare professionals. Hence, patients who had serious outcomes that were identified in the ED were excluded from the analysis (section 2.2.1). Only patients that had the event of interest occurring outside of the ED were considered as outcomes in this study, as a risk stratification tool that detected serious outcomes unobserved in the ED would be most clinically-useful and sensible. Outcomes were assessed through a multi-step and systematic approach consisting of medical records review at local and regional hospitals, telephone follow-ups, and matching records at provincial coroner's offices.

The study was approved at the Ottawa Hospital Research Institute (the coordinating center for the study) by the Ottawa Hospital Research Ethics Board[64]; ethics was also approved by all the ethics boards from each of the study sub-sites. No further ethics approval was required for the thesis.

2.2 Data Preparation

All data preparation and cleaning were performed using SAS version 9.4 (SAS Institute, Cary. U.S.). The analytic dataset for this thesis included 4034 patients enrolled from October 2010 to March 1, 2014. Patients who had outcomes in the ED or were lost to follow-up were excluded from the analysis. Patients were deemed loss to follow-up if they had unknown outcome status at 30 days after ED visit and could not be reached by telephone follow-up after 5

separate attempts. The data preparation was done blinded to the predictor-outcome relationship to avoid biasing the analyst and increasing the risk of Type I error.

2.2.1 Data Cleaning and Recoding of Variables

A preliminary list of candidate predictors was prepared based on data collected in the case report forms, as well as subject-matter expertise. The list of candidate predictors, along with a brief description, is presented in **Table 4**. The distribution of each candidate predictor was examined for characteristics of data sparseness, skewness, and overall shape of the distribution. Each candidate variable was also inspected for extreme values; such values were examined to identify possibly erroneous values. For continuous variables, descriptive statistics and visual inspection of histograms were used to identify extreme observations. Values that appeared to be outside the range of biological plausibility were reviewed and corrected, where possible; otherwise they were set to missing. Patient characteristics such as demographics, clinical features, and ED management, were summarized using means, ranges, and standard deviations for continuous variables, whereas percentages were used for categorical variables.

Table 4. Preliminary list of candidate predictors based on clinical knowledge and data availability

VARIABLE NAME	SCALE	DEFINITION
Demographics		
Age	Continuous	Patient's age at date of visit (units: year)
Sex	Dichotomous	Patient's sex
Visit History		
Systolic blood pressure at triage	Continuous	Value of systolic blood pressure measured during assessment at ED triage (units: mmHg)
Lowest systolic blood pressure in the ED	Continuous	Lowest systolic blood pressure measurement recorded over duration of ED visit (units: mmHg)
Highest systolic blood pressure in the ED	Continuous	Highest systolic blood pressure measurement recorded over duration of ED visit (units: mmHg)
Diastolic blood pressure at triage	Continuous	Value of diastolic blood pressure measured during assessment at ED triage (units: mmHg)
Lowest diastolic blood pressure in the ED	Continuous	Lowest diastolic blood pressure measurement recorded over duration of ED visit (units: mmHg)
Highest diastolic blood pressure in the ED	Continuous	Highest diastolic blood pressure measurement recorded over duration of ED visit (units: mmHg)
Lowest heart rate in the ED	Continuous	Lowest heart rate measurement recorded over duration of ED visit (units: beats / minute)
Highest heart rate in the ED	Continuous	Highest heart rate measurement recorded over duration of ED visit (units: beats / minute)
Pulse rate at triage	Continuous	Pulse rate at ED triage (units: beats / minute)
Respiratory rate at triage	Continuous	Respiratory rate measured during assessment at ED triage (units: breaths/ minute)
Oxygen saturation at triage	Continuous	Oxygen saturation measured during assessment at ED triage (units: percentage)
Hemoglobin	Continuous	Hemoglobin measured in the ED (units: g/L)
Hematocrit	Continuous	Hematocrit measured in the ED (units: proportion)
Creatinine	Continuous	Creatinine measured in the ED (units: μmol/L)
Blood urea nitrogen	Continuous	Blood urea nitrogen measured in the ED (units: mmol/L)
Troponin	Continuous	Troponin measured in the ED (units: μg/L)
Medical History		
Coronary artery disease	Dichotomous	Patient history of coronary artery disease
Valvular heart disease	Dichotomous	Patient history of valvular heart disease
Cardiomyopathy	Dichotomous	Patient history of cardiomyopathy
Ventricular arrhythmia	Dichotomous	Patient history of ventricular arrhythmia

VARIABLE NAME	SCALE	DEFINITION
Atrial fibrillation / flutter	Dichotomous	Patient history of atrial fibrillation or atrial flutter
Pacemaker	Dichotomous	Patient history (or current) of having cardiac pacemaker
Hypertension	Dichotomous	Patient history of hypertension
Diabetes	Dichotomous	Patient history of diabetes
Transient ischemic attack	Dichotomous	Patient history of transient ischemic attack
Stroke	Dichotomous	Patient history of stroke
Gastrointestinal bleed	Dichotomous	Patient history of gastrointestinal bleed
Dementia	Dichotomous	Patient history of dementia
Pulmonary embolism	Dichotomous	Patient history of pulmonary embolism
Pulmonary hypertension	Dichotomous	Patient history of pulmonary hypertension
Deep vein thrombosis	Dichotomous	Patient history of deep vein thrombosis
Implantable cardioverter-defibrillator (ICD)	Dichotomous	Patient history (or current) of having implantable cardioverter-defibrillator
Malignancy	Dichotomous	Patient history of malignancy (palliative/ treatment within 6 months)
Congestive heart failure	Dichotomous	Patient history of congestive heart failure
Event Details		
Palpitations	Dichotomous	Palpitations prior to syncope event
Physical exertion prior to event	Dichotomous	Level of exertion prior to syncope event. High exertion defined by standing, walking, or high activity prior to syncope event; low exertion defined by sitting or lying prior to syncope event.
Predisposition for vasovagal	Dichotomous	Predisposition for vasovagal syncope (warm-crowded place, prolonged standing, fear, pain/emotion)
Prodrome	Dichotomous	Prodrome prior to syncope (lightheadedness/ dizziness, nausea, vomiting, abdominal cramps, or sweating)
Orthostatic symptoms	Dichotomous	Orthostatic symptoms prior to syncope
Symptoms post-event	Dichotomous	Presence of symptoms post-event; including palpitations, chest pain or angina equivalent, abdominal pain, headache, focal neurological or speech deficits, or shortness of breath
Witnessed by someone	Dichotomous	Syncope episode witnessed by someone else
MD suspected diagnosis	Categorical (4 levels)	Final ED diagnosis of cardiac syncope, vasovagal syncope, orthostatic syncope, or unknown cause
ECG		
QRS duration	Continuous	QRS duration value measured on first ED ECG (units:

VARIABLE NAME	SCALE	DEFINITION
		milliseconds)
QRS axis	Continuous	QRS axis value measured on first ED ECG (units: milliseconds)
Corrected QT interval	Continuous	Corrected QT value measured on first ED ECG (units: milliseconds)
Non-sinus rhythm on ED ECG	Dichotomous	Non-sinus rhythm detected on ED ECG
Right bundle Branch Block	Dichotomous	Right bundle branch block detected on ED ECG
Left bundle branch block	Dichotomous	Left bundle branch block detected on ED ECG
Left anterior fascicular block	Dichotomous	Left anterior fascicular block detected on ED ECG
Left posterior fascicular block	Dichotomous	Left posterior fascicular block detected on ED ECG
Right ventricular hypertrophy	Dichotomous	Right ventricular hypertrophy detected on ED ECG
Left ventricular hypertrophy	Dichotomous	Left ventricular hypertrophy detected on ED ECG
Left axis deviation	Dichotomous	Left axis deviation detected on ED ECG
Right axis deviation	Dichotomous	Right axis deviation detected on ED ECG
Non-specific ST-T changes	Dichotomous	Non-specific ST-T wave changes detected on ED ECG
Repolarization abnormalities	Dichotomous	Repolarization abnormalities detected on ED ECG
Secondary ST-T changes	Dichotomous	Secondary ST-T wave changes detected on ED ECG
Atrioventricular block	Categorical (4 levels)	Type of Atrioventricular block detected on ED ECG (defined as I, II- subtype 1, II- subtype-2, or III)

Frequency tables and plots were constructed to identify abnormal entries for categorical variables. The distribution of observations for each categorical predictor was assessed. Categories were collapsed or reassigned if there were limited observations within certain categories. Using subject-matter expertise, relevant composite variables were created by combining several variables to create meaningful clinical predictors.

Continuous variables with skewed distributions were transformed with the goal of generating approximate normal distributions to improve the central limit theorem assumptions which underlie many statistical tests[66]. Log transformations were used to generate approximate normal distributions for right-skewed variables, while square root transformations were used for left-skewed distributions.

2.2.2 Refining list of Candidate Predictors

A list of eligible candidate predictor variables for consideration in the multivariable model was prepared from the initial list of potential variables. The list was prepared with consideration of proportion of missing values, data sparseness, inter-rater agreement, and collinearity with other candidate predictor variables. Refining the list of eligible candidate predictors using this approach does not increase the risk of statistical over-fitting since the predictor-outcome relationship is not used to inform this process[28,30].

First, candidate predictors with substantial missing values (>25%) were excluded as such variables were likely to be of limited use in the clinical setting[29]. Although multiple imputation was used to deal with missing values (see section 2.3.1 and 2.4.1), imputation of predictors with substantial missing values may lessen the face validity of the prediction tool, and may increase the risk of producing an unstable and unreliable model[30].

Second, distributions for both categorical and continuous variables were assessed to identify variables with sparse distributions. Variables with sparse distributions and little variation were excluded from the list of candidate predictors, as these variables have limited utility when included in a clinical decision tool[67]. Categorical variables that had fewer than 100 observations in the smallest category were identified as having insufficient variation. We established a

minimal threshold of 100 observations based on the assumption that the expected prevalence of the event of interest would be approximately 5%. Thus, a total category of 100 patients would have an expected frequency of only 5 serious outcomes, which was considered insufficient for subsequent inclusion in a multivariable regression model.

Third, inter-rater reliability of selected candidate predictors was assessed. We planned to assess inter-observer agreement among 10% of patients for appropriate predictors where there was a subjective component in the data collection process (e.g., congestive heart failure in medical history). These inter-observer assessments were performed independently by a second emergency physician with patient consent. The inter-observer agreement was determined by calculating the kappa value for each variable. Candidate predictors with a kappa value ≤0.4 were excluded, as only variables that can be reliably assessed should be included in a prediction model[37-40]. The inter-rater agreement was not determined for the other candidate predictors because these variables were ascertained using objective medical measurements (e.g., automated machines to measure vitals such as heart rates in the ED).

Finally, collinearity was assessed among the list of candidate predictor variables using a variable clustering procedure (SAS PROC VARCLUS) and using variance inflation factors (VIFs) in SAS PROC REG[68]. The variable clustering procedure divides numerical variables into hierarchical clusters; each cluster represents variables that display linear dependencies with other variables within the same cluster[68]. The minimum proportion of variance explained by a cluster component was set at 0.70. The VIFs measure the inflation in the variances of the parameter estimates that result from the inclusion of multiple collinear predictors in the same model. Variables with variance inflation factors >2.5 were identified as potentially collinear. Variables

with near linear dependencies were identified and we ensured that variables involved in near linear dependencies would not be included in the same final prediction model[67].

After this process, the prepared list of eligible candidate predictors was evaluated one more time based on subject-matter expertise and clinical judgement and any variables considered inappropriate were excluded, prior to preparing the final list of candidate predictor variables. Throughout the data preparation process, we emphasized the necessity to restrict the number of candidate predictors considered and sought to develop a parsimonious list of candidate predictors that were both clinically-sensible and statistically-viable. The overall purpose of this procedure was to reduce the risk of statistical overfitting and Type I error, ultimately decreasing the chance of including spurious predictors in the final prediction model[28,30].

2.2.3 Investigating the Missing Data Mechanisms

We investigated the missing data mechanisms using our observed data to help inform the most appropriate approach to handle missing data in our dataset. We compared the patient characteristics of those with and without any missing data to investigate and potentially rule out the missing completely at random (MCAR) assumption. This was done by comparing predictor variables between those with complete data and those with missing data in at least one predictor variable. Chi-squared tests were used to evaluate this association for categorical predictors and t-tests were used for continuous predictors.

2.3 Multivariable Modeling for the Traditional Approach

The multivariable regression analysis for the traditional model was completed in four main steps: multiple imputation of predictors, bivariable screening of candidate predictors, categorization of continuous predictors, and stepwise multivariable logistic regression analysis.

The traditional approach is largely defined by data-driven methodology and hence some have argued that it leads to significant statistical overfitting and produces prediction models that perform poorly when applied to external populations[28,30]. However, there are numerous decision tools derived using similar methodology to the traditional approach that have continued to successfully validate on external populations, and have had profound implications in influencing clinical decision-making worldwide[38,40,69].

The data analysis of the traditional method was performed using SAS version 9.4 (SAS Institute, Cary. U.S.) and R version 3.1.1 (R Foundation for Statistical Computing, Vienna. Austria).

2.3.1 Multiple Imputation

Multiple imputation for the traditional approach was performed using PROC MI. The starting values were calculated using the Expectation Maximization (EM) algorithm. Markov Chain Monte Carlo (MCMC) was used to generate a sequential chain in the data augmentation iterations of the imputation procedure[70]. MCMC is a general method for estimating the Bayesian posterior distribution, which is defined as the distribution of unobserved observations based on the observed data[70].

We performed multiple imputation using an imputation model that contained the analysis variables (including the outcome of interest) and auxiliary variables, which are variables that are

not considered for the prediction model but may contribute information about missing values, to support the imputation process. After the first 200 "burn-in" iterations were discarded, each imputation dataset was generated after 100 iterations of the imputation procedure, producing ten imputation datasets for subsequent multivariable logistic regression. All continuous variables were retained in the continuous scale for the imputation procedure. All imputed values for categorical variables were rounded to the nearest plausible categorical value.

After each iteration of the imputation procedure, the imputation model was re-estimated using random draws from the posterior distribution of the parameters. This approach allows the imputation procedure to account for the inherent elements of uncertainty associated with generating estimated values for missing data. The stability of the imputation procedure was assessed using trace and autocorrelation plots. Trace plots graph the values of the coefficients of the imputation model throughout the iterations of data augmentation[70]. Autocorrelation plots depict the correlation between regression coefficients for different lag intervals of the data augmentation procedure[70].

The frequency distribution for each candidate predictor was compared between the imputed and original dataset to detect any major deviations that may have been introduced in the imputation procedure. The first imputation dataset generated from the multiple imputation procedure was used for all subsequent exploratory analysis.

2.3.2 Bivariable Screening

Bivariable tests were conducted to determine the significance of association of each predictor with the outcome of interest. The bivariable tests were conducted using the un-imputed dataset for convenience. The following bivariable statistical tests were used: chi-square tests with

continuity correction for nominal variables or Fisher's exact test where appropriate, two-tailed t-tests for continuous variables, and Mann-Whitney U test for ordinal variables[37-40]. Predictors that were significantly associated with the outcome at a 0.05 level of significance were considered for inclusion in the multivariable model[37-40].

2.3.3 Categorization of Continuous Variables

Continuous variables were categorized using cut-points determined by clinical and practical rationale, and verified using statistical methods. First, cut-points were established according to widely recognized definitions where they exist, e.g., age >75 which is commonly considered as the threshold for old age in emergency medicine literature. Second, cut-points were determined based on clinical reasoning, e.g., QRS duration ≤130 milliseconds is generally considered within normal limits; thus values exceeding 130 milliseconds are considered abnormal. Third, cut-points were determined using the upper limit of the reference interval in a healthy population (e.g.: troponin values greater than the 99th percentile of the reference population considered as having elevated troponin). Finally, convenient cut-points were specified by choosing 'round numbers' that were easily memorized and applied, such as categories divided by even 10 unit intervals.

To statistically verify the appropriateness of the selected cut-points, the functional relationship between each continuous predictor and the probability of the outcome (on the logit scale) was examined by grouping subjects with similar continuous values (e.g., grouped into deciles), and plotting these observed data along with an estimated restricted cubic spline transformation using methods outlined in section 2.4.3[28]. Graphs plotting the sensitivity, specificity, and Youden index (a general measure of diagnostic test performance, defined as

sensitivity + specificity – 1) for all potential values of continuous variables were constructed to help verify that cut-points were reasonable[71]. The point of intersection of the sensitivity and specificity, or optimization of the Youden index, was considered as the 'optimal' statistical cut-points; the 'optimal' cut-point was compared to the selected cut-point[71].

After categorization, the frequency distributions of the dichotomized variables were examined to verify that there were an adequate number of patients above and below the cut-point. If there were fewer than 5 expected events in the smallest category, we first considered whether a change in the cut-point resolved the issue; if not, the variable was excluded from consideration for multivariable modeling.

2.3.4 Variable selection

Candidate predictors that were strongly associated with the outcome ($p<0.05$) in the bivariable analysis were considered for inclusion in a multivariable logistic regression model. Stepwise backwards elimination was used to make the model more parsimonious. Variables with a significance of $p<0.05$ were retained in the model.

The backwards elimination procedure was applied to each imputation dataset to verify if the automated variable selection procedure selected different variables in the various imputation datasets. It is possible that the stepwise variable selection procedure could identify different variables for inclusion in the final model when applied across the different imputation datasets due to the random variation introduced in the imputation procedure. In the event that the stepwise variable selection procedure yielded different variables across imputation datasets, the variable dataset that was selected most frequently across imputation datasets was identified as

the final model. The final model was then applied to each of the 10 imputed datasets, and the combined point and variance estimates were estimated using rules developed by Rubin[35].

The model derived from the backward elimination procedure was also re-estimated using the original (e.g., non-imputed) data to examine whether there were dramatic differences in the regression parameters or model performance measures between the models estimated from the imputed and non-imputed datasets.

2.4 Multivariable Modeling for the Modern Approach

The methodology for the modern approach followed guidelines established by Harrell[28] and Steyerberg[30]. The modern approach emphasizes the need to reduce the risk of statistical overfitting and the use of non-linear transformations to flexibly model continuous variables. The model development process avoided a data-driven approach in favour of using clinical knowledge to pre-specify the prediction model. The model development process was characterized by three major steps: multiple imputation of predictors, pre-specification of a limited set of candidate predictors, and flexible modeling of continuous variables using restricted cubic splines. The data analysis of the modern approach was performed using R version 3.1.1 (R Foundation for Statistical Computing, Vienna. Austria).

2.4.1 Multiple Imputation

Multiple imputation was completed using the "aregImpute" function from the Hmisc Package in R. The aregImpute uses a sophisticated imputation procedure that offers many features to improve the quality of the imputations. For example, it automatically transforms continuous and categorical variables to maximize the correlation between variables and uses

flexible non-linear functions to model the associations between variables during imputation. In particular, aregeImpute models continuous variables using restricted cubic splines with 3 default knots, while categorical variables are expanded as dummy variables representing all categories within a categorical variable. Bootstrap resampling is used when fitting the imputation model to account for elements of uncertainty associated with this process. The aregImpute procedure employs predictive mean matching to generate imputed values that are realistic and consistent with the values seen in the original data[72]. To avoid generating repeated imputations with the same value in the predictive mean matching process, the aregImpute function randomly samples from the multinomial distribution of the probabilities to avoid over-representing specific values in the dataset.

The same variables used in the imputation model for traditional approach were used in the imputation of the modern approach using aregImpute (section 4.1.1). Ten multiple imputations datasets were generated, after the first 3 "burn-in" imputation iterations were discarded. Seventy-five bootstrap resamples were generated for each iteration of the imputation process. The frequency distribution for each candidate predictor was compared between the imputed and original dataset to detect any major deviations that may have been introduced in the imputation procedure.

2.4.2 Model Pre-specification

All variables for the modern approach were specified immediately after data cleaning and prior to conducting bivariable tests of association for the traditional approach (section 2.3.2).

The first step was to define a ranked list of candidate predictor variables. We identified the 25 most important predictor variables among the list of candidate predictors (after refining,

data cleaning, and treatment for missing values) based on clinical literature and content expertise. We developed a survey using Google Forms and circulated this survey via e-mail to our collaborative network of syncope researchers. The survey is attached as **Appendix 1**. The network consisted of 6 physicians with expertise in Emergency Medicine and Cardiology. The physicians were instructed to select 15 variables that they deemed were important predictors of serious outcomes in adult ED syncope patients. Our dataset justified fitting a model that included a maximum of 14 parameters (rationale described below); physicians were thus asked to select 15 important variables as this number corresponded approximately to the maximum number of potential variables that could be included in the final model. From this list of 15 variables, the physicians were asked to identify the 5 most important variables, and to then rank the remaining 10 variables in order of importance. We took this approach, as opposed to having the physicians identify and rank all 15 variables in order of importance, because we anticipated that it would be more challenging to differentiate among the 5 variables deemed most important.

The responses from the 6 physicians were compiled and an overall list of the 15 most important predictors was determined, ranked by their order of importance. The list was compiled by calculating the mean rank for each potential predictor pooled across the six survey responses. In particular, for each physician, the 5 variables identified as being the most important were assigned a rank of 1. The 10 remaining variables were assigned a rank of 6 to 15, corresponding to the rankings designated by the physician. Variables that were not selected as being among the top 15 most important variables by a physician were assigned a rank of 16 by default. We then calculated the mean ranking for each variable across the six survey responses, and ordered the variables in ascending order according to their mean rank. Variables with the lowest mean rank were considered to be the most important and were strongly considered for inclusion in the

prediction model. The full ranked list was reviewed to inform the model specification in the modern approach.

2.4.3 Modelling Predictors and adjusting Model Complexity

The next step was to pre-define the degree of complexity of the model, e.g., the number of parameters or degrees of freedom that can be included in the model. There were 147 events in the present syncope study, which allowed fitting a maximum of 14 parameters in the model.

Continuous predictors were modelled using restricted cubic spline functions with a default allocation of 3 knots (representing 2 degrees of freedom). We used the RMS package in R (developed by Harrell *et al.)*[28]. The three knots were located at fixed quantiles, specifically at the 0.10, 0.50, and 0.90 percentiles, of the predictor's marginal distribution rather than based on inspection of the observed relationship with the outcome to reduce the risk of overfitting[28].

The importance of each predictor variable in the presence of the other variables was then determined based on the partial model likelihood ratio chi-squared statistic and the final number of knots was assigned to each variable based on its predictive importance. Following the approach developed by Harrell, continuous predictors with higher chi-squared statistics were allocated more knots and thus degrees of freedom, while less important predictors were allocated fewer; categorical predictors with lower chi-squared statistics were collapsed into dichotomous variables using the category with the highest frequency[28]. These decisions were based purely on the importance of each predictor as opposed to statistical significance testing. Allocating degrees of freedom based on partial tests of association was a fair procedure because all variables were retained in the model, regardless of their statistical significance; and the partial chi-squared statistics do not reveal the degree of nonlinearity in the observed data[28]. After the degree of

freedom allocation was done based on importance of each predictor, the final model was estimated using the logistic regression procedure in R. The final model was estimated by combining the results across the ten imputation datasets using the rules developed by Rubin[35]. All inferences regarding statistical significance, confidence interval estimates, and statistical tests were ascertained from this full model as it provides the most reliable estimates of these measures[28].The full model was also re-estimated using the original (e.g., non-imputed) data to examine whether there were any discrepancies between the models estimated from the imputed and non-imputed datasets.

We assessed the non-linearity assumption and explored the functional relationship between the continuous predictors and the outcome by generating graphs that plot the observed data along with the fitted restricted cubic spline functions from the full model. We performed this assessment after the final model was fit to avoid biasing the analyst, who may have inadvertently changed the degree of complexity used to model continuous variables after visually examining the predictor-outcome relationship.

2.5 Assessment of Model Performance

In both the traditional and modern approaches, the assumptions for the estimated logistic regression model were first assessed using regression diagnostics, including identification of outliers and influential observations. Then, model performance was assessed using measures of overall predictive ability, discrimination, and calibration. Since the traditional and modern approach were analyzed using different statistical software packages, some performance measures and graphs were generated in one approach but not the other, based largely on the built-in functionality of the respective software packages.

2.5.1 Outliers and Influential Points

Outliers and influential observations were identified using DFBETA statistics. DFBETA statistics identify influential points by measuring how individual regression coefficients change after the omission of each observation in turn, scaled by their standard errors[73]. Observations that had DFBETA values >0.20 were identified as influential, and these observations were checked for data extraction or entry errors.

2.5.2 Overall Performance

We first assessed the overall model performance using the Nagelkerke's R^2 and the Brier Score. The Nagelkerke's R^2 quantifies overall model fit for binary outcomes by comparing the logarithm of predictions to the actual outcome[30]. The Brier score quantifies model fit by measuring the squared difference between the predictions and the outcomes[30]. The Brier score is typically less severe than Nagelkerke's R^2 in penalizing false predictions that have predicted probabilities close to 0 and 100%[30]. A scaled Brier score was also calculated since the regular scaled Brier score can be difficult to interpret as its range varies depending on the incidence of the outcome[30].

2.5.3 Discrimination and Calibration

ROC curves were constructed to graphically illustrate the model discrimination. Discrimination box plots comparing the estimated mean probabilities stratified by those who did and did not suffer serious outcomes were also generated to assess model discrimination. Models with better discrimination should show less overlap in the predicted probabilities between the

two groups. The discrimination slope, a measure that quantifies the mean difference of predictions between those with and without outcomes, was also calculated.

Calibration plots were created to graphically illustrate model calibration. In a calibration plot, predicted probabilities are plotted on the x-axis with observed proportions of events on the y-axis[30]. To create calibration plots for our model, subjects with similar predicted probabilities were grouped (e.g., based on deciles) and the mean observed outcomes were calculated for each decile group[30]. The calibration plot consisted of points representing each group, rather than each observation. Good calibration was characterized by small differences between the observed and predicted probabilities. Furthermore, the Hosmer-Lemeshow goodness of fit test was calculated to assess model calibration[30].

2.6 Internal Validation and Shrinkage

Internal validation was performed using 500 bootstrap samples in both the traditional and modern approaches to assess model stability, determine the level of shrinkage, and calculate optimism-corrected performance measures.

2.6.1 Internal Validation

Harrell *et al.* presented an algorithm to assess the degree of overfitting resulting from the model building process; this algorithm is especially useful in evaluating overfitting in models derived using stepwise variable selection procedures[74]. In this algorithm, the model is estimated separately using each bootstrap sample and subsequently evaluated in the original sample to calculate optimism. We used this procedure to calculate optimism-corrected performance measures (i.e.: c-statistic, R^2) for both the traditional and the modern models.

Confidence intervals for performance measures were calculated during the internal validation by using the non-parametric properties of the bootstrap re-sampling procedure. In particular, the 95% confidence intervals for performance measures were calculated based on the 2.5% and 97.5% percentiles of the performance measures from the bootstrap samples[30].

For the traditional approach, the consistency and stability of the stepwise variable selection procedure was also evaluated in this internal validation procedure. We assessed the frequency with which variables were selected in the multivariable model in each bootstrap sample. If variables from the traditional model were selected frequently in the bootstrap samples, these results would suggest that the automated variable selection procedure was fairly consistent.

2.6.2 Shrinkage

The model shrinkage estimator was calculated in the bootstrap resampling procedure calculating the pooled optimism in the calibration slopes across the bootstrap samples, and subtracting this value from one[28,75]. The regression coefficients from the logistic regression model were adjusted for shrinkage by multiplying the original coefficients by the shrinkage estimator; thereby making the regression coefficients less optimistic and thus more realistic when applied to future data. The regression model intercept was also re-calibrated after the shrinkage-adjusted regression coefficients were estimated. Shrinkage estimators can range from values of 0 to 1; values approaching one suggest a lower degree of overfitting. Current guidelines suggest that models with significant statistical overfitting, defined by a shrinkage estimator <0.9, should consider incorporating shrinkage into the regression model by calculating the shrinkage-adjusted regression coefficients[72].

2.7 Model Presentation

Once a statistical prediction model has been derived, it is typically presented in a format that can be used to easily calculate risks for individual patients in the clinical setting. Similar to the statistical considerations in developing clinical prediction models (section 1.4.1.3), some presentation methods emphasize simplicity with the trade-off of more imprecise predictions, while other presentation methods consist of more complex formats that offer more accurate and precise predictions[30]. Some common methods of prediction model presentation include a point scoring system, a regression formula, a nomogram, or a computational aid administered through an electronic application-based platform[30]. It is important to re-emphasize in this section that prediction models should be externally validated before implementation in the clinical setting.

2.7.1 Traditional Model: Point Scoring System

The model obtained from the traditional approach was presented using a point scoring system. Point scoring systems are easy to interpret because they translate each variable into a discrete point value that can be easily added to obtain a total risk estimate; however, this comes at the danger of less precise predictions since the derivation of the point scoring system inherently employs rounding of the regression coefficients. The traditional model was translated into a point scoring system by dividing all shrunken beta coefficients by the smallest beta coefficient, and rounding to the nearest integer[30,76]. The expected risk estimates for the point score values were derived using a logistic regression model with each patient's corresponding point score value as the sole predictor in the model. An additional calibration plot comparing the expected and observed probabilities at each point score level was constructed to assess the calibration of the scoring system.

The sensitivities and specificities at each cumulative point score threshold were calculated (e.g., sensitivity and specificity of the decision tool at a score of ≥2) to visualize these test parameters at various discrimination thresholds. We explored whether a feasible cut-point could be established to guide clinical decision making once the model is validated.

2.7.2 Modern Model: Nomogram

2.7.2.1 Approximating the Full Model

The full model specified for the modern approach contained 14 degrees of freedom, representing a possible maximum of 14 variables. This model was considered the 'gold standard', capable of producing honest and accurate standard errors and p-values[28]. Since this model would likely be cumbersome and impractical to implement in the clinical setting without an electronic aid, we sought to develop a more parsimonious model to be presented as a nomogram for use in the clinical setting (Section 2.7.2.2). The reduced model serves only as an approximation to the 'gold-standard' full model, and does not serve to replace the full model.

Harrell[28] recommended a "step-down" procedure that removes variables of limited predictive performance from the prediction model. The step-down procedure differs from conventional stepwise variable selection procedures because it removes variables based on pooled Wald chi-square statistics for the full model, as opposed to the Wald chi-square values for individual predictors that is characteristic of conventional stepwise variable selection[28]. As a result, the step-down procedure yields a stopping rule that is less arbitrary compared to conventional stepwise variable selection[28].

The step-down procedure was performed on the full model using the "fastbw" function in the RMS package. Variables were removed based on the pooled model Akaike's information

criterion (defined as the Wald chi-square statistic minus two times the degrees of freedom) to approximate the full model to the desired level of complexity. The reduced model was derived by applying the step-down procedure on the shrinkage-adjusted full model, thereby introducing shrinkage in the reduced model by default[28]. The reduced model was generated by removing variables until the full model Akaike's information criterion fell below a threshold of zero. Basic model performance measures were calculated and internal validation was performed to assess whether the reduced model adequately approximated the full model.

2.7.2.2 Nomogram

An approximated modern model was presented as a nomogram using the "nomogram" function in the RMS package. A nomogram is a graphical representation of a regression equation that can effectively describe complex non-linear relationships, and intuitively describe an individual's risk probability based on their specific predictor profile. Although nomograms can provide very accurate predictions using diverse regression equations, some researchers have criticized nomograms for their initial difficulty to understand[30].

A nomogram calculates the linear predictor and the risk estimate using the same process as the regression formula. The user calculates the number of scoring points for each variable based on the graphical tool; then the scoring points for all the individual variables are added together to obtain the linear predictor, which is then translated into a predicted risk estimates.

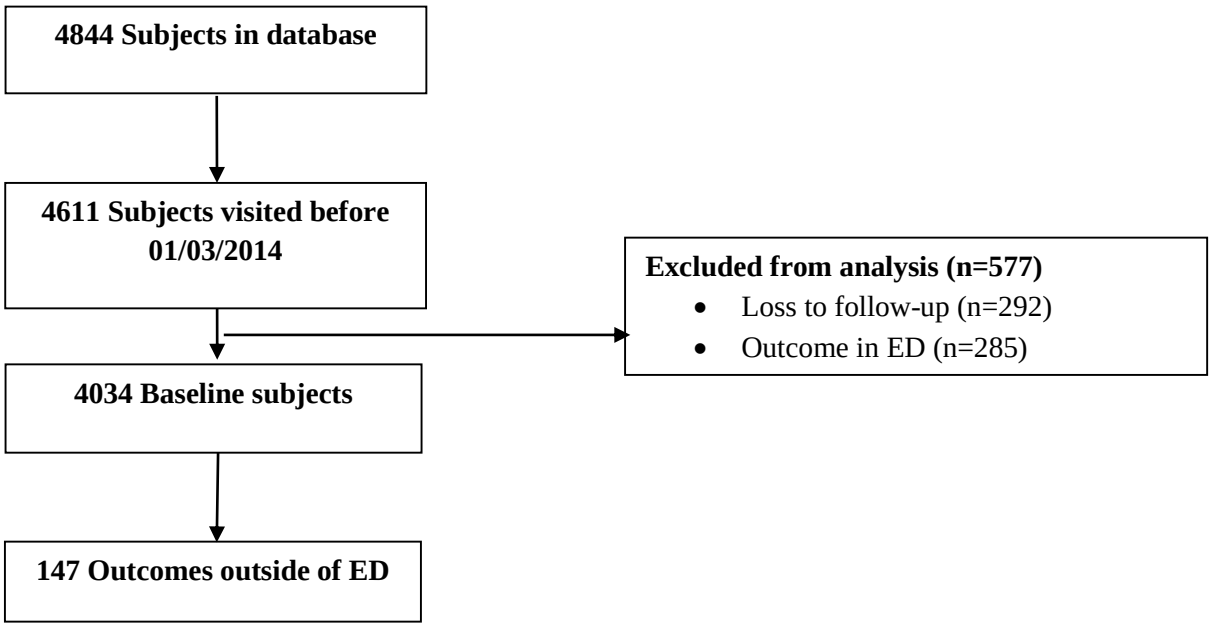
CHAPTER 3: RESULTS: DATA PREPARATION

This chapter presents the results from the data preparation process used to generate the final analytical dataset used in both the traditional and modern approaches.

3.1 Study Participants

The patient flow diagram is presented in **Figure 2**. Four thousand six hundred and eleven (4611) eligible patients were enrolled in our study between October 2010 and March 1, 2014. Patients who had outcomes in the ED (n=285) or who were lost to follow-up (n=292) were excluded from the analysis. Thus, 4034 patients were included in this study; of these, 147 patients experienced the event of interest outside the ED.

Figure 2. Study flowchart describing inclusion and exclusion of syncope patients



The characteristics of the 4034 study patients are described in **Table 5**. The study patients had an average age of 53.6 (SD= 23.0), consisted of 55.6% females, and had an 11.8%

prevalence of coronary artery disease. Among the syncope patients, 64.3% arrived to the ED by ambulance. Once in the ED, 95.1% had an ECG performed and 85.5% had blood tests completed. After ED disposition, 61 (of the 147 total patients who suffered a serious outcome) had the serious outcome outside of the hospital, while the other 86 suffered a serious outcome while hospitalized.

Table 5. Study patient characteristics, Emergency Department management, and outcomes

Characteristics	**N = 4,034 (%)**
Age (Mean, SD)	53.6 (23.0)
Range	16 to 102
Female	2241 (55.6)
Arrival by ambulance	2592 (64.3)
Medical History	
Hypertension	1294 (32.1)
Diabetes	402 (10.0)
Coronary artery disease	476 (11.8)
Atrial fibrillation/flutter	291 (7.2)
Valvular heart disease	137 (3.4)
Congestive heart failure	152 (3.8)
Management	
Electrocardiogram performed	3838 (95.1)
Blood tests performed	3449 (85.5)
Admitted to hospital	383 (9.5)
Serious Outcomes	
Serious Outcome while hospitalized	86 (2.1)
Serious Outcome outside the hospital	61 (1.5)

3.2 Candidate Predictors

3.2.1 Data Cleaning and Recoding of Variables

The list of potential predictors identified based on data availability and subject-matter expertise consisted of 59 predictors, 20 continuous and 39 categorical variables (Table 4).

Histograms showing the distributions of the continuous potential predictors are presented in **Appendix 2**. Five (5) extreme values for respiratory rate (>65 breaths/minute), 4 extreme values for QRS duration (< 25 of >250 milliseconds), and 6 extreme values for corrected QT interval (<200 milliseconds) were detected and set to missing. Patients with extreme systolic blood pressure measurements in the ED (>200 mmHg), blood urea nitrogen values (>75mol/L), and QRS axis values (<-180 and >180 milliseconds) were checked for data errors; five errors were identified and corrected. The variables blood urea nitrogen and creatinine were right skewed and a log transformation was used to successfully generate an approximate normal distribution for these variables; the log transformations were used in all subsequent analyses. Oxygen saturation was left skewed, but a square root transformation was unsuccessful in generating an approximate normal distribution.

Frequency tables of categorical variables were inspected and no erroneous values were identified. Two composite variables representing medical history were defined: history of cerebrovascular disease (presence of stroke or transient ischemic attack), and history of heart disease (presence of valvular heart disease, cardiomyopathy, congestive heart failure, coronary artery disease, ventricular arrhythmia, pacemaker/ICD, or non-sinus rhythm by ECG). In addition, two composite variables were created from relevant ECG measurements: bifascicular block (presence of a right bundle branch block, and either a left anterior fascicular block or left posterior fascicular block), and bundle-branch/first degree atrioventricular block (presence of a first degree atrioventricular block, and either a right bundle branch block or left bundle branch block).

The categorical variable that evaluates the MD suspected diagnosis of syncope had four categories: vasovagal syncope, orthostatic syncope, cardiac syncope, or unknown. We sought to

re-categorize this variable by collapsing the categories into clinically-meaningful groups. MD suspected diagnosis was thus recoded as a three-level categorical variable, consisting of vasovagal syncope, cardiac syncope, and orthostatic syncope/unknown. The orthostatic/unknown group was used as the reference category for all analyses.

ED ECGs were interpreted and extracted by ED physicians and/or cardiologists. We sought to combine the various interpretations to define one value for each variable. For ECGs that were interpreted by both ED physicians and cardiologist, the cardiologist's interpretation was used. For ECGs interpreted by only one clinician, we used that clinician's interpretation for all subsequent analyses.

3.2.2 Refining List of Candidate Predictors

In order to prepare the final list of eligible candidate predictor variables, we examined sparseness of distributions, presence of missing values, inter-rater reliability, and collinearity. The results are presented in **Table 6**. These results, together with a review using clinical considerations, were then used to prepare the final list of eligible candidate predictors for multivariable modeling.

Table 6. Eligibility of potential predictors based on sparseness of distributions, missing values, inter-rater reliability, and collinearity

Predictor	Continuous (min, max); Categorical (smallest %)	Missing (n, %)	Kappa Value (95% CI)	Collinearity (variable, R^2)
Continuous Variables				
Demographics				
Age	16 to 102	1 (0.0)		
Visit History				
Systolic blood pressure at triage	53 to 249	119 (3.0)		Lowest SBP in ED (R^2=0.76); Highest SBP in ED(R^2=0.74)
Diastolic blood pressure at triage	29 to 134	142 (3.5)		Lowest DBP ED (R^2=0.65); Highest DBP ED (R^2=0.62)
Lowest systolic blood pressure in the ED	48 to 222	30 (0.7)		SBP at Triage (R^2=0.86); Highest SBP ED (R^2=0.74)
Lowest diastolic blood pressure in the ED	18 to 186	33 (0.8)		DBP at Triage (R^2=0.83); Highest DBP ED (R^2=0.62)
Highest systolic blood pressure in the ED	67 to 249	30 (0.7)		SBP at Triage(R^2=0.86); Lowest SBP in ED (R^2=0.76)
Highest diastolic blood pressure in the ED	23 to 209	33 (0.8)		DBP at Triage (R^2=0.83); Lowest DBP in ED (R^2=0.65)
Pulse rate at triage	20 to 168	149 (3.7)		Lowest heart rate in ED (R^2=0.79); Highest heart rate in ED (R^2=0.81)
Respiratory rate at triage	8 to 60	812 (20.1)		
Oxygen saturation at triage	51 to 100	82 (2.0)		
Lowest heart rate in the ED	20 to 151	31 (0.8)		Pulse rate at triage (R^2=0.88); Highest heart rate in ED (R^2=0.81)
Highest heart rate in the ED	40 to 178	32 (0.8)		Lowest heart rate in ED (R^2=0.79); Pulse rate at triage (R^2=0.88)
Investigation				
Hemoglobin	61 to 183	662 (16.4)		Hematocrit (R^2=0.99)

Predictor	Continuous (min, max); Categorical (smallest %)	Missing (n, %)	Kappa Value (95% CI)	Collinearity (variable, R^2)
Creatinine	12 to 1403	729 (18.0)		Blood urea nitrogen (R^2=0.79)
Hematocrit	0.19 to 0.55	661 (16.4)		Hemoglobin (R^2=0.99)
Blood urea nitrogen	0.5 to 96.0	958 (23.8)		Creatinine (R^2=0.79)
Troponin	169 (4.3)	2102 (52.1)		
ECG				
QRS duration	30 to 206	221 (5.5)		
QRS axis	-150 to 507	219 (5.4)		
Corrected QT interval	302 to 682	223 (5.5)		
Categorical Variables				
Demographics				
Sex	44.5	0 (0.0)		
Medical History				
Cerebrovascular disease	6.3	7 (0.2)		
Heart disease	20.6	254 (6.3)		
Hypertension	32.1	4 (0.1)		
Diabetes	10.0	6 (0.2)		
Gastrointestinal bleed	1.1 ‡	7 (0.2)		
Dementia	3.3	4 (0.1)		
Pulmonary embolism	0.8 ‡	4 (0.1)		
Pulmonary hypertension	0.4 ‡	4 (0.1)		
Deep vein thrombosis	1.3 ‡	4 (0.1)		
Malignancy	2.3 ‡	6 (0.2)		
MD Data Form				
MD suspected diagnosis	6.1	4 (0.1)	0.60 (0.48, 0.72)	
Witnessed syncope	28.7	44 (1.1)	0.81 (0.71, 0.91)	
Palpitations prior to syncope	5.4	75 (1.9)	0.48 (0.27, 0.69)	
Physical exertion prior	39.6	48 (1.2)	0.66 (0.53,	

Predictor	Continuous (min, max); Categorical (smallest %)	Missing (n, %)	Kappa Value (95% CI)	Collinearity (variable, R^2)
to syncope			0.79)	
Predisposition to vasovagal syncope	41.2	63 (1.6)	0.48 (0.33, 0.64)	
Prodrome prior to syncope	24.2	69 (1.7)	0.56 (0.40, 0.72)	
Orthostatic symptoms	20.2	166 (4.1)	0.46 (0.29, 0.64)	
Symptoms post-event	22.0	356 (8.8)	0.41 (0.26, 0.57)	
ECG				
Bifascicular block	1.2‡	253 (6.3)		
Bundle-branch/ atrioventricular block	1.4 ‡	473 (11.7)		
Left bundle branch block	2.5	254 (6.3)		
Left ventricular hypertrophy	5.3	253 (6.3)		
Right axis deviation	1.9 ‡	252 (6.3)		
Old ischemia	4.4	254 (6.3)		
Right ventricular hypertrophy	0.4 ‡	253 (6.3)		
Left axis deviation	5.4	253 (6.3)		
ST-T changes ischemic of unknown age	2.3 ‡	255 (6.3)		
Non-specific ST-T changes	9.4	255 (6.3)		
Early repolarization changes	3.5	255 (6.3)		
Secondary ST-T changes	1.0 ‡	257 (6.4)		

‡Indicates sparse variable distribution (n<100)
SBP= systolic blood pressure
DBP= diastolic blood pressure

3.2.2.1 Variable Distribution Dispersion

Gastrointestinal bleed (n= 45), pulmonary embolism (n= 34), pulmonary hypertension (n= 17), deep vein thrombosis (n= 51), malignancy (n= 93), bifascicular block (n= 48), bundle-branch/atrioventricular block (n= 58), right ventricular hypertrophy (n= 14), right axis deviation (n= 75), ST-T changes ischemic of unknown age (n= 91), and secondary ST-T wave changes (n= 39), were excluded from the list of candidate predictors because they had fewer than 100 observations in their smallest categories. All continuous variables had sufficient variation to be considered as viable predictor variables.

3.2.2.2 Missing Values

All variables, with the exception of troponin, had <25% of missing data and were thus considered for inclusion in the list of candidate predictors. The troponin assay was only completed in 48% of the patients, leaving 52% of the patients with missing data for troponin. Despite the high level of missing values in the troponin measurement, we strongly believed it was a valuable predictor for a syncope prediction model and should be retained as a candidate variable. Thus, we dichotomized troponin even though it was recorded as a continuous predictor; values exceeding the site-specific 99^{th} percentile of the reference were deemed to have abnormal troponin values. Patients with no troponin assay completed were assigned to have normal troponin values. This assumption was supported by a subgroup analysis assessing the characteristics of patients who received a troponin tests compared to patients who did not receive the test, presented in **Appendix 3.** Troponin elevation is a marker for ischemic cardiac damage, which is more commonly found in older individuals with a medical history of poor cardiac health. The results from the subgroup analysis suggest that patients who did not receive a

troponin assay were younger, had a lower frequency of cardiovascular disease (coronary artery disease, valvular heart disease, congestive heart failure, etc.), had a higher frequency of vasovagal MD suspected diagnosis, had a lower frequency of cardiac MD suspected diagnosis, and were at lower risk of developing the event of interest. The findings from the subgroup analysis support the assumption that patients with no troponin assay completed in the ED had normal troponin values. All subsequent analyses (in both the traditional and modern method), treated troponin as a dichotomous variable.

3.2.2.3 Inter-rater agreement

Inter-rater agreement was measured using the kappa statistic for the following variables: cause of syncope, witness, palpitations, physical exertion, predisposition for vasovagal, prodrome, and orthostatic symptoms (Table 6). All variables with a measured kappa statistic had values ranging from 0.46 to 0.81 (Table 6). None of the variables were excluded due to low inter-rater reliability according to our defined minimal acceptability level of 0.4.

3.2.2.4 Collinearity

The variable clustering procedure identified 5 sets of correlated variables, consisting of systolic blood pressure (triage systolic blood pressure, highest ED systolic blood pressure, lowest ED systolic blood pressure), diastolic blood pressure (triage diastolic blood pressure, highest ED diastolic blood pressure, lowest ED diastolic blood pressure), heart rate (lowest heart rate in the ED, highest heart rate in the ED, pulse rate at triage), red blood cell measurements (hemoglobin, hematocrit), and renal function (blood urea nitrogen, creatinine).

The collinearity discovered within the variable groups was consistent with biological rationale and clinical expertise. Candidate predictors were not excluded based on collinearity

alone; rather, this information was used to ensure that multiple variables from one collinear group were not included in the final prediction models.

3.2.2.5 Exclusions based on subsequent Clinical Rationale

After eliminating variables using these statistical considerations, we conducted a final review using clinical judgement to arrive at the list of candidate predictor variables meeting all acceptability criteria. Hypertension and diabetes were excluded because their common prevalence in practice made them less sensible for inclusion in a risk tool. Moreover, dementia and symptoms post-event were excluded since these variables lacked clinical sensibility. Finally, three additional ECG variables: left axis deviation, non-specific ST-T wave changes, and early repolarization abnormalities were excluded because they were not believed to be clinically useful as standalone variables.

3.2.2.6 Summary of Candidate Predictors

The final list of eligible candidate predictor variables for multivariable modeling is presented in **Table 7**. The list consisted of 33 variables (19 continuous, 14 categorical). Variables are presented in related subgroups based on the sets of correlated variables identified in section 3.2.2.4. The variable with the highest missing rate was blood urea nitrogen with 23.8%, while the majority of the variables had a missing rate < 7%.

Table 7. Final list of candidate predictors considered for syncope prediction model

Predictor	Scale	Range/ levels	Missing (n, %)
Visit History			
Sex	Dichotomous	Male, Female	
Age	Continuous	16 to 102	1 (0.0)
Systolic blood pressure			
SBP at Triage	Continuous	53 to 249	119 (3.0)
Lowest SBP in ED	Continuous	48 to 222	30 (0.7)
Highest SBP in ED	Continuous	67 to 249	30 (0.7)
Diastolic blood pressure			
DBP at Triage	Continuous	29 to 134	142 (3.5)
Lowest DBP in ED	Continuous	18 to 186	33 (0.8)
Highest DBP in ED	Continuous	23 to 209	33 (0.8)
Heart rate			
Lowest HR in the ED	Continuous	20 to 151	31 (0.8)
Highest HR in the ED	Continuous	40 to 178	32 (0.8)
Pulse rate at triage	Continuous	20 to 168	149 (3.7)
Respiratory rate at triage	Continuous	8 to 60	812 (20.1)
Oxygen saturation at triage	Continuous	51 to 100	82 (2.0)
Medical History			
Cerebrovascular disease	Dichotomous	Yes, No	7 (0.2)
Heart disease	Dichotomous	Yes, No	254 (6.3)
Investigation			
Red blood cell			
Hemoglobin	Continuous	61 to 183	662 (16.4)
Hematocrit	Continuous	0.19 to 0.55	661 (16.4)
Renal function			
Creatinine	Continuous	12 to 1403	729 (18.0)
Blood urea nitrogen	Continuous	0.5 to 96	958 (23.8)
Troponin	Dichotomous	Yes, No	0 (0.0)
MD Assessment			
MD suspected diagnosis	3-level categorical	Cardiac, Vasovagal, Orthostatic/unknown	4 (0.1)
Witnessed by someone	Dichotomous	Yes, No	44 (1.1)
Palpitations	Dichotomous	Yes, No	75 (1.9)
Physical exertion prior to event	Dichotomous	High, low exertion	48 (1.2)
Predisposition for vasovagal	Dichotomous	Yes, No	63 (1.6)
Prodrome	Dichotomous	Yes, No	69 (1.7)
Orthostatic symptoms	Dichotomous	Yes, No	166 (4.1)

Predictor	Scale	Range/ levels	Missing (n, %)
ECG			
QRS duration	Continuous	30 to 206	221 (5.5)
QRS axis	Continuous	-150 to 507	219 (5.4)
Corrected QT interval	Continuous	302 to 682	223 (5.5)
Left bundle branch block	Dichotomous	Yes, No	254 (6.3)
Left ventricular hypertrophy	Dichotomous	Yes, No	253 (6.3)
Old ischemia	Dichotomous	Yes, No	254 (6.3)

SBP= systolic blood pressure
DBP= diastolic blood pressure
HR= heart rate

3.2.3 Missing Data Mechanism

Among the 4043 patients included in our study, 2052 (50.9%) had missing values for at least one of the 33 candidate predictors listed in Table 7. The large proportion of missing cases emphasizes the need to address missing data issues using imputation procedures; otherwise, up to half of the participants would be excluded from subsequent multivariable analysis.

The characteristics of patients with complete data on all predictors, and those with a missing value in at least one predictor are presented in **Appendix 4a** and **Appendix 4b**. Many significant differences were found. For example, patients with missing data were younger; had lower blood pressure values (systolic blood pressure at triage, highest systolic blood pressure in the ED, highest diastolic blood pressure in the ED); lower heart rates in the ED; higher oxygen saturation levels; and higher QRS axes. Patients with missing data were more likely to be female, have a vasovagal MD suspected diagnosis, syncope episode witnessed by someone else, predisposition for vasovagal syncope, and prodrome. Those with missing data were less likely to have a history of vascular disease, a cardiac MD suspected diagnosis, and orthostatic symptoms.

Overall, patients with missing data were systematically different than patients with complete cases. These results indicate that MCAR was not a realistic missing data mechanism in

our dataset. Details regarding the imputation models and results for the imputation used in both the traditional and modern approach are presented in Chapters 4 and 5 respectively.

CHAPTER 4: RESULTS: TRADITIONAL APPROACH

In this chapter, we present the results from the derivation of the clinical prediction tool using the "traditional approach". In particular, the results presented in this section include the multiple imputation, bivariable screening, categorization of continuous variables, and variable selection used to develop the prediction model under the considerations of the traditional approach. The results describing model performance; internal validation and shrinkage; and model presentation as a point scoring system are also presented in this chapter.

4.1 Initial Modelling Steps

4.1.1 Multiple Imputation

The imputation model included all the candidate predictors considered for the prediction tool, along with the event of interest. In addition to these variables, the following auxiliary variables were included in the imputation model: dementia, hypertension, diabetes, symptoms post-event, left anterior fascicular block, left axis deviation.

Trace plots from the multiple imputation are presented in **Appendix 5**. Trace plots that fluctuate randomly (e.g., display no evident trends) indicate that the coefficients of the imputation model are unrelated throughout the sequential iterations of the data augmentation process, suggesting that convergence for the imputation procedure has likely been achieved. The autocorrelation plots from the imputation procedure are presented in **Appendix 6**. Correlation values that approach zero indicate that negligible correlation exists between the parameter values of the defined lag iterations, suggesting that the imputation procedure has likely converged.

Overall, these results suggest that the multiple imputation procedure was stable and that convergence was achieved.

To examine the effect of the multiple imputation procedure on the analytic dataset, frequency distributions for each candidate predictor were compared between the original data and the first imputation dataset. The results are presented in **Appendix 7a** and **Appendix 7b**. They show that there were no substantial differences in the frequency distribution of categorical variables between the original and imputed dataset using PROC MI, and that the mean and standard deviations of continuous variables were comparable between the original and imputed dataset.

4.1.2 Bivariable Screening

The results of the bivariable tests of association with each predictor and the outcome are presented in **Table 8a** and **Table 8b**. From the original list of 33 candidate predictors (19 continuous, 14 categorical), a total of 23 predictors (14 continuous, 9 categorical) had significant associations with the outcome and were considered for inclusion in the multivariable model. The variables having significant bivariable associations with the outcome are indicated in Table 8a and Table 8b.

Table 8a. Bivariable screening for continuous variables to determine the unadjusted strength of association between each continuous predictor and the serious outcome

	No Serious outcome (n=3887)	**Serious outcome (n=147)**	
Variable	**Mean (sd)**	**Mean (sd)**	**p-value**
Demographics			
Age*	52.9 (23.0)	71.6 (16.7)	<0.0001
Visit History			
Systolic blood pressure at triage*	125.3 (21.4)	133.3 (33.0)	0.005
Lowest systolic blood pressure in the ED	115.0 (19.2)	115.2 (25.2)	0.91
Highest systolic blood pressure in the ED*	137.1 (22.1)	149.1(29.5)	<0.0001
Diastolic blood pressure at triage	73.7 (13.1)	74.2 (17.7)	0.71
Lowest diastolic blood pressure in the ED*	65.3 (12.8)	61.6 (17.3)	0.01
Highest diastolic blood pressure in the ED*	80.4 (13.7)	83.6 (17.6)	0.03
Lowest heart rate in the ED	69.9 (14.0)	67.9 (17.7)	0.17
Highest heart rate in the ED	84.0 (16.6)	87.2 (19.7)	0.06
Pulse rate at triage	77.4 (16.8)	79.1 (21.3)	0.38
Respiratory rate at triage*	17.2 (2.6)	18.0 (3.8)	0.04
Oxygen saturation at triage*	96.6 (3.2)	94.4 (4.3)	<0.0001
Investigation			
Hemoglobin*	134.8 (17.3)	127.9 (21.2)	0.0002
Hematocrit*	0.40 (0.05)	0.38 (0.06)	0.0003
Creatinine (log-transformed)*	4.37 (0.42)	4.58 (0.44)	<0.0001
Blood urea nitrogen (log-transformed)*	1.72 (0.49)	2.00 (0.56)	<0.0001
ECG			
QRS duration*	93.0 (17.6)	111.1 (31.7)	<0.0001
QRS axis*	35.6 (42.3)	16.0 (60.4)	0.0003
Corrected QT interval*	431.7 (30.7)	463.6 (43.3)	<0.0001

*significant in bivariable screening and considered for inclusion in subsequent multivariable regression model

Table 8b. Bivariable screening for categorical variables to determine the unadjusted strength of association between each categorical predictor and the serious outcome

	No Serious outcome (n=3887)	**Serious outcome (n=147)**	
Variable	**n (row %)**	**n (row %)**	**p-value**
Demographics			
Sex*			0.0016
Male	1709 (95.3)	84 (4.7)	
Female	2178 (97.2)	63 (2.8)	
Medical History			
History of vascular disease*			0.0210
No	3641(96.5)	131 (3.5)	
Yes	239 (93.7)	16 (6.3)	
History of heart disease*			<0.0001
No	2894 (98.1)	55 (1.9)	
Yes	748 (90.0)	83 (10.0)	
Investigation			
Troponin*			<0.0001
No	3755 (97.2)	110 (2.9)	
Yes	132 (78.1)	37 (21.9)	
MD Data Form			
MD suspected diagnosis*			<0.0001
Cardiac	194 (79.2)	51 (20.8)	
Vasovagal	2166 (99.1)	19 (0.9)	
Other	1526 (95.3)	76 (4.7)	
Witness			0.47
No	1112 (96.0)	46 (4.0)	
Yes	2733 (96.5)	99 (3.5)	
Palpitations			0.29
No	3604 (96.3)	137 (3.7)	
Yes	213 (97.7)	5 (2.3)	
Physical exertion			0.42
Low Exertion	2415 (96.5)	87 (3.5)	
High Exertion	1425 (96.0)	59 (4.0)	
Predisposition for vasovagal*			<0.0001
No	2182 (94.5)	128 (5.5)	
Yes	1642 (98.9)	19 (1.1)	
Prodrome*			<0.0001
No	915 (93.7)	62 (6.4)	
Yes	2906 (97.3)	82 (2.7)	
Orthostatic symptoms			0.51
No	2937 (96.2)	116 (3.8)	
Yes	788 (96.7)	27 (3.3)	

	No Serious outcome (n=3887)	Serious outcome (n=147)	
Variable	**n (row %)**	**n (row %)**	**p-value**
ECG			
Left bundle branch block*			<0.0001
No	3563 (96.8)	117 (3.2)	
Yes	85 (85.0)	15 (15.0)	
Left ventricular hypertrophy*			0.03
No	3450 (96.7)	119 (3.3)	
Yes	199 (93.9)	13 (6.1)	
Old ischemia			0.23
No	3481 (96.6)	123 (3.4)	
Yes	167 (94.9)	9 (5.1)	

*significant in bivariable screening and considered for inclusion in subsequent multivariable regression model

Significant and important associations were identified during the bivariable tests of association. Patients who had the outcome of interest were older, which is sensible given that the risk of many elements included in the composite serious outcome, such as death, myocardial infarction, and arrhythmia, are known to increase with age. With respect to medical history, patients who had a serious event were more likely to have pre-existing cardiovascular diseases, as evidenced by the higher proportion of cerebrovascular disease and heart disease in those who had the event of interest. It is reasonable that syncope patients with pre-existing cardiovascular conditions are more likely to suffer serious events, which are largely cardiovascular-related[77]. Similarly, patients with the event of interest were more likely to have an MD suspected diagnosis of cardiac syncope, and less likely to have a MD suspected diagnosis of vasovagal syncope. Vasovagal syncope is considered low-risk for the development of serious outcomes[77]. Likewise, patients with a predisposition to vasovagal syncope were less likely to suffer a serious event.

Patients who suffered the event of interest also generally had worse vital signs measured in the ED (higher triage and maximum-recorded systolic blood pressure, and lower oxygen saturation). These patients also had poorer investigation results, such as lower hemoglobin,

higher creatinine, and elevated troponins, which are suggestive of impaired renal and cardiac function. Furthermore, patients who had the event of interest had worse ECG characteristics, such as a higher corrected QT intervals and a larger incidence of left bundle branch block, which are characteristic of cardiac abnormalities. It is sensible that syncope patients who developed the outcome of interest had worse health status observed during their ED visits, increasing their susceptibility to developing serious outcomes within 30-days after ED disposition.

4.1.3 Categorization of Continuous Variables

The functional forms of the relationship between continuous predictors and the probability of the event (on the logit scale) are presented in **Figure 3**. The solid line represents the restricted cubic spline function with the corresponding 95% confidence intervals indicated using the dotted lines. The triangular points represent "bins" or percentile groups in the observed data, with approximately 400 patients in each group. The arrows denoted on the x-axis represent the locations of the knots in the restricted cubic spline functions. Graphs that plot the test characteristics (sensitivity, specificity, Youden index) at varying thresholds of the continuous variables are presented in **Figure 4**. The solid red line represents the sensitivity, the dotted blue line represents the specificity, and the solid purple line represents the Youden index. We used these statistical plots to verify the clinically-specified cut-points presented in **Table 9**.

Figure 3. Functional association between continuous predictors and the probability (logit scale) of the serious outcomes

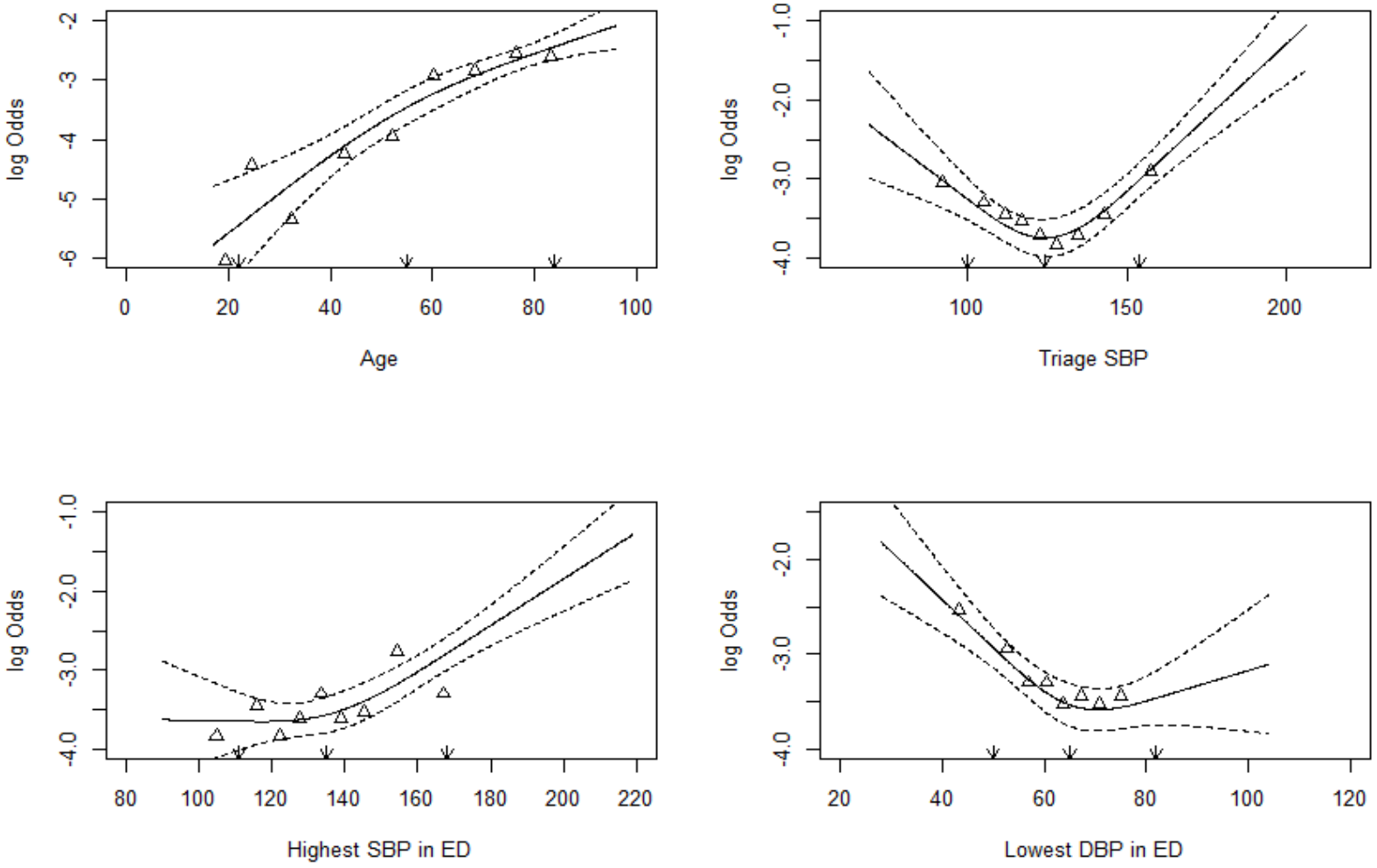


Figure 3. Functional association between continuous predictors and the probability (logit scale) of the serious outcomes (cont'd)

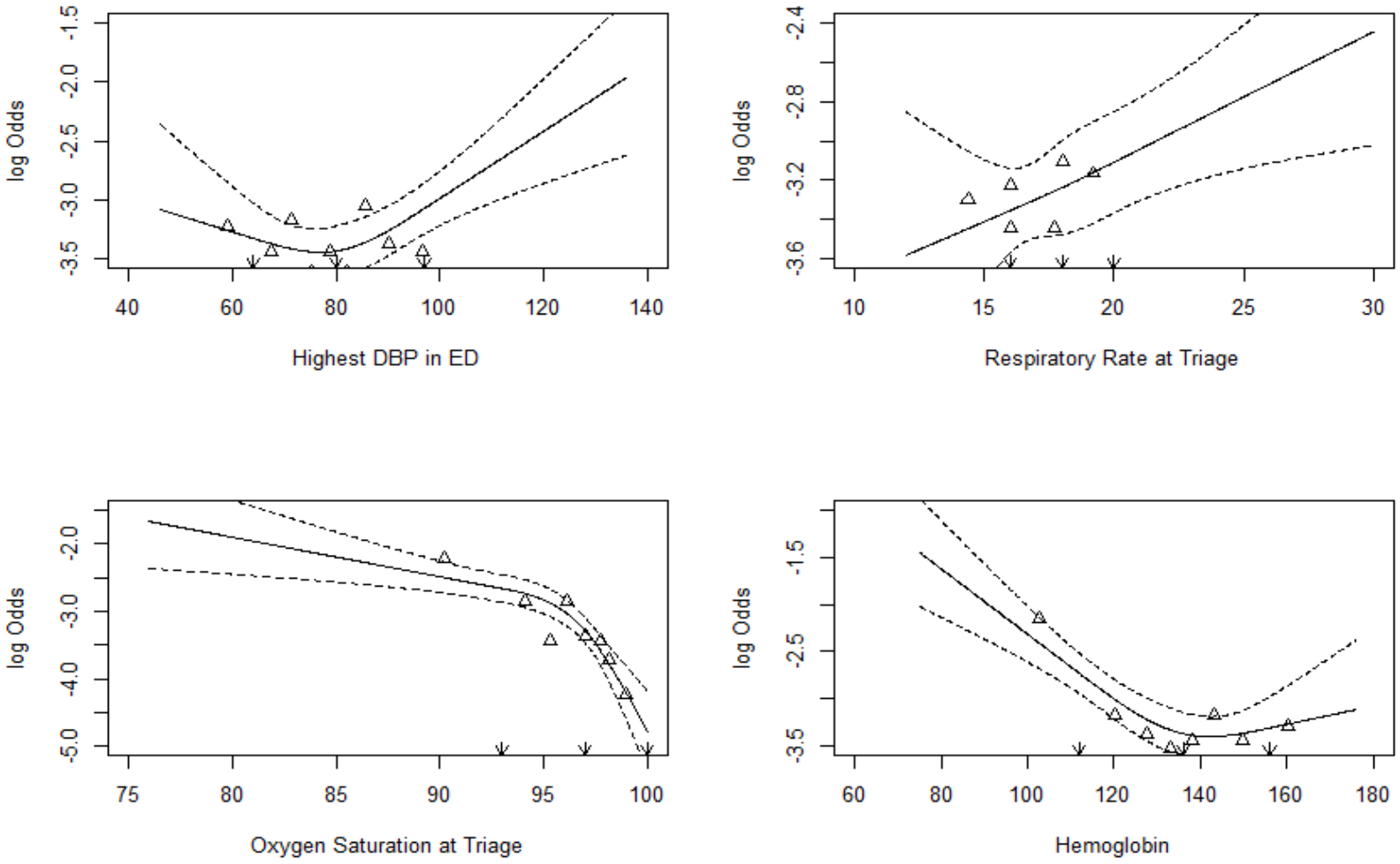


Figure 3. Functional association between continuous predictors and the probability (logit scale) of the serious outcomes (cont'd)

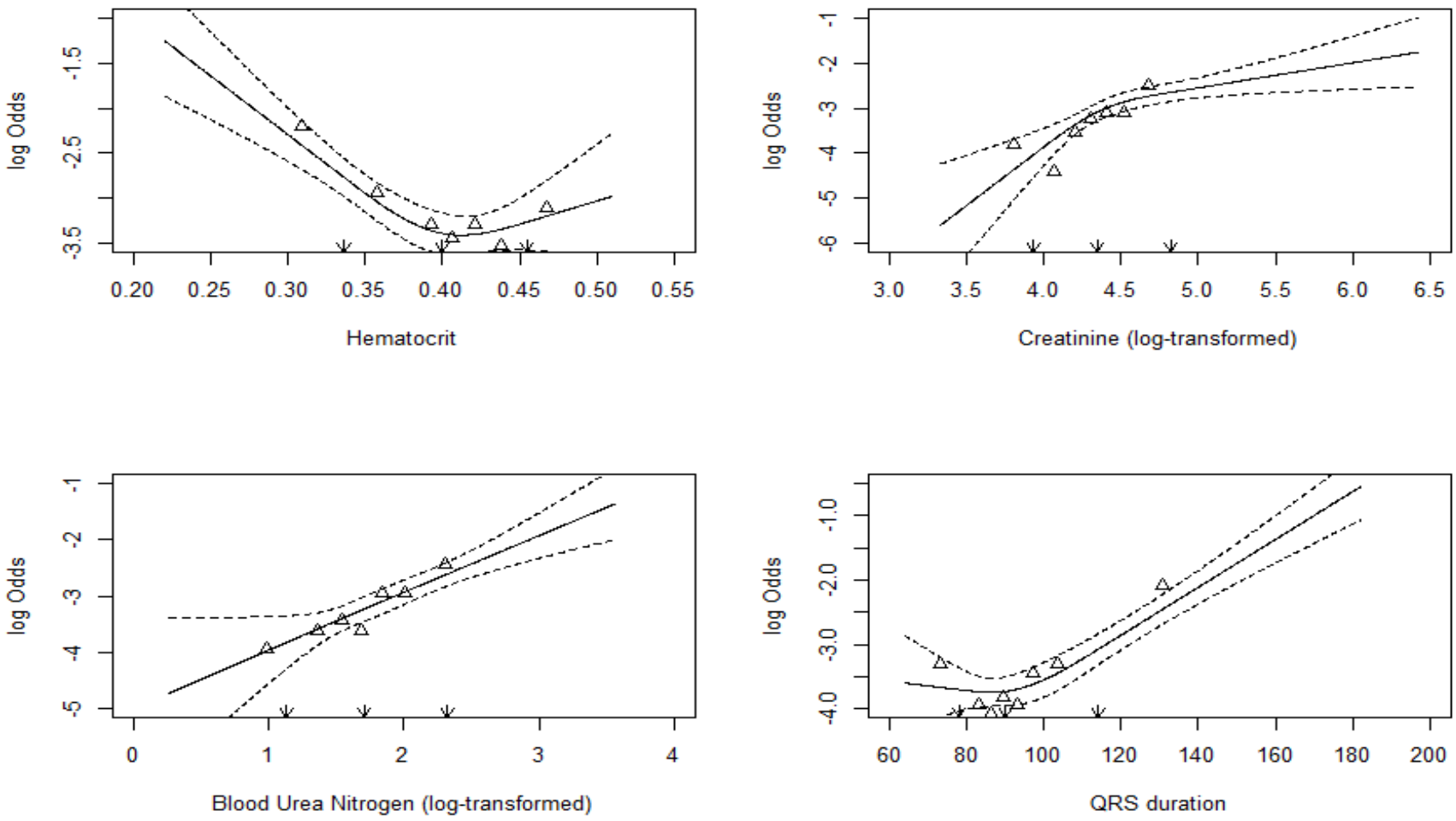


Figure 3. Functional association between continuous predictors and the probability (logit scale) of the serious outcomes (cont'd)

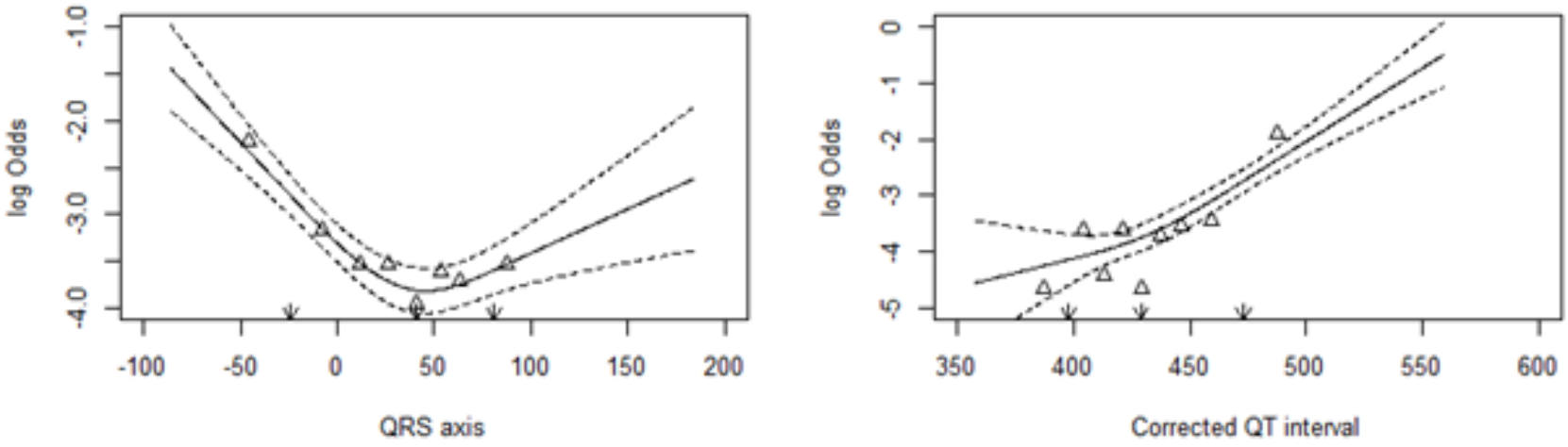


Figure 4. Graphs plotting test characteristics (sensitivity, specificity, Youden index) across various thresholds of the continuous variables

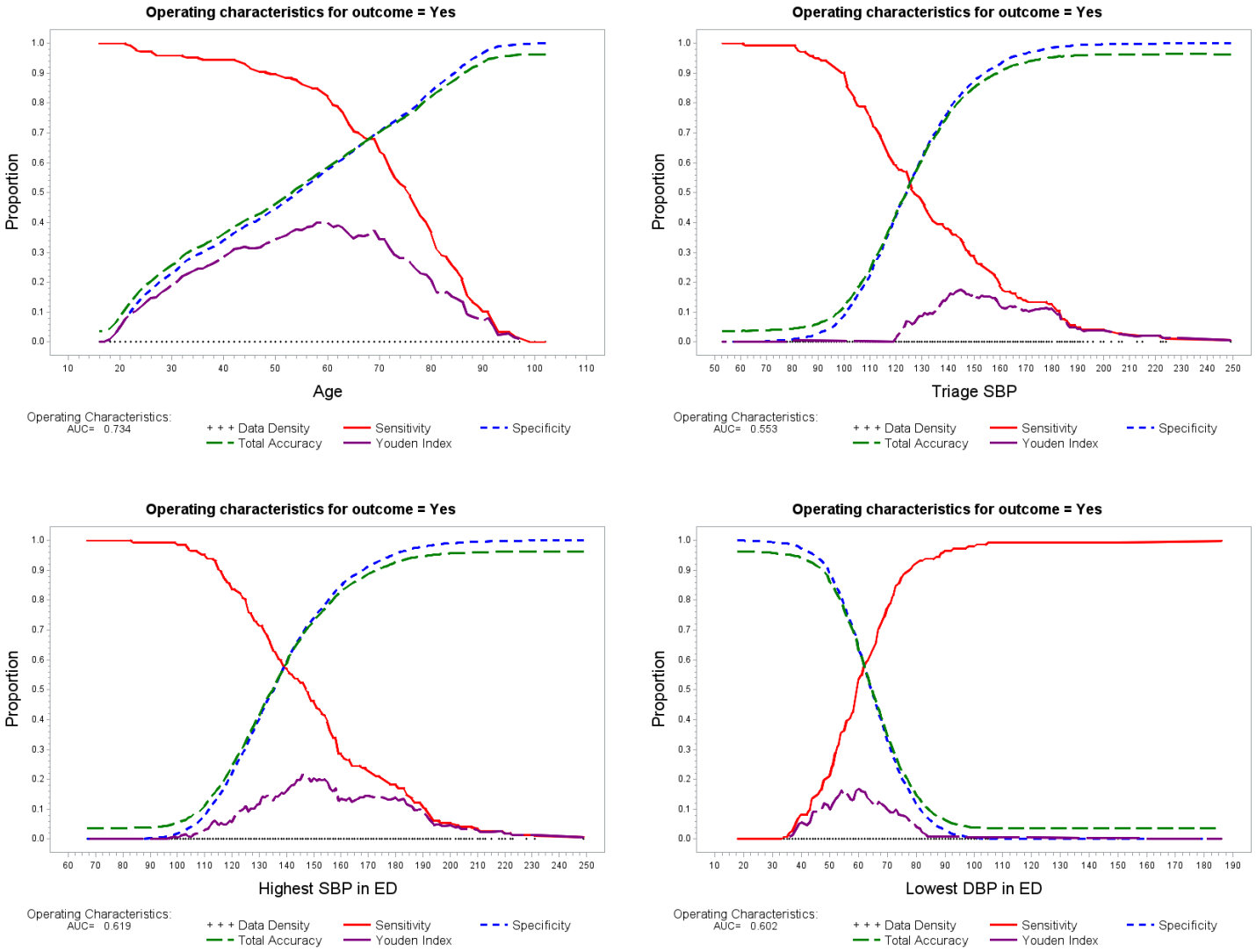


Figure 4. Graphs plotting test characteristics (sensitivity, specificity, Youden index) across various thresholds of the continuous variables (cont'd)

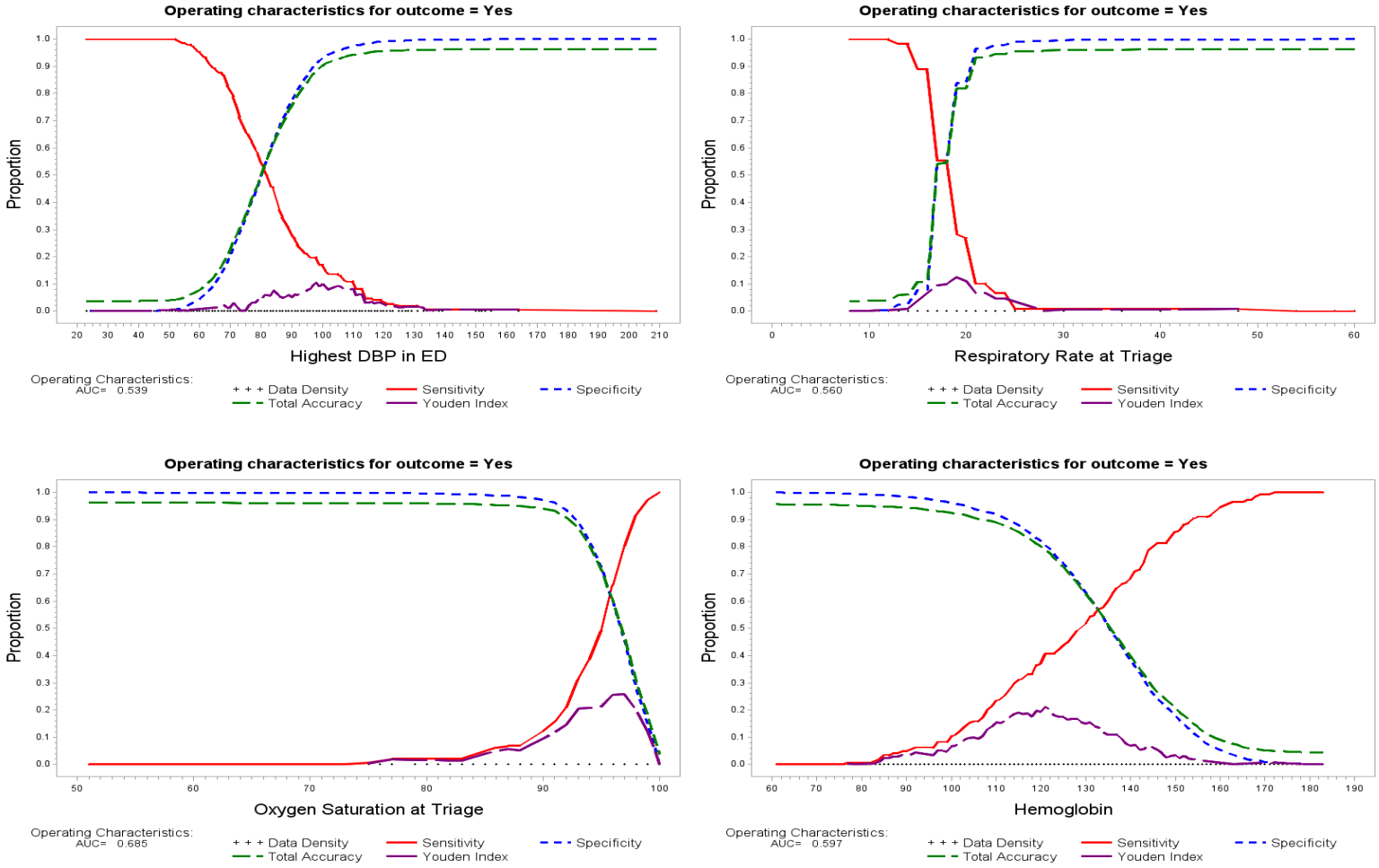


Figure 4. Graphs plotting test characteristics (sensitivity, specificity, Youden index) across various thresholds of the continuous variables (cont'd)

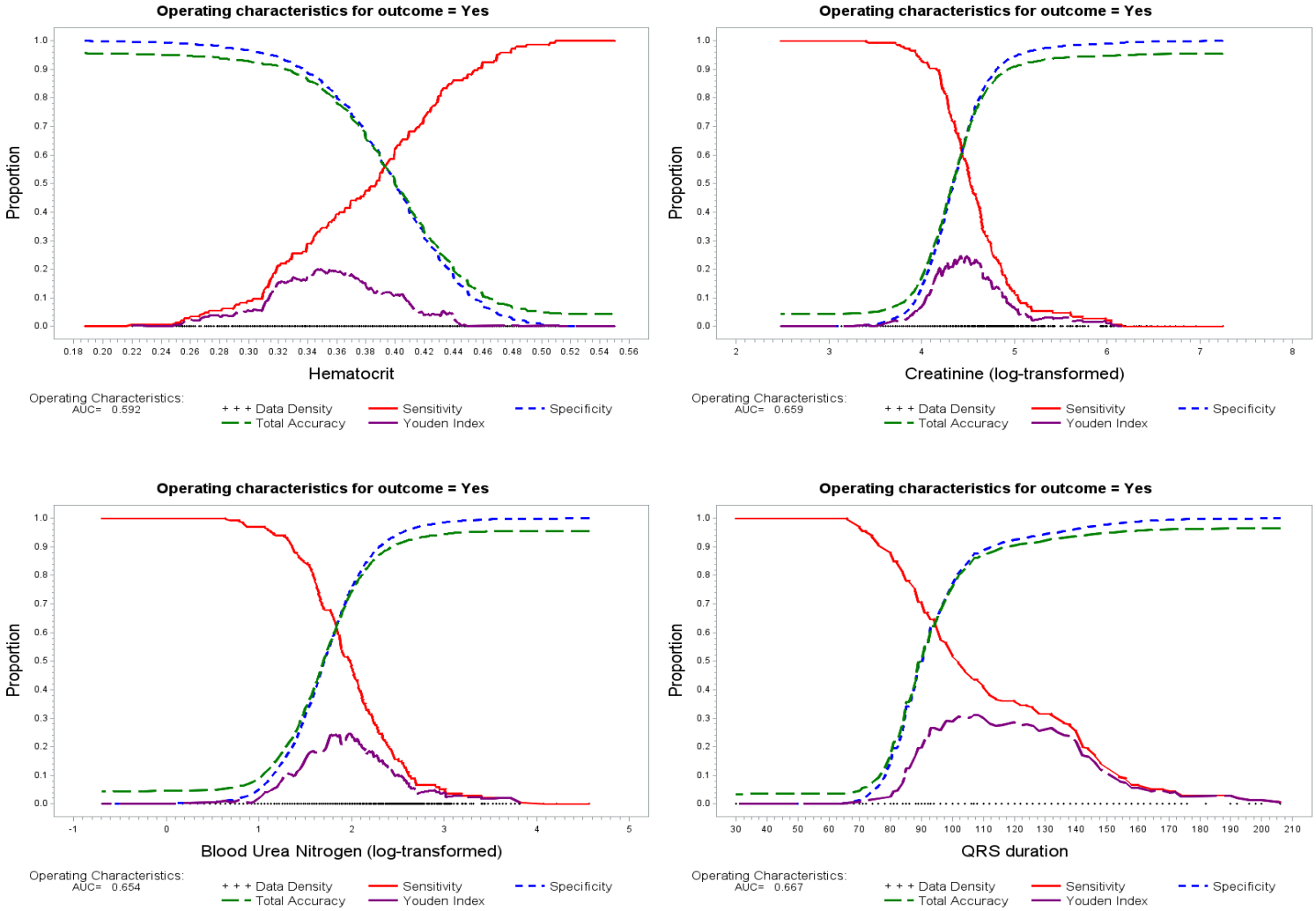
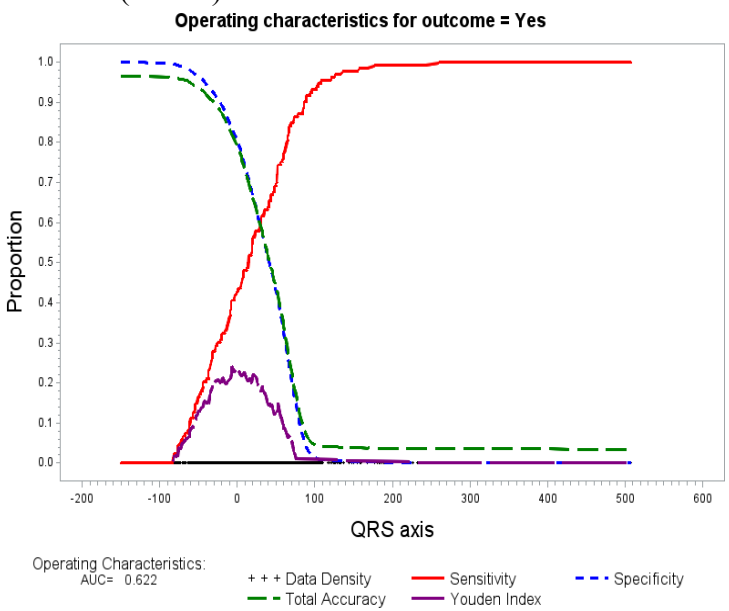


Figure 4. Graphs plotting test characteristics (sensitivity, specificity, Youden index) across various thresholds of the continuous variables (cont'd)



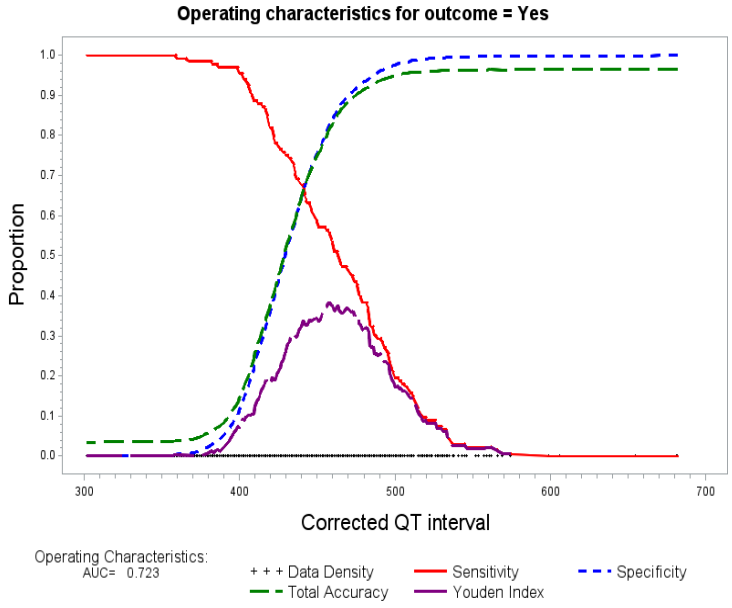


Table 9. Cut-points used to categorize continuous variables based on clinical and practical considerations

Variable	**Cut-point**	**Smallest Frequency (n, %)**
Age	>75 years	994 (23.4)
Extreme systolic blood pressure*	<90 or >180 mmHg in any SBP recording	494 (12.3)
Lowest diastolic blood pressure in the ED	<50 mmHg	379 (9.4)
Highest diastolic blood pressure in the ED	>110 mmHg	103 (2.6)
Respiratory rate at triage	>20 breaths/minute	121 (3.0)
Oxygen saturation at triage**	<89%	74 (1.8)
Hemoglobin**	<80 g/L	21 (0.5)
Hematocrit	<0.3	126 (3.1)
Creatinine (log-transformed)	>5 µmol/L	187 (4.6)
Blood Urea Nitroge (log-transformed)	>2.48 mmol/L	185 (4.6)
QRS duration	>130 milliseconds	228 (5.7)
QRS axis	<-30 or >110 milliseconds	350 (8.7)
Corrected QT interval	>480 milliseconds	287 (7.5)

*****Systolic blood pressure measurements include extremes in Triage SBP, highest SBP in ED, and lowest SBP in ED

** Excluded from subsequent multivariable model due to insufficient patients below the clinically-relevant cut-point

According to Figure 3, age, respiratory rate at triage, blood urea nitrogen (log-transformed), QRS duration, and corrected QT interval, demonstrated moderately positive linear associations with the risk of developing the event of interest. In addition, highest SBP in the ED, highest DBP in the ED and creatinine (log-transformed) demonstrated positive linear associations with the outcome of interest, but these relationships appeared to deviate from linearity. Nonetheless, the functional relationships supported cut-points that described patients with values exceeding a certain threshold being at higher risk of developing the event of interest. The cut-points defined using clinical rationale reflected this relationship as individuals with age >75 years, respiratory rate at triage >20 breaths/minutes, blood urea nitrogen creatinine (log-

transformed) >2.48 mmol/L, QRS duration >130 milliseconds, corrected QT interval >480 milliseconds, highest systolic blood pressure in ED >180 mmHg, highest diastolic blood pressure in ED >110 mmHg, and creatinine (log-transformed) >5 μmol/L, were considered at higher risk for developing the outcome of interest after the categorization procedure. With the exception of highest systolic and diastolic blood pressures in the ED, these clinically-specified cut-points corresponded to the 'optimal' statistical cut-points outlined in the graphs of Figure 4. The 'optimal' cut-point values identified by these graphs for the highest systolic and diastolic blood pressures in ED were not clinically-meaningful.

Lowest diastolic blood pressure in ED, oxygen saturation at triage, hemoglobin, hematocrit, demonstrated roughly negative relationships with the outcome of interest, although these relationships appeared to deviate from linearity. The functional relationships supported the use of specifying cut-points where patients with values below the threshold were considered at greater risk for developing the outcome of interest, which are consistent with our clinically-specified cut-points. The cut-points for lowest diastolic blood pressure in ED, oxygen saturation at triage, and hematocrit generally corresponded with values that optimized sensitivity, specificity, and Youden as illustrated in Figure 4. According to Figure 4, the 'optimal' statistical cut-point identified for hemoglobin was approximately 130 g/L, which was incompatible with clinical rationale and practice.

Triage systolic blood pressure and QRS axis demonstrated U-shaped functional associations with the risk of developing the event of interest (Figure 3). The observed functional relationships supported the use of clinically-specified cut-points that identified extreme values in these variables as being at higher risk in developing the outcome of interest. In particular, values of <90 or >180 mmHg for triage systolic blood pressure and < -30 or >110 milliseconds for QRS

axis were identified as cut-points. The graphs outlining the sensitivity, specificity, and Youden index at different thresholds (Figure 4) are not informative for these variables, since they only consider one value to be the 'optimal' statistical cut-point, which inherently fails to consider the U-shaped functional distribution of these variables.

We believed that having both triage systolic blood pressure (cut-point of <90 or >180 mmHg) and highest ED systolic blood pressure (cut-point of >180 mmHg) in the same prediction model would be impractical and difficult to interpret. We sought to create a pragmatic definition of systolic blood pressure that incorporated all the different systolic blood pressure measurements (triage systolic blood pressure, lowest ED systolic blood pressure, and highest ED systolic blood pressure). We therefore decided to designate that any systolic blood pressure measurement <90 or >180 mmHg observed in the ED was considered as an extreme systolic blood pressure. We believed this approach would make the definition of extreme systolic blood pressure more clinically-practical, interpretable, and user-friendly.

The frequency distributions of dichotomized variables were examined to ensure that an adequate number of patients were above and below the cut-point. Oxygen saturation at triage and hemoglobin were excluded from multivariable modeling because there were insufficient patients (<100) below the clinically-relevant cut-points. We investigated other cut-points for these variables but did not find clinically-sensible cut-points with an adequate distribution.

4.1.4 Variable Selection

The models that were identified in the backward elimination procedure of the multivariable logistic regression applied to each of the imputation datasets are presented in **Table 10**. The same variables were identified in each imputation dataset with the exception of

QRS axis (<-30 or >110) which was not selected in two imputation datasets. The final model (presented in section 4.2) was designated as the one that was selected most frequently across the ten imputation datasets.

Table 10. Backwards elimination applied to each imputation dataset to assess consistency of variable selection

	Imputation Dataset									
Variable Selected	1	2	3	4	5	6	7	8	9	10
History of heart disease	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Elevated troponin (>99%ile normal population)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MD suspected diagnosis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Predisposition to vasovagal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Extreme SBP (<90 or >180 mmHg)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
QRS duration (>130 milliseconds)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Corrected QT interval (>480 milliseconds)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
QRS axis (<-30 or >110)	✓		✓	✓	✓		✓	✓	✓	✓

4.2. Final Multivariable Logistic regression Model From Traditional Approach

The regression coefficients, standard errors, p-values, and corresponding odds ratios of the final multivariable logistic regression model are presented in **Table 11**. In the multivariable model, we observed that abnormal troponin, MD suspected diagnosis of cardiac syncope, extreme systolic blood pressure, and corrected QT interval were strongly associated with an increased occurrence of the outcome of interest, as demonstrated by the relatively large adjusted

odds ratios (OR >2.0). A history of heart disease, QRS duration, and abnormal QRS axis were also relatively strongly associated with an increase risk in the outcome of interest as demonstrated by adjusted odds ratios > 1.5.

Table 11. Conventional and shrinkage-adjusted estimates for the final syncope prediction model derived in the traditional approach

	Conventional Estimates		**Shrinkage-adjusted Estimates**		**Conventional and Shrinkage Estimates**	
Parameter	**Parameter Estimate**	**Odds Ratio (95% CI)**	**Parameter Estimate**	**Odds Ratio (95% CI)**	**Standard Error**	**p-value**
History of heart disease	0.56	1.75 (1.17 to 2.63)	0.51	1.67 (1.11 to 2.50)	0.21	0.007
Elevated troponin (>99%ile normal population)	1.24	3.47 (2.16 to 5.57)	1.13	3.10 (1.93 to 4.98)	0.24	<0.0001
MD suspected diagnosis - Vasovagal	-1.05	0.35 (0.20 to 0.61)	-0.95	0.39 (0.22 to 0.67)	0.28	0.0002
MD suspected diagnosis - Cardiac	1.16	3.20 (2.09 to 4.91)	1.06	2.88 (1.88 to 4.42)	0.22	<0.0001
Predisposition to vasovagal	-0.60	0.55 (0.32 to 0.94)	-0.55	0.58 (0.33 to 0.99)	0.28	0.030
Extreme systolic blood pressure (<90 or >180 mmHg)	0.82	2.26 (1.51 to 3.40)	0.74	2.10 (1.40 to 3.16)	0.21	<0.0001
QRS duration (>130 milliseconds)	0.68	1.98 (1.16 to 3.39)	0.62	1.86 (1.09 to 3.19)	0.27	0.013
Abnormal QRS axis (<-30 or >110)	0.52	1.68 (1.04 to 2.71)	0.47	1.60 (0.99 to 2.59)	0.24	0.033
Corrected QT interval (>480 milliseconds)	0.96	2.62 (1.64 to 4.18)	0.88	2.40 (1.50 to 3.83)	0.24	<0.0001

A MD suspected diagnosis of vasovagal syncope and a predisposition to vasovagal syncope displayed odds ratio values <1.0 with values of approximately 0.5, indicating that these predictors were strongly associated with a decreased risk of developing the event of interest. The direction and magnitude of the predictor-outcome relationships observed in the final multivariable logistic regression model are clinically-sensible.

To examine the sensitivity of the final model to the imputation procedure, the same predictors were also specified in a multivariable logistic regression model applied to the raw dataset and 354 observations were excluded due to having missing values in at least one predictor variables. No major differences in the model parameters and fit statistics were observed between the models estimated using the imputed and raw datasets (**Appendix 8**).

4.3 Model Performance

4.3.1 Outliers and Influential Points

Twenty eight (28) influential observations with DFBETA values exceeding 0.20 were identified. All 28 of these influential observations had the outcome of interest. There were some noticeable patterns in the profiles of these influential observations. Twelve (12) patients were identified as being influential with respect to the regression coefficient for MD suspected diagnosis; these observations were identified as having either vasovagal syncope or orthostatic/unknown syncope in the ED (both considered low-risk categories compared to cardiac syncope) but still suffered a serious outcome after ED disposition. Similarly, eleven (11) patients were identified as being influential with respect to the predisposition to vasovagal syncope coefficient, which can be explained by the fact that these patients had a serious outcome

even though they had a predisposition for vasovagal syncope (typically a low-risk condition for the development of serious outcomes). We investigated the characteristics and variable values for the influential observations, and reviewed the patient charts and case report forms for many of these influential observations. We did not find any evident data measurement, extraction, or entry errors. We deemed that these influential cases were clinically sensible and biologically plausible, so we retained these observations to be included in fitting the final model.

4.3.2 Overall Performance, Discrimination, and Calibration

A summary of the model performance measures are presented in **Table 12**. The Nagelkerke's R^2 was 29% which suggests that the model was able to adequately describe the observed data. A Nagelkerke R^2 value of 1 suggests that the model perfectly predicts the outcome; however the interpretation of pseudo- R^2 measures (such as Nagelkerke's R^2) cannot be easily compared between models[78]. The Brier score was 0.031, which indicates that the model was informative; Brier scores that approach 0 suggest more informative models[30]. The scaled Brier score was 0.13, which supports the notion that the model was informative.

Table 12. Summary of model performance measurements for the final multivariable logistic regression model derived in the traditional approach showing estimate and 95% bootstrap confidence interval

	Original	**Optimism**	**Optimism-corrected measure**
c-statistic	0.880 (0.0852 to 0.904)	0.015	0.865 (0.837 to 0.889)
Nagelkerke's R^2	0.288 (0.243 to 0.343)	0.04	0.248 (0.203 to 0.303)
Brier	0.031 (0.028 to 0.037)	-0.001	0.032 (0.027 to 0.036)
Shrinkage Factor	1.00	0.09	0.91
Hosmer-Lemeshow Goodness of Fit * p-value Chi-square Degrees of freedom	 0.11 13.1 7	-	-
Discrimination Slope*	0.14	-	-

*Measures not corrected for optimism

The ROC curve plotting the true positive rate against the false positive rate at various thresholds is presented in **Figure 5**. The ROC curve follows a smooth trajectory near the upper left-hand corner of the plot, which indicates good model discrimination. The corresponding c-statistic was 0.88, which supports the notion that the prediction model was able to discriminate well between patients who did and did not suffer the event of interest. Previous researchers have indicated that a c-statistic greater than approximately 0.8 has utility in predicting the responses of individual subjects[28]. A box plot comparing the estimated mean risk probabilities between patients who did and did not suffer a serious outcome is presented in **Figure 6**. The discrimination slope indicates that patients with the event of interest had a 14% higher average estimated probability of developing the event of interest compared to patients who did not have the event of interest.

Figure 5. ROC curve plotting true positive rate against false positive rate at various thresholds to illustrate discrimination of the traditional model

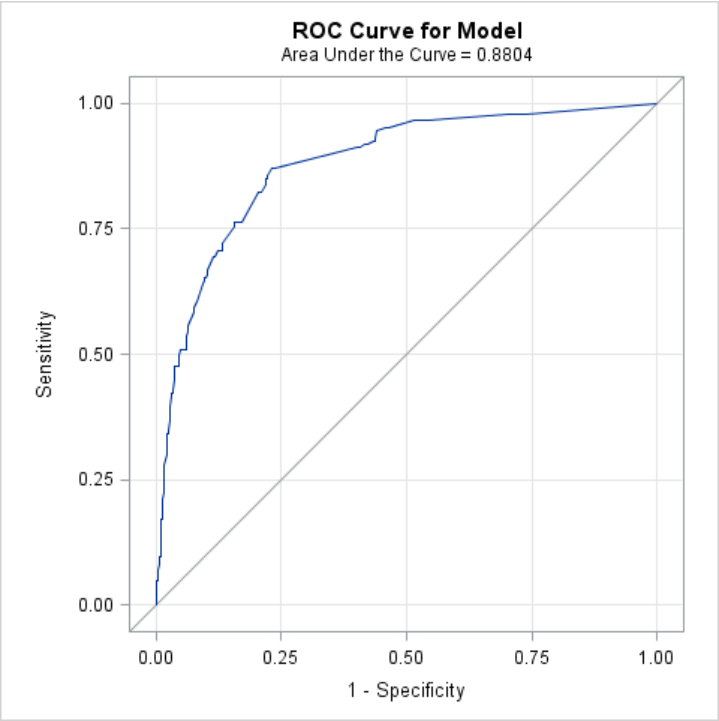


Figure 6. Discrimination box plot comparing the mean estimated risk probabilities between patients who did and did not have a serious outcome for the traditional model

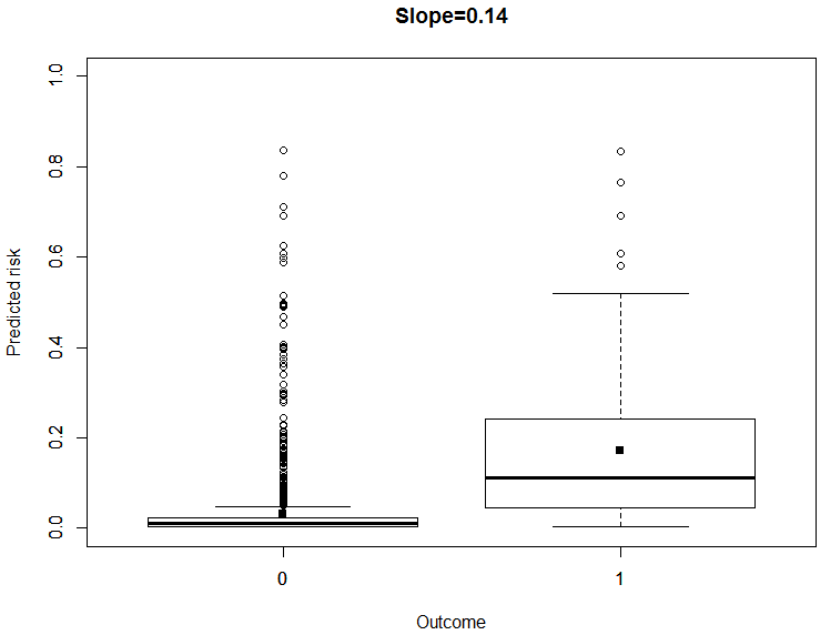
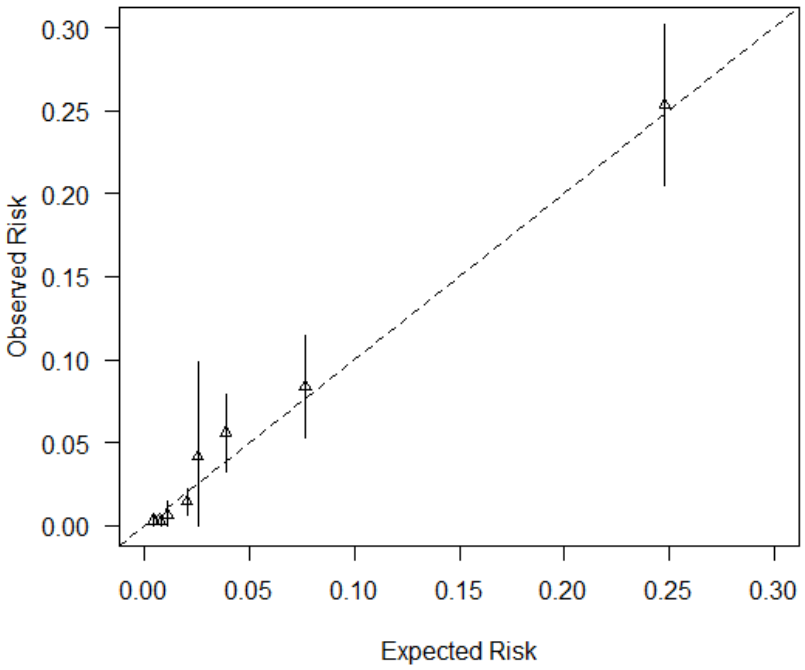


Figure 7 presents the calibration plot with the predicted probabilities plotted on the x-axis and the observed probabilities plotted on the y-axis; points along the dotted 45 degree line indicate perfect calibration. All the points in Figure 7 are situated around the 45 degree line, which suggests that the model had acceptable calibration. In addition, the Hosmer-Lemeshow goodness of fit test yielded a non-significant p-value of 0.11, which further suggests that observed and predicted probabilities are not significantly different. The points in Figure 7 are largely clustered at the lower risk levels, which can be attributed to the fact that the incidence of the outcome of interest is relatively low (3.6%).

Figure 7. Calibration plot comparing the predicted and observed risk estimates for the traditional model



4.4 Internal Validation and Shrinkage

The stepwise variable selection procedure used to derive the traditional model (section 2.3.4) was repeated in each of the 500 bootstrap samples during the internal validation process. The optimism and optimism-corrected model performance measures are presented in Table 12. The optimism for the c-statistic was calculated to be 0.015, representing approximately 1.7% of the original performance measure. The optimism for the R^2 statistic was 0.04, representing 16% of the original performance measure. The relatively low performance measurement optimism values suggest that the prediction model derived in the traditional approach was only slightly overfitted to the derivation dataset.

We calculated a shrinkage factor of 0.91 using the bootstrap internal validation procedure. The shrinkage factor indicate that 9% of the apparent model performance can be attributed to non-replicable 'noise' associated with statistical overfitting[74]. These results suggest that the prediction model derived in the traditional model may have been slightly overfitted to the derivation dataset. Shrinkage-adjusted regression coefficients and odds ratios were calculated and the results are presented in Table 11.

The results outlining the consistency of the stepwise variable selection procedure across the bootstrap samples are presented in **Table 13**. The eight variables included in the final multivariable logistic regression model in the traditional approach were most frequently selected in the bootstrap samples; they were selected in 57.2 to 100% of bootstrap samples. These results suggest that the traditional model was reasonably consistently selected across bootstrap samples; however, there was some inconsistency with at least one different variable selected in 57.2% of the samples, highlighting the inherent instability of the stepwise variable selection procedure used to generate the traditional model.

Table 13. Consistency of variable selection in 500 bootstrap samples of internal validation procedure

Variable	**Percent Selected**
MD suspected diagnosis*	100.0
Troponin (> 99%ile normal population)*	99.2
Corrected QT interval >480 milliseconds*	97.4
Extreme Systolic Blood Pressure (<90 or >180 mmHg)*	93.8
QRS duration >130 milliseconds*	82.4
History of heart disease*	73.8
Abnormal QRS axis (<-30 or >110)*	57.4
Predisposition to vasovagal*	57.2
Respiratory Rate at Triage >20/minute	45.4
Highest Diastolic Blood Pressure in ED>110 mmHg	44.2
Hematocrit <0.3	18.0
Highest heart rate in ED >110 beats/min	15.6
Sex	14.6
Left bundle branch block	11.6
Creatinine >150 μmol/L	11.6
Blood urea nitrogen >12 mmol/L	11.0
Lowest diastolic blood pressure in ED <50 mmHg	10.4
Age >75 years	9.6
Prodrome	8.0
History of vascular disease	7.8
Left ventricular hypertrophy	6.8

*Variable included in final multivariable model developed using traditional approach

4.5 Model Presentation: Point Scoring System

The final multivariable logistic regression model was translated into a point scoring system which is presented in **Table 14a**. The point scoring system assigned point values ranging from -2 to 2 for each of the variables included in the final prediction model. The total scores for the point scoring system can range from -3 to 11. The corresponding estimated risks associated with each point score value are presented in **Table 14b**, along with the suggested risk categories designated by physicians and subject-matter experts.

Table 14a. Point scoring system translated from the traditional model

Parameter	**Points**
History of heart disease	1
Elevated troponin (>99%ile normal population)	2
MD suspected diagnosis - Vasovagal	-2
MD suspected diagnosis - Cardiac	2
Predisposition to vasovagal	-1
Extreme systolic blood pressure (<90 or >180 mmHg)	2
QRS duration (>130 milliseconds)	1
Abnormal QRS axis (<-30 or >110)	1
Corrected QT interval (>480 milliseconds)	2

Table 14b. Expected risk for serious outcomes at each total point score level

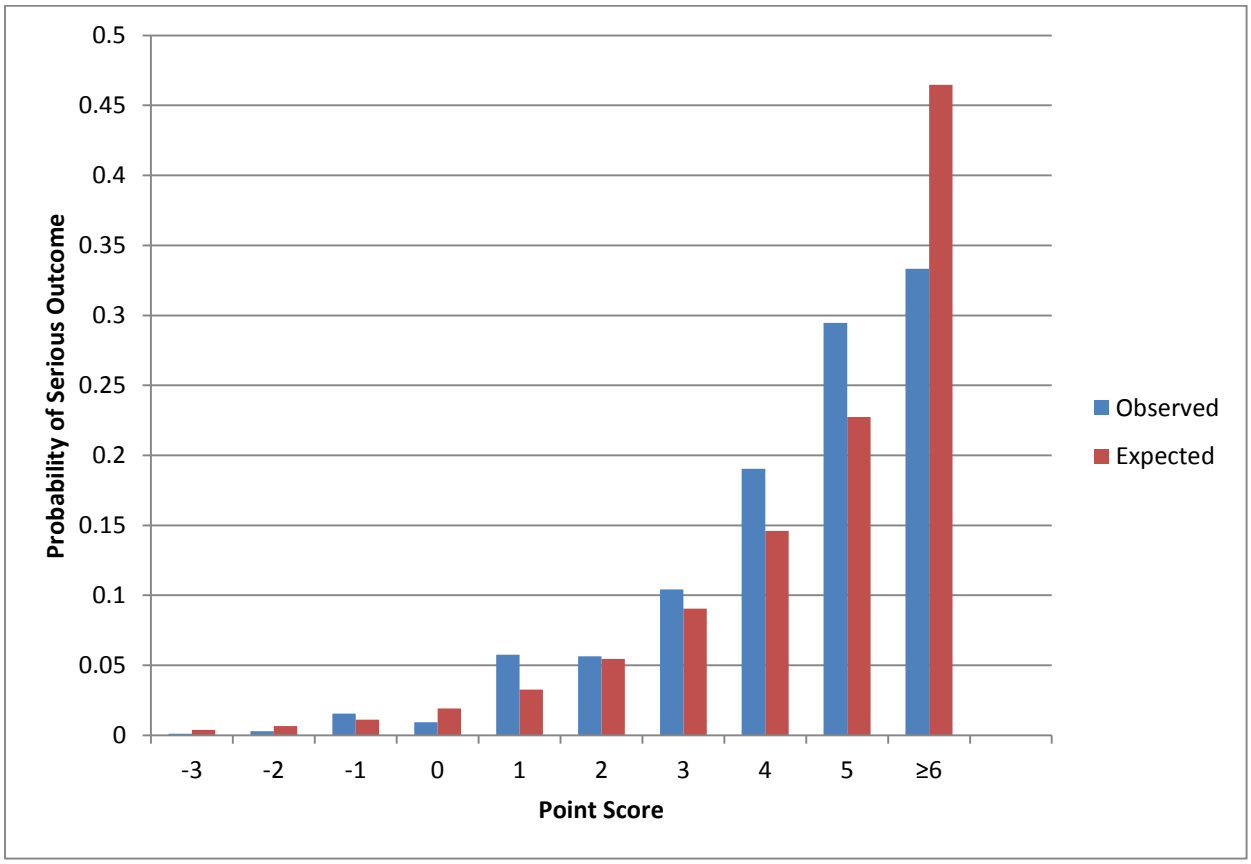
Total Score	Estimated Risk	Risk Category
-3	0.4%	Very Low
-2	0.7%	Very Low
-1	1.2%	Low
0	1.9%	Low
1	3.1%	Medium
2	5.1%	Medium
3	8.1%	Medium
4	12.9 %	High
5	19.7%	High
6	28.9%	Very High
7	40.3 %	Very High
8	52.8 %	Very High
9	65.0 %	Very High
10	75.5 %	Very High
11	83.6%	Very High

4.5.1 Calibration of Point Scoring System

A calibration plot comparing the expected and observed probabilities across the different plausible point score values is presented in **Figure 8**. Point score values exceeding 6 were grouped together since there were few patients in these categories. We observe that the expected

probabilities correspond well with the observed probabilities at the lower levels from point scores of -3 to approximately 3. Higher point score values (≥4) seemed to have a larger deviation between the expected and observed probabilities. The decreased calibration at these higher score values may be attributed to having a limited number of patients at these levels, reducing the ability to make reliable predictions in these categories.

Figure 8. Calibration plot comparing the expected and observed risk estimates across the various point score values



4.5.2 Sensitivity and Specificity at each point score

The cumulative sensitivities and specificities at each point score value is presented in **Table 15**. Based on clinical considerations, a point score of -1 was considered a feasible cut-point with a sensitivity of 97.7% and a specificity of 45.1%, and a corresponding observed and expected probability of having the serious outcome being 1.6% and 1.2% respectively.

Table 15. Classification performance of the point scoring system at varying point score thresholds

Point Score	**Sensitivity**	**Specificity**	**Expected probability of Serious Outcome***	**Observed probability of a Serious Outcome**	**Number of Patients at Score**
-3	1.000	0.000	0.4%	0.1%	900
-2	0.992	0.253	0.7%	0.3%	702
-1	0.977	0.451	1.2%	1.6%	385
0	0.932	0.558	1.9%	0.9%	744
1	0.879	0.765	3.1%	5.8%	261
2	0.765	0.835	5.1%	5.7%	248
3	0.659	0.901	8.2%	10.4%	192
4	0.508	0.949	12.9 %	19.1 %	84
5	0.386	0.968	19.7%	29.5%	95
6	0.174	0.970	28.9%	43.4%	23
7	0.099	0.991	40.3 %	25.0 %	32
8	0.038	0.998	52.8 %	28.6 %	7
9	0.023	0.999	65.0 %	40.0 %	5
10	0.008	1.000	75.5 %	50.0 %	2
11	0.000	1.000	83.6%	-	0
Missing**	-	-	-	4.2%	354

*Expected probability measurements are shrinkage-adjusted
**Missing on at least one predictor value of the final multivariable logistic regression model

CHAPTER 5: RESULTS: MODERN APPROACH

5.1 Initial Modelling Steps

5.1.1 Multiple Imputation

The imputation models used for the traditional and modern approaches used the same variables (described in section 4.1.1) but different software and estimation procedures. All dichotomous variables in the imputation models were modelled using one degree of freedom, while MD suspected diagnosis, the only 3-level categorical variable, was modelled using two degrees of freedom. Continuous variables were modelled using restricted cubic splines with 3 knots, except for triage respiratory rate, oxygen saturation, QRS duration, and the log transformations of creatinine and blood urea nitrogen. These continuous variables were modelled using linear trends since they had narrow marginal distributions that would not allow a reliable restricted cubic spline function with 3 knots to be fit.

The multiple correlation coefficient representing the proportion of variation explained by the variables in the imputation model ranged from an R^2 of 0.06 for palpitations prior to syncope to 0.96 for hemoglobin; variables with higher correlation (R^2 values approaching 1) generally have more reliable imputations since the other predictors provide meaningful information when estimating the missing values. Multiple imputation is still an appropriate procedure even if the R^2 value is small; however, imputations on variables with low R^2 indicate that the values generated from the imputation procedure may not be significantly better than a random guess[28]. The imputation-completed distributions of both continuous and categorical variables were compared to those in the original dataset (Appendix 7a and Appendix 7b). The mean, range, and standard deviations for imputation-completed continuous variables were compared to those in the original

dataset; there were no substantial differences observed in these measurements. There were no substantial differences in the frequency distributions of the categorical variables between the raw dataset and the imputed dataset.

5.1.2 Model Pre-specification

We received responses from all 6 physicians who were contacted to complete the survey to pre-specify the variables for the prediction model in the modern approach. Five of these surveys were fully completed, while one survey was partially completed. In the partially complete survey, the physician identified 11 important variables to be considered for inclusion in the prediction model instead of the instructed 15; the results from this incomplete survey were still included in the analysis.

The rankings from each physician's survey are presented in **Table 16**. Each value in the table represents the rank that the physician assigned to the corresponding variable; a value of '1' indicates that the variable was selected among the top 5 most important variables. Table 16 demonstrates that there was little consensus among the respondents with respect to the most important variables. Heart disease was the only variable identified by all respondents as being among the top 5 most important variables. Non-sinus rhythm, systolic blood pressure, age, MD suspected diagnosis, left bundle branch block, were identified as being among the top 5 most important predictor variables by at least 50% of the respondents. Other variables that were ranked as being important predictors include palpitations, heart rate, corrected QT interval, troponin, prodrome, and physical exertion prior to event. Diastolic blood pressure and QRS axis were two variables that were not selected by any of the respondents as being among the 15 most important predictors.

Table 16. Rankings obtained from survey of physicians to identify most important variables for syncope prediction model

	Physician respondents					
Predictor	**1**	**2**	**3**	**4**	**5**	**6**
Visit History						
Sex			1			12
Age	1			1	6	1
Systolic blood pressure		1	1	8	8	1
Diastolic blood pressure						
Heart rate		1	11	13	1	
Respiratory rate at triage		9	12			
Oxygen saturation at triage		10	13	14	9	
Medical History						
Cerebrovascular disease		6	10		10	14
Heart disease	1	1	1	1	1	1
Non-sinus rhythm*	1	1	7		1	1
Investigation						
Hemoglobin/Hematocrit		14	14	9		
Creatinine/Bun		15				9
Troponin	1			1		10
MD Assessment						
MD suspected diagnosis	7	1		1	1	
Witnessed by someone				12		11
Palpitations	9	8	1	7	12	
Physical exertion prior to event	8	7		6		15
Predisposition for vasovagal	11				15	6
Prodrome	10			1	11	13
Orthostatic symptoms				10		7
ECG						
QRS duration		13	9			8
QRS axis						
Corrected QT interval		12	6		7	1
Left bundle branch block	1		1		1	
Left ventricular hypertrophy	6		8		13	
Old ischemia		11	15	11	14	

*Non-sinus rhythm and heart disease were combined with heart disease based on clinical considerations after the survey was administered and completed

The ranked list of predictor variables is presented in **Table 17**. Using the summary of the survey responses along with clinical and subject-matter expertise, the predictors to be included in

the full model were identified. Due to the limited number of events available for the analysis, and the fact that some variables required multiple degrees of freedom to account for restricted cubic splines, only 10 variables could be included in the pre-specified model. This consisted of three continuous variables: systolic blood pressure, age, and corrected QT interval. The continuous variables represent a total of 6 degrees of freedom in the model, since each variable was modelled with a 3-knot restricted cubic spline. MD suspected diagnosis was included in the full model as a three-level categorical variable representing two degrees of freedom. Heart disease, palpitations, troponin, prodrome, physical exertion prior to event, and vascular disease represent were included as dichotomous variables, representing a total of 6 degrees of freedom. Overall, the fully specified model included a total of 14 degrees of freedom.

Table 17. List of predictor variables in order of importance based on mean ranking

Predictor	**Mean Ranking**
Heart disease*	1.0
Non-sinus rhythm*	4.5
Systolic blood pressure*	5.8
Age*	6.8
MD suspected diagnosis*	7.0
Left bundle branch block	8.5
Palpitations*	8.8
Corrected QT interval*	9.7
Heart rate	9.7
Troponin*	10.0
Prodrome*	11.2
Physical exertion prior to event*	11.3
Vascular disease*	12.0
Left ventricular hypertrophy	12.5
Sex	12.8
Oxygen saturation at triage	13.0
QRS duration	13.0
Predisposition for vasovagal	13.3
Orthostatic symptoms	13.5
Old ischemia	13.8
Respiratory rate at triage	14.2
Hemoglobin/Hematocrit	14.2
Witnessed by someone	14.5
Creatinine/Bun	14.7
Diastolic Blood Pressure	16.0
QRS axis	16.0

*included in final syncope prediction model

Non-sinus rhythm was identified as the second most important predictor based on the survey responses. After the survey was completed, the non-sinus rhythm variable was included in the definition of another variable (heart disease) based on clinical considerations. For this reason, it was not included in the final model as a separate variable. Left bundle branch block and heart rate were two variables that were not included in the full model despite their overall high ranking in the survey responses. Left bundle branch block was expected to contribute little

additional predictive value in a multivariable model that included heart disease and corrected QT interval, as left bundle branch block is usually associated with existing heart disease. We chose to include heart disease and corrected QT interval in the model, as opposed to left bundle branch block, because we believed that these variables provided a more comprehensive ability to capture the risk of developing the event of interest in adult ED syncope patients. We did not include heart rate in the model since we believed it was a potentially unreliable predictor that has been observed to fluctuate dramatically throughout a patient's ED visit.

5.1.3 Predictor Importance and Adjusting Model Complexity

The importance of each predictor in the presence of the other variables based on the partial model likelihood ratio chi-square statistic is presented graphically in **Figure 9**, and summarized in **Table 18**. The higher the chi-square value, the stronger the predictive ability of the variable, after all other variables have been accounted for in the model. In the full model, the MD suspected diagnosis was a powerful predictor with a likelihood ratio chi-square statistic of approximately 60. Corrected QT interval, troponin, and systolic blood pressure were also strong predictors, with likelihood ratio chi-square statistics of 35.4, 24.3, and 21.1 respectively. History of heart disease and age were moderately strong predictors, while vascular disease, palpitations prior to syncope, prodrome, and physical exertion were found to have negligible predictive ability.

Figure 9. Partial model likelihood ratio chi-square statistic of each predictor included in the modern model

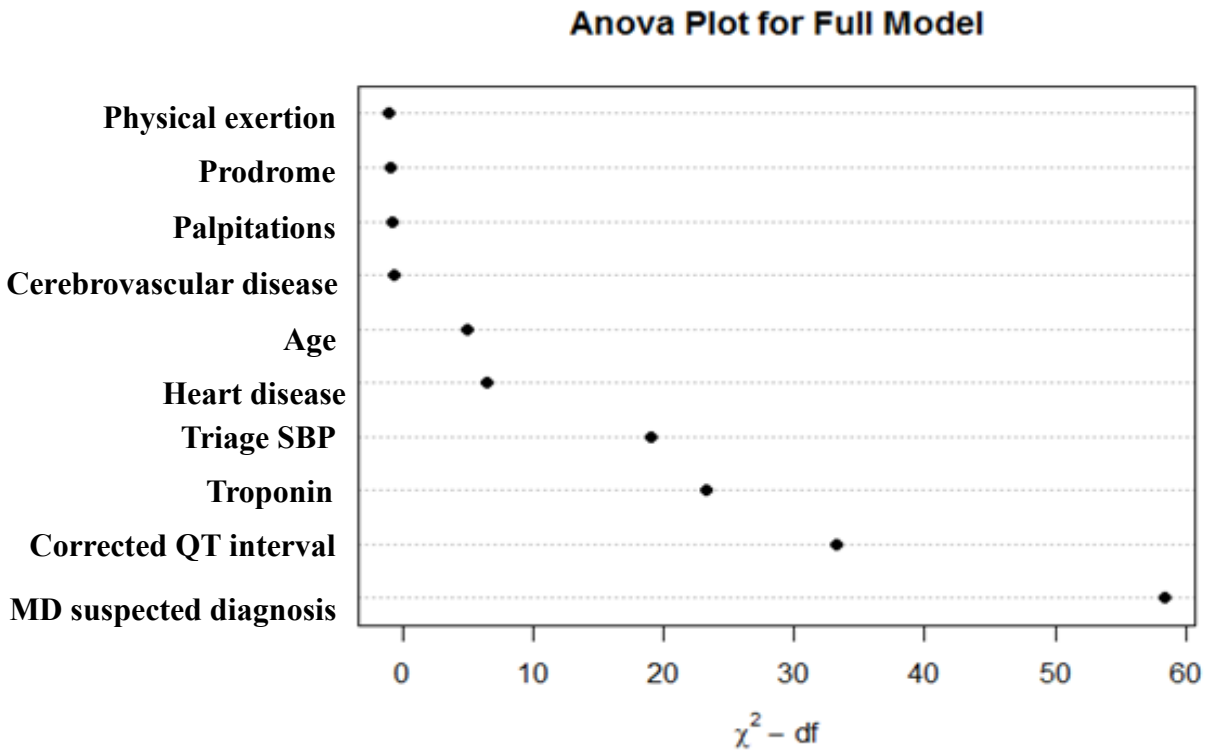


Table 18. Partial model likelihood ratio chi-square statistic of each predictor included in the modern model, including overall assessment of non-linearity

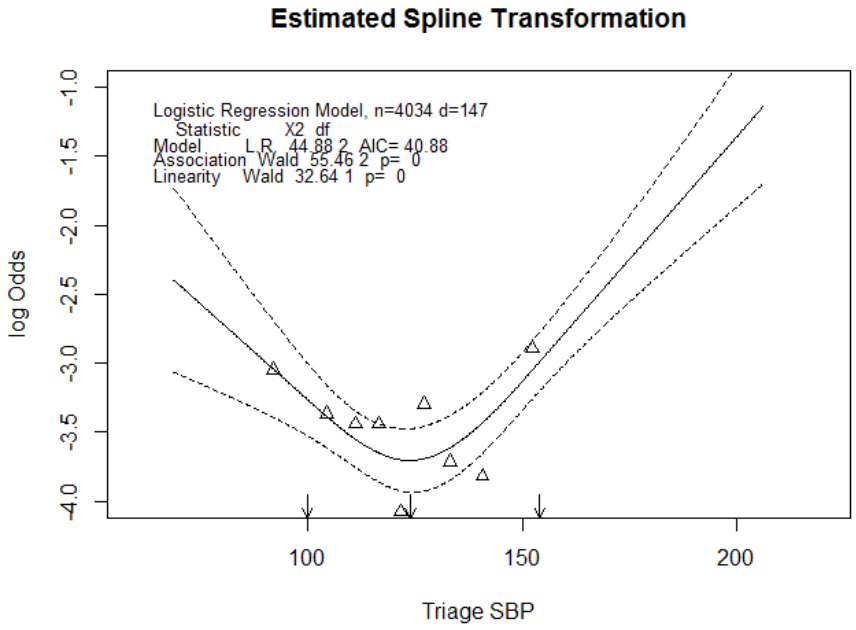
Variable	**Chi-square likelihood statistic**	**d.f.**	**p-value**
Age	6.97	1	0.03
Nonlinear	2.00	1	0.16
Triage Systolic Blood Pressure	21.05	2	<0.0001
Nonlinear	15.68	1	0.0001
Heart Disease	7.41	1	0.01
Vascular Disease	0.33	1	0.56
Troponin	24.30	1	<0.0001
MD diagnosis in the ED	60.38	2	<0.0001
Palpitations	0.22	1	0.64
Physical Exertion	0.00	1	0.95
Prodrome	0.12	1	0.73
Corrected QT interval	35.24	2	<0.0001
Nonlinear	4.09	1	0.04
Total Nonlinear	21.63	3	0.0001
Total	247.22	14	<0.0001

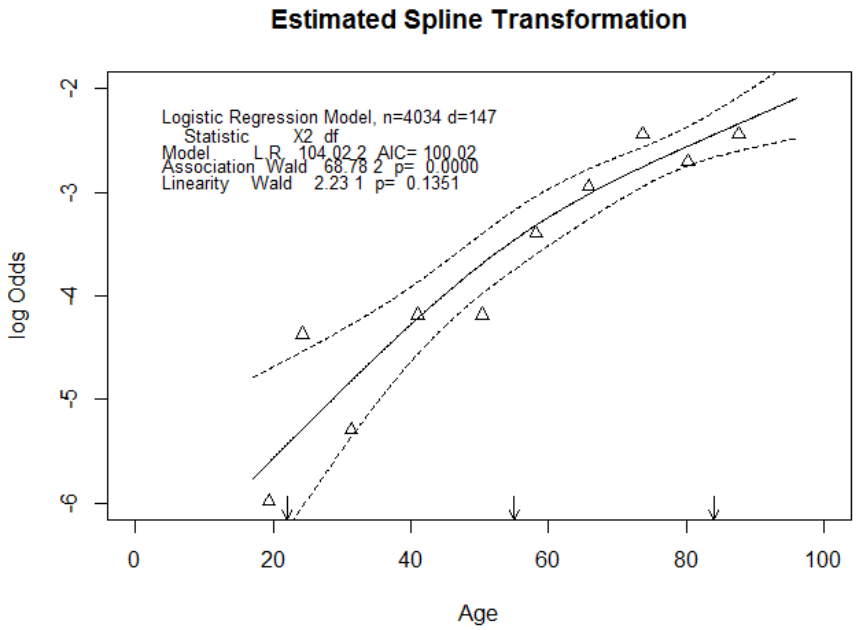
Based on the anova plot in Figure 9, all variables specified with more than one degree of freedom in the initial model (MD suspected diagnosis, corrected QT interval, systolic blood pressure, and age) were strong and important predictors, and we therefore decided to retain the level of complexity for these variables. Thus, MD suspected diagnosis was modelled as a 3-level categorical variable, and each of the continuous variables (systolic blood pressure, age, and corrected QT interval) using 3-knot restricted cubic splines. All variables were retained in the prediction model regardless of statistical significance.

5.1.4 Modelling Continuous Predictors

As shown in Table 18, there was substantial nonlinearity detected in the continuous variables (age, systolic blood pressure, and corrected QT interval) with an overall non-linearity likelihood ratio chi-square statistic of 21.6 (p=0.0001). In particular, strong non-linearity was detected in triage systolic blood pressure (likelihood ratio chi-square statistic of 15.7); substantial non-linearity was detected for corrected QT interval (chi-square statistic of 4.0), and mild non-linearity was detected for age (chi-square statistic of 2.0). The estimated restricted cubic spline transformations relating each continuous predictor to the probability of the outcome on the logit scale are presented in **Figure 10**. The statistics printed in the upper left-hand corner of the plots differ from those in Figure 9 and Table 18 since the plots only consider the bivariable association between the continuous predictor and the outcome, while the statistics in Figure 9 and Table 18 quantify associations adjusting for the other variables in the model.

Figure 10. Estimated relationship between continuous predictors (Triage SBP, Age, and Corrected QT interval) and the log odds of serious outcomes modelled using restricted cubic spline transformations





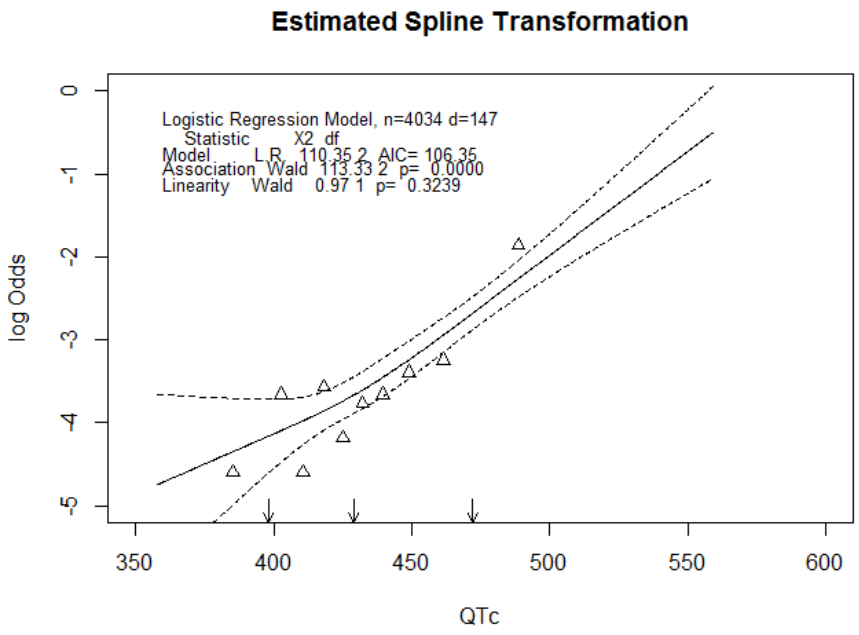


Figure 10 illustrates that systolic blood pressure and the log odds of a serious outcome display a U-shaped relationship with the risk being lowest at the median (approximately 124). The relationship with age appears monotonic with risk increasing initially but then leveling off at around 60 years. The relationship with corrected QT interval on the logit scale appears roughly linear; with the odds of a serious outcome increasing with increasing corrected QT interval.

5.2 Multivariable Logistic Regression Model from Modern Approach

The regression coefficients, standard errors, and p-values for the full model are presented in **Table 19**. The corresponding odds ratios and confidence intervals are also presented. For simplicity in this table, the continuous variables are presented as odds ratios comparing the upper and lower quartiles (75^{th} and 25^{th} percentiles), as it is otherwise difficult to interpret coefficients for restricted cubic splines. From Table 19, we observe that history of heart disease, elevated troponin, and MD suspected diagnosis have relatively large odds ratios, indicating that these

variables are strongly associated with the occurrence of severe outcomes even after adjustment for the other variables included in the prediction model. The odds ratios for the continuous variables should be interpreted with due consideration for the shape of the predictor-outcome relationship. For example, the odds ratio for SBP (0.97) indicates that a patient with an SBP value equal to the 75th percentile has 3% lower odds of an event than a patient with an SBP equal to the 25th percentile, however, Figure 10 indicates that the odds of an event does not increase linearly with SBP. The full equation describing the prediction model from the modern approach is a presented in **Appendix 9**. Compared to the regression coefficients listed in Table 19, the full equation includes the expanded terms for the restricted cubic splines.

Table 19. Conventional and shrinkage-adjusted estimates for the final syncope prediction model in the modern approach

	Conventional Estimates		**Shrinkage-adjusted Estimates**		
Predictor	**Parameter Estimate**	**Odds Ratio (95% CI)**	**Parameter Estimate**	**Odds Ratio (95% CI)**	**p-value**
History of heart disease	0.58	1.79 (1.18 to 2.71)	0.54	1.72 (1.14 to 2.62)	<0.01
History of vascular disease	-0.18	0.84 (0.46 to 1.53)	-0.17	0.85 (0.46 to 1.54)	0.56
Elevated troponin (>99%ile normal population)	1.19	3.28 (2.05 to 5.27)	1.12	3.06 (1.90 to 4.90)	<0.01
MD suspected diagnosis - Vasovagal	-1.20	0.30 (0.18 to 0.51)	-1.13	0.32 (0.19 to 0.56)	<0.01
MD suspected diagnosis - Cardiac	1.15	3.16 (2.05 to 4.86)	1.08	2.94 (2.03 to 4.57)	<0.01
Palpitations	-0.24	0.80 (0.29 to 2.12)	-0.22	0.80 (0.30 to 2.15)	0.64
Physical Exertion	0.01	1.01 (0.70 to 1.47)	0.01	1.01 (0.70 to 1.47)	0.95
Prodrome	-0.07	0.67 (0.63 to 1.39)	-0.07	1.06 (0.72 to 1.56)	0.73
Age (74 vs. 31)	0.90	2.44 (1.27 to 4.80)	0.84	2.32 (1.19 to 4.52)	0.03
Systolic Blood Pressure (138 vs. 110)	-0.03	0.97 (0.77 to 1.19)	-0.03	0.97 (0.78 to 1.20)	<0.01
Corrected QT interval (450 vs. 411)	0.23	1.25 (0.92 to 1.71)	0.21	1.24 (0.91 to 1.69)	<0.01

The full model was also fit on the raw dataset, without imputation for missing data, and the regression coefficients, standard errors, and p-values are presented in **Appendix 10**. The model was estimated using a complete-case analysis, including 3533 of the 4034 patients in the analyses. The regression coefficients were similar to those from the multiple imputation data in Table 19. The standard errors are larger for the model estimated from the raw dataset, which can

be attributed to the smaller effective sample size used in the non-imputed dataset. Overall, these results suggest that minimal instability was introduced in the imputation process.

5.3 Model Performance

5.3.1 Outliers and Influential Points

Forty four (44) influential observations with DFBETA values exceeding 0.20 were identified. Of these, 42 were observations from patients who suffered serious outcomes. There were some noticeable trends in the profile of the outlier observations. Seven patients were identified as being influential with respect to the age coefficient; these patients were young (under 35 years old) but ended up suffering a serious outcome, which is unexpected since older patients tend to have the highest risk. Four influential patients were identified as influential with respect to the systolic blood pressure coefficients: they had extreme blood pressures values (<85 or > 220 mmHg) and had suffered serious outcomes. These patients were likely identified as influential since they contributed to the elevated risk profiles observed at extreme systolic blood pressure levels. For the variable MD suspected risk, eight influential points were identified; these observations were identified as having vasovagal syncope in the ED (a typically low-risk category), but suffered an outcome following ED disposition nonetheless. Similarly to the investigation of outliers for the traditional model (section 4.3.1), we investigated the influential observations for evident data measurement, extraction, or entry errors, but did not identify any erroneous entries or observations that were biologically implausible. We thus retained these influential observations when estimating the final model.

5.3.2 Overall Performance, Discrimination, and Calibration

All model performance measures for the full model are presented in **Table 20**. The Nagelkerke's R^2 was 28.5% which implies that approximately 28.5% of the variation in the outcome can be explained by the model. The Brier score was 0.030 (scaled Brier was 0.13), which together indicate that the model was informative.

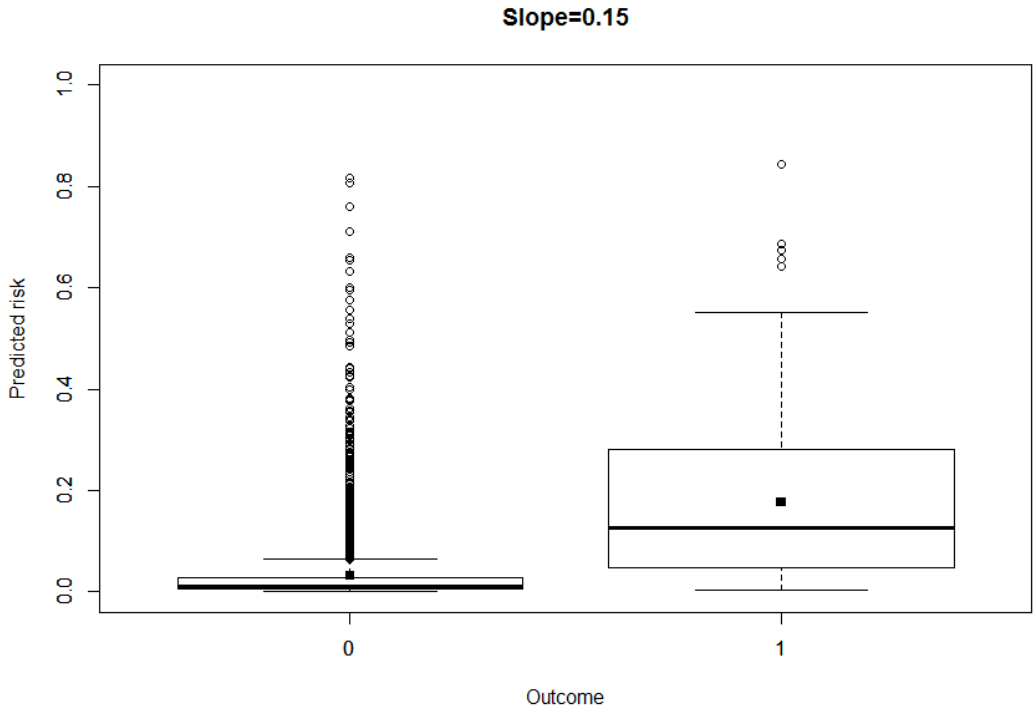
Table 20. Summary of model performance measurements for the final multivariable logistic regression model derived in the modern approach

	Full Model			Reduced Model (approximation of full model)		
	Original	Optimism	Optimism-corrected	Original	Optimism	Optimism-corrected
c-statistic (95% CI)	0.878 (0.850-0.903)	0.011	0.867 (0.839-0.892)	0.875	0.010	0.865
Nagelkerke's R^2 (95% CI)	0.285 (0.246-0.354)	0.026	0.259 (0.220-0.328)	0.280	0.022	0.258
Brier score (95% CI)	0.030 (0.026-0.034)	-0.001	0.031 (0.027-0.035)	0.031	-0.001	0.032
Shrinkage Factor	1.000	0.06	0.94	1.00	0.05	0.95
Hosmer-Lemeshow Goodness of Fit*		-	-	-	-	-
p-value	0.78					
Chi-square	4.8					
Degrees of freedom	7					
Discrimination Slope*	0.15	-	-	-	-	-

*Measures not corrected for optimism

The c-statistic was 0.88 and suggests that the prediction model discriminates well between those who do and do not suffer serious outcomes. **Figure 11** presents a box plot comparing the estimated mean probabilities between those patients who did and did not suffer serious outcomes. The calibration slope indicates that patients with a serious outcome had a 15% higher average estimated probability of developing a serious outcome, compared to patients who did not suffer a serious outcome.

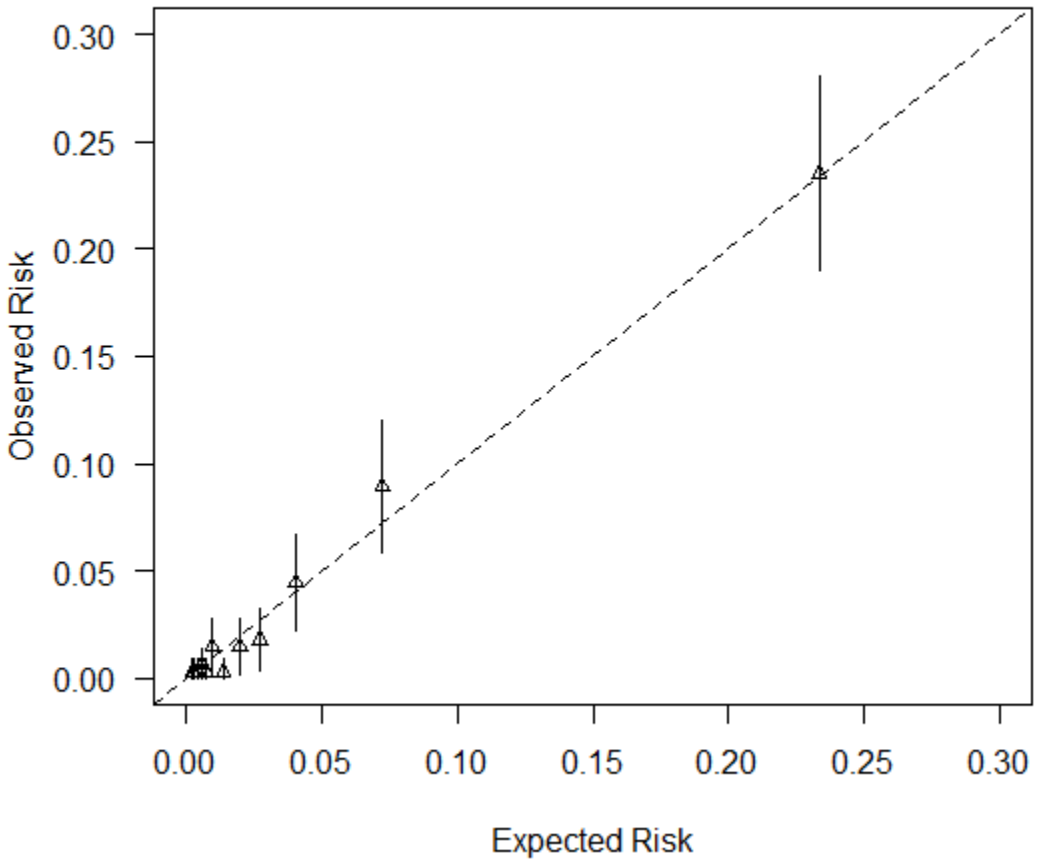
Figure 11. Discrimination box plot comparing the mean estimated risk probabilities between patients who did and did not have a serious outcome for the modern model



A calibration graph plotting the relationship between observed and expected risk is presented in **Figure 12**. In the calibration plot, it is observed that the risk deciles are clustered closely together at the lower risk estimates (e.g., at expected risk values <0.10). This uneven distribution of observed risk can be attributed to the relatively small incidence of the outcome of

interest. In the calibration plot, the observed values agree well with the expected probabilities, as shown by the points that are plotted closely along the 45 degree dashed line. The Hosmer-Lemeshow goodness-of-fit test yielded a p-value of 0.78, indicating acceptable model fit. Overall, the results suggested that the model is well calibrated, meaning that the model provided accurate risk estimates that adequately described the risk observed in our study population.

Figure 12. Calibration plot comparing the predicted and observed risk estimates of developing the event of interest for the modern model



5.4 Internal Validation and Shrinkage

Internal validation was carried out by evaluating the full model (described in section 5.2) in each of the 500 bootstrap samples. The original model was fully pre-specified and no automated variable selection procedure was performed; thus, the performance of the full model was assessed in each of the bootstrap samples. The optimism for each of the performance measures is listed in Table 20 in the "Full Model" section. The optimism for the c-statistic was calculated to be 0.011, representing approximately 1.3% of the original performance measure. The optimism for the R^2 statistic was 0.026, representing 9.1% of the original performance measure. The relatively low optimism values suggest that the prediction model derived in the modern approach was not substantially overfitted to the data, despite the relatively small number of events. Overall, the model was robust and we did not identify any major statistical issues in the prediction model.

A shrinkage factor of 0.94 was calculated through the bootstrap internal validation procedure, further suggesting that the prediction model was not substantially overfitted to our data. Shrinkage-adjusted regression coefficients and odds ratios were calculated and are presented in Table 19. It can be observed that the shrinkage-adjusted regression coefficients are less extreme and closer to the null value. The p-values for the shrinkage-adjusted regression coefficients are the same as the conventional estimates. The full equation describing the shrinkage-adjusted prediction model is presented in **Appendix 11.**

5.5 Model Approximation and Presentation

After applying the step-down procedure to the shrinkage-adjusted full model, we derived a reduced model that contained 5 predictors: triage systolic blood pressure, history of heart disease, troponin, MD suspected diagnosis, and corrected QT interval. **Table 21** lists the regression coefficients for the reduced model along with the corresponding odds ratios. Similar to Table 19, the continuous variables were presented by comparing the values of the upper to lower quartiles. We omit presenting any standard errors or p-values for this reduced model since statistical inferences based on automated variable selection procedure are difficult to interpret[28,30]. The equation describing the shrinkage-adjusted reduced model, along with the expanded terms for the restricted cubic splines, is presented in **Appendix 12**.

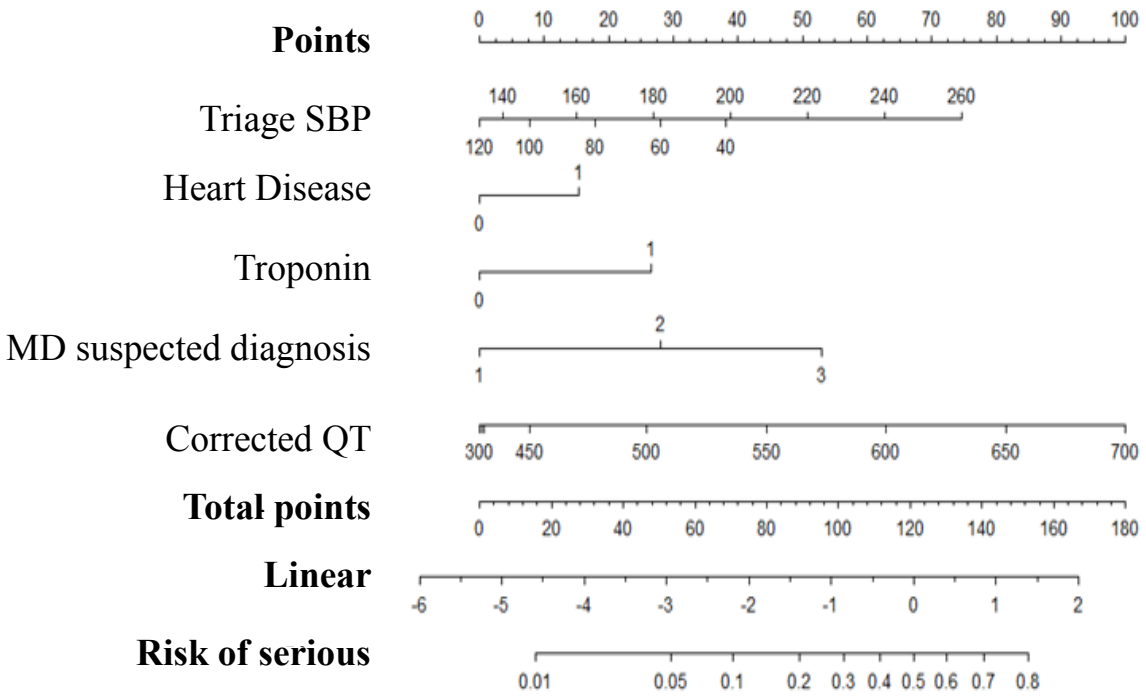
Table 21. Shrinkage-adjusted estimates for the reduced model derived from the modern approach using the step-down procedure

Parameter	**Estimate**	**Odds Ratio**
History of heart disease	0.67	1.95
Elevated troponin (>99%ile normal population)	1.16	3.18
Vasovagal MD suspected diagnosis	-1.22	0.30
Cardiac MD suspected diagnosis	1.09	2.98
Systolic Blood Pressure (138 vs. 110)	-0.01	0.99
Corrected QT interval (450 vs. 411)	0.29	1.34

The optimism and optimism-corrected model performance measures of the reduced model, calculated using bootstrap internal validation, are presented in Table 20. After accounting for the step-down procedure, the optimism-corrected performance measures of the reduced model were found to be similar to the full model. The shrinkage factor was also comparable between the full and reduced models. Overall, these results suggest that the reduce model provides an adequate approximation to the full model.

The reduced model relating triage SBP, history of heart disease, troponin, MD suspected diagnosis, and corrected QT interval duration to the risk of a serious outcome was presented in a nomogram (**Figure 13**). The nomogram allows clinicians to obtain the predicted risk of a serious outcome for an individual patient by calculating the number of points associated with each predictor variable, adding the points for all the individual variables to obtain the linear predictor, and then reading the predicted risk from the diagram. Alternatively, the prediction model (Appendix 12) can also be programmed as an equation in Microsoft Excel or presented in a web-based application to automatically calculate the risk estimates based on the specified predictor profile.

Figure 13. Nomogram to assess the risk of a patient developing the event of interest translated from the reduced modern model. (1= Yes, 0=No for dichotomous variables; for MD suspected risk, 1=Vasovagal, 2=Orthostatic/unknown, 3= cardiac)



CHAPTER 6: DISCUSSION

6.1 Summary of Findings

In this thesis, we developed clinical prediction models for emergency medicine physicians to identify syncope patients at risk of developing serious outcomes after ED disposition. The models were developed based on two different methodological approaches namely a "traditional" approach and a "modern" approach. The approaches differed in fundamental ways, including the way continuous predictors were modelled and the way that variables were selected to be included in the final prediction model.

In the traditional approach, a long list of candidate variables were identified based on review of the literature, clinical and subject-matter expertise, and data availability. As there were too many variables compared to the number of events, the variable list was reduced using a screening procedure that examined the bivariable tests of association between each variable and the outcome of interest; only variables found to be statistically significant were considered for multivariable modeling. In the multivariable model, continuous variables were categorized using cut-points based on clinical rationale and verified using statistical methods. Backward elimination, a stepwise variable selection procedure, was used to arrive at the final prediction model that included only variables statistically significant at the 5% level. Using this process, the traditional model identified eight (8) significant predictors: history of heart disease, elevated troponin, MD suspected diagnosis, predisposition to vasovagal, systolic blood pressure (<30 or >180 mmHg), QRS duration (>130 milliseconds), QRS axis (<-30 or >110), and corrected QT interval (>480 milliseconds). The model-fit was acceptable, as demonstrated by the optimism-corrected Nagalkerke's R^2 of 0.25, which can be interpreted roughly as 25% of the variation in

the outcome explained by the model. The model also had acceptable discrimination and calibration (optimism-corrected c-statistic=0.865, Hosmer-Lemeshow test p=0.11).

In the modern approach, the prediction model was developed by pre-specifying all predictors using clinical knowledge. Continuous predictors were modelled using flexible restricted cubic spline transformations. The final prediction model included history of heart disease, elevated troponin, MD suspected diagnosis, systolic blood pressure, corrected QT interval, history of vascular disease, palpitations, physical exertion, prodrome, and age. The goodness of fit was acceptable as illustrated by the optimism-corrected Nagelkerke's R^2 of 0.26. It also displayed good discrimination and calibration (optimism-corrected c-statistic=0.867, Hosmer-Lemeshow test p= 0.78). **Table 22** presents a summary of the results and performance measures for the two prediction models. We observe many similarities between the models. History of heart disease, elevated troponin, MD suspected diagnosis, systolic blood pressure, and corrected QT interval were included in both prediction models. These variables were also among the strongest predictors in both models; heart disease had a significant p-value of 0.0167 and 0.0065 for the traditional and modern models respectively, while the other variables common to both models had highly significant p-values <0.0001. Although we acknowledge that using p-values as a metric to compare the relative importance of predictors in models derived using stepwise variable selection may provide somewhat misleading results, we nonetheless believe that the variables common to both the modern and traditional models were among the strongest predictors in our prediction models[28,30]. The overall performance measures (i.e. Nagelkerke R^2 and Brier Score) and discrimination measures (i.e. c-statistic and discrimination slope) were comparable between the traditional and modern approach (Table 22). The modern model appeared to be better calibrated compared to the traditional model, as demonstrated by the

calibration plots (Figure 7 and Figure 12). The Hosmer-Lemeshow goodness of fit chi-square test statistic (X^2) in the traditional model was larger (X^2=13.1), indicating weaker calibration compared to the modern model (X^2=4.8). The superior calibration of the modern model may be attributed to the modelling of continuous predictors using flexible non-linear transformations and the inclusion of a larger number of predictors that may have served to improve the precision of the risk estimates. Using the shrinkage factor to assess overfitting, we found that the traditional model had a slightly higher level of statistical overfitting (9% or shrinkage factor of 0.91) compared to the modern model (6% or shrinkage factor of 0.94). These results can be interpreted as 9% of the model fit in the traditional approach is attributed non-replicable 'noise', while only 6% of the model fit in the modern approach is attributed to 'noise'[74]. It was expected that the model derived from the traditional approach would display more statistical overfitting, given that the traditional approach employed data-drive techniques to select variables included in the final prediction model, while the modern method was designed to avoid using data-driven techniques and ultimately overfitting.

Table 22. Overall comparison of prediction models: odds ratios, 95% confidence intervals, and performance measures for the models derived from the traditional versus the modern approach

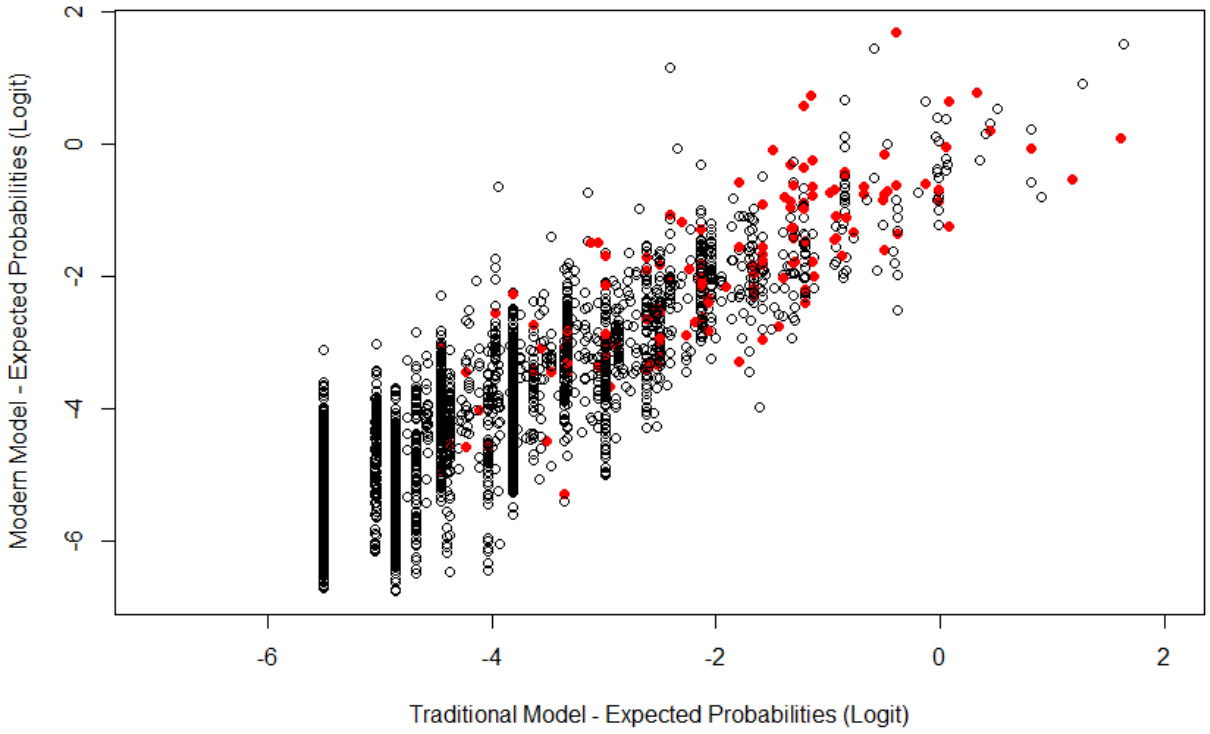
	Traditional Approach		**Modern Approach**	
Parameter Estimates	**Odds Ratio (95% CI)**	**p-value**	**Odds Ratio (95% CI)**	**p-value**
History of heart disease	1.7 (1.2 to 2.5)	0.0167	1.7 (1.1 to 2.6)	0.0065
Elevated troponin (>99%ile normal population)	3.5 (2.1 to 5.6)	<0.0001	3.1 (1.9 to 4.9)	<0.0001
Vasovagal MD suspected diagnosis	0.4 (0.2 to 0.7)	<0.0001	0.3 (0.2 to 0.6)	<0.0001
Cardiac MD suspected diagnosis	3.6 (2.3 to 5.5)	<0.0001	2.9 (2.0 to 4.6)	<0.0001
Predisposition to vasovagal	0.5 (0.3 to 0.9)	0.0203		
Systolic blood pressure*	2.3 (1.5 to 3.4)	<0.0001	1.0 (0.8 to 1.2)	<0.0001
QRS duration*	1.9 (1.1 to 3.2)	0.0046		
QRS axis*	1.7 (1.1 to 2.8)	0.0567		
Corrected QT interval*	2.9 (1.8 to 4.5)	<0.0001	1.2 (0.9 to 1.7)	<0.0001
History of cerebrovascular disease			0.9 (0.5 to 1.5)	0.5636
Palpitations			0.8 (0.3 to 2.2)	0.6387
Physical exertion			1.0 (0.7 to 1.5)	0.9461
Prodrome			1.1 (0.7 to 1.6)	0.7322
Age*			2.3 (1.2 to 4.5)	0.0307
Optimism-Corrected Performance Measures				
c-statistic	0.865 (0.837 to 0.889)		0.867 (0.839 to 0.892)	
Discrimination Slope	0.14		0.15	
Nagelkerke's R^2	0.248 (0.203 to 0.303)		0.259 (0.220 to 0.328)	
Brier Score	0.032 (0.027 to 0.036)		0.031 (0.027 to 0.035)	
Shrinkage Factor**	0.91		0.94	
Hosmer-Lemeshow Goodness of Fit**				
p-value	0.11		0.78	
Chi-square	13.1		4.8	
Degrees of freedom	7		7	

*estimates for continuous variables compare categorization levels in traditional approach, and values of inter-quartile ranges in the modern approach
** performance measure not adjusted for optimism

A graphical comparison of the predicted values from the traditional and modern models is presented in **Figure 14**. Each point represents a patient included in our study; black points represent patients who did not have the event of interest, while red points represent patients who had the event of interest. Values are plotted on the logit scale, with the 45 degree line indicating

perfect agreement between the traditional and modern models. According to Figure 14, we observe a strong positive association between the predicted probabilities of the two models. This figure also illustrates that the proportion of patients with the event of interest increases as predicted risk increases; these findings are expected as both the traditional and modern model displayed acceptable model calibration in their individual assessments of model performance. Of particular interest, the traditional model had discrete levels of predicted risk as demonstrated by the distinct vertical 'bands' of points. On the other hand, the modern model produced a wide range of possible predicted risk estimates, as demonstrated by the lack of horizontal 'bands' of points. These findings are sensible given that the traditional model can only take a combination of categorical variables as the input, resulting in discrete levels of plausible predicted risk. Meanwhile, the modern approach can take both categorical and continuous variables as the input, which can have implications in generating more realistic and precise risk estimates. **Appendix 13** presents the same figure except the expected values are plotted on the probability scale (range from 0 to 1) for ease of interpretation.

Figure 14. Overall comparison of the predicted probabilities from the traditional and modern models (logit scale)



It is important to note that, in this thesis, we were able to assess and compare only the internally-validated performance of the models. We were unable to definitively compare accuracy of the two models as the only way to robustly assess and compare the true performance of a prediction model is through external validation (e.g., using data that were not used in the derivation of the models).

6.2 Interpretation in Context of Other Studies

Many of the variables identified in our prediction models have also been identified in other syncope prediction models. A summary of other clinical decision tools estimating the risk of serious outcomes among adult ED syncope patients was presented in Table 1. It is interesting to compare the predictors that were included in our traditional and modern models relative to these competing models. A history of heart/cardiovascular disease was identified as a predictor in 8 of the 9 existing syncope prediction model studies, although the definitions of heart/cardiovascular disease were fairly heterogeneous[8,10,17,54-58]. Systolic blood pressure (defined as extreme values >160 or < 90 mmHg) was explicitly identified as important predictors in the San Francisco Syncope Rule and the Syncope Risk Scale developed by Sun *et al.* [10,58]. In addition, the Boston Syncope Criteria developed by Grossman *et al.* inherently included the systolic blood pressure measurement in the broad category of "persistent abnormal vital signs in ED" [10,17,58]. Troponin was identified as a predictor in one other existing syncope prediction model[58].

Many of the existing syncope prediction models identified an abnormal ECG as a predictor of serious outcomes[8,10,17,54-57]. There have been some criticisms that an abnormal ECG is a non-specific variable that is evaluated subjectively and heterogeneously in the clinical setting[13,79]. In the development of our prediction models, we sought to identify specific elements of an abnormal ECG, which could be interpreted more objectively and reliably. QRS duration, QRS axis, and Corrected QT interval were three variables included in our prediction models and can be considered synonymous with an abnormal ECG.

No other existing syncope prediction models included a subjective variable such as MD suspected risk. We speculate that other syncope prediction models were derived with the

intention of including primarily objectively-assessed variables and limit the role of physician discretion, ultimately providing more objectively-derived risk measurements. However, recent studies have demonstrated that the current syncope prediction tools did not perform any better than clinical judgement in identifying serious outcomes after ED disposition, suggesting that clinical judgement remains a crucial component in ascertaining the risk of serious outcomes in syncope patients[80]. In this study, we thus sought to develop a prediction model that captured elements of both clinical judgements and objective measurements, since we believed it would develop a more accurate, robust, and pragmatic decision tool. However, we do acknowledge potential concerns regarding the consistency and generalizability of these subjective variables when evaluated by physicians of varying experience across different centres. In addition, it can be argued that the inclusion of subjective variables continues to place too much diagnostic responsibility on the physician. Nonetheless, there have been other successful prediction models that have demonstrated the utility of including a subjectively-assessed variable based on the clinician's impression after clinical examination. In particular, one of the strongest predictors in the Wells' Criteria for pulmonary embolism, an internationally adopted prediction model used in assessing the risk of pulmonary embolism, contains a subjective measurement outlining whether the physician believes that pulmonary embolism is the most likely diagnosis[69]. Similarly, one of the strongest predictors in the Wells' Criteria for deep vein thrombosis also contains a subjective measurement indicating whether the physician believes that a deep vein thrombosis is the most likely diagnosis[81].

Predisposition to vasovagal syncope, which was included in the traditional, but not the modern model, had also been included in one other existing prediction model[57]. Several variables were included in the modern model, but not the traditional model: age, prodrome,

palpitations prior to syncope and physical exertion. Whereas age is a commonly included predictor in other syncope prediction models[8,54-56,58], prodrome was identified as a component in only two other prediction models, while palpitations prior to syncope and physical exertion were identified as independent predictors in one other study[55,57]. It was not surprising that the variables included in the modern model were consistent with other existing syncope prediction models, as the variables for the modern model were specified based on clinical knowledge, which was inevitably influenced by previous studies.

Overall, the variables included in the prediction models derived in the present study were consistent with the predictors included in other prediction models confirming the face validity of the derived models. These findings provide preliminary support that our prediction models display reproducible results that are consistent with other studies conducted at other centres.

The model performance measures of the existing syncope clinical prediction tools were presented in Table 1. As outlined in section 1.6.5, the evaluation and reporting of the model performance measures of existing syncope prediction model were largely sub-optimal; thus, it was difficult to compare the statistical performance of our prediction models compared to other existing models. The c-statistic was the most commonly-reported performance measures. Among those existing prediction models that reported the c-statistic, it ranged from 0.61 in the Syncope Risk Score[58] to 0.897 in OESIL[55]. We calculated an optimism-corrected c-statistic of 0.87 in both our traditional and modern models, suggesting that our model had good discrimination compared to other syncope prediction models. However, we exercise caution when comparing the c-statistic among the various studies, due to the heterogeneity in populations and definitions of outcomes among the studies and the fact that the reported statistics were often calculated from the derivation datasets, without adjusting for overfitting[28].

6.3 Methodological Considerations

The methodology in the traditional and modern approach differed in two fundamental aspects; namely, modelling of continuous predictors and choosing the variables to be included in each prediction model. The traditional approach produced a model that was easy to interpret and was presented as a simple point scoring system; meanwhile, the modern approach produced a more complex model that was more difficult to interpret and necessitated presentation as a nomogram, or electronic formula.

With regards to the modelling of continuous predictors in the traditional approach, variables were dichotomized using discriminative cut-points. The categorization was straightforward to implement, incorporated clinical judgement where possible, and led to easy to understand and interpret risk factors. However, many of the continuous variables do not have well-reported cut-points, and the assignment of cut-points may have been somewhat arbitrary for such variables. Thus, we have concerns whether the categorized continuous variables adequately approximate the true predictor-outcome relationship; we also have reservations about whether the cut-points will generalize to other populations, especially for those variables that incorporated elements of data-driven statistical techniques to define the cut-point values.

In the modern approach, continuous variables were modelled using flexible non-linear transformations. By using restricted cubic spline transformations, we were able to model the predictor-outcome relationship without having to make any major assumptions about the functional relationship with these variables. However, the functions describing the spline transformations were complex, difficult to interpret, and challenging to present clearly. We believe that modelling the continuous variables using flexible non-linear transformations, rather than categorization, allows for more accurate and biologically plausible modelling of the

predictor-outcome relationship, ultimately resulting in a prediction model with better predictive performance. Our ability to assess the added value associated with modelling continuous variables using non-linear transformations, as opposed to categorization, was complicated by having different variables included in the traditional and modern models.

The variables included in the traditional model were largely based on statistical tests of association; namely, bivariable screening and stepwise variable selection. After developing a list of candidate predictors based on clinical rationale and data availability, the specification of the traditional model was a simplistic process since it selected variables based solely on calculated p-values. However, we have concerns about the generalizability of the model derived from this approach since such data-driven approaches in prediction model development have been documented to be susceptible to Type I error and statistical overfitting[28,30].

In comparison with the traditional model, the specification of the modern model was a more complex process. The survey responses used to identify the most important predictors revealed that there was little consensus among clinicians about the most important risk factors for serious outcomes after syncope. This complicates the pre-specification of the prediction model. Some of the variables included in the model were identified by only a few respondents. Thus, we acknowledge that the variables included in the final modern model may be sensitive to the specific respondents chosen to complete the survey. Nonetheless, our subject-matter experts were successful in identifying strong and important predictors in our dataset, such as MD suspected diagnosis, triage systolic blood pressure, Corrected QT interval, troponin, history of heart disease, and age, using solely clinical knowledge. While pre-specification of predictors can be beneficial in protecting against statistical overfitting and Type I error, this may not be possible unless the health condition under study is homogeneous and/or well-studied; pre-selecting

variables for syncope which is a heterogeneous and understudied condition seemed to be more challenging. Ultimately, to determine whether the pre-specification of variables reduced the effect of statistical overfitting and Type I error requires a subsequent external validation study. Thus, the present thesis cannot address the important question of how pre-specification contributes to the reliability and performance of a prediction model.

Multiple imputation and internal validation were used in both the traditional and modern approaches. Multiple imputation was conducted under the assumption that the missing data mechanism was missing at random (MAR). The imputation process enabled us to retain all the patients in the analysis and preserve statistical power. We observed no major differences between the imputation procedures implemented in the traditional model (using SAS Proc MI) and the modern model (using the R-function aregImpute). Internal validation was a useful tool in assessing the robustness of the prediction models. In particular, the internal validation procedure was crucial in identifying any statistical issues in the model-fitting process, assessing the stability of the stepwise variable selection procedures, generating optimism-corrected performance measures to provide more realistic estimates, and providing insight into the degree of overfitting attributed to the model development process.

6.4 Limitations

A major limitation is the extent of missing data on the troponin variable which ended up being selected as an important predictor in both models. Troponin assays were complete in only 48% of our study patients. To avoid using an automated imputation procedure for a variable with such degree of missingness, we assumed that troponin levels were normal for those patients who did not have the assay complete. We acknowledge that this approach may have led to the

misclassification of patients who did not have a troponin assay completed in the ED, but did in fact have an abnormal underlying troponin level. To examine the potential for misclassification, we compared characteristics of those patients who did and did not have a troponin assay completed. We found that those who did not have a troponin assay completed in our study were generally younger, healthier, and had fewer cardiovascular-related comorbidities, rendering these patients unlikely to have abnormal troponin levels.

A second limitation is that we did not include any interaction terms in either the traditional or the modern models. The methodology outline d in the modern approach emphasizes the use of interaction terms to explore non-additivity of the predictor terms in a regression model[28,30]. Interaction terms model situations where the effect of one predictor depends on the value of another variable. The inclusion of interaction terms in a regression model can be useful in improving model performance if the interaction term allows for the underlying biological phenomena to be modelled more accurately[28,30]. However, interaction terms increase the complexity of the model and each interaction term accounts for additional degrees of freedom in the model[28]. Since we had a relatively small effective sample size that could only justify fitting a maximum of 14 degrees of freedom, we would have considered including interaction terms only if there were important interactions based on clinical or biological rationale. After discussion with the subject-matter experts, no important interactions could be identified and we therefore proceeded without the inclusion of any interactions.

6.5 Strengths

Our study was the largest prospective study deriving a syncope clinical decision tool to date. To our knowledge, the next largest syncope prediction model derived using prospective

data was the San Francisco Syncope Rule with a sample size of 684 patients, of which 79 suffered serious outcomes[10]. Having a large dataset in the derivation of a clinical prediction model has several advantages; namely, it increases statistical power, improves the generalizability of the results, and reduces the risk of statistical overfitting[28,30]. In this study, we also collected a diverse array of predictor variables representing elements of patient demographics, medical history, clinical features, and clinical investigations. As a result, we were able to generate a clinically meaningful and relevant prediction model that potentially captures the multi-facetted nature of a patient's characteristics, ultimately having the potential to translate to more robust predictions.

A second major strength of this study was the high level of collaboration and discussion among the subject-matter experts and the methodologists. In each step of the derivation process, both methodological and clinical considerations were used in making decisions, ensuring that the modeling was both statistically and clinically sensible. In the initial data preparation phases, great effort was put forth by both methodologists and clinicians in developing a parsimonious list of candidate predictors that were statistically-viable and clinically-sensible. We believe that this process was critical in limiting the threat of having spurious predictors included in the final prediction models, particularly in the data-driven traditional approach. We also had good collaboration and input from clinical experts when identifying the most important predictors to be included in the modern model.

6.6 Proposed Future Research

This study presents two new clinical prediction models, with a high degree of internal validity, and acceptable discrimination and calibration, to potentially aid risk-stratification and

influence decision-making in the treatment of adult ED syncope patients. However, before implementing the decision tools in a clinical setting, the prediction models must be validated in an external cohort of patients. External validation is a crucial step in prediction model development since it demonstrates whether the prediction model has robust and acceptable performance when applied to a new population; prediction models often do not perform as well when tested on new groups of patients[20,26]. In the validation phase of this study, we will validate both the traditional and modern model to assess and compare their model performance, ultimately providing insight in determining if one model provides more accurate predictions than the other. An additional future project will involve assessing the preliminary acceptance and clinical usefulness of the clinical prediction models by surveying ED physicians.

6.7 Conclusion

In this study, we successfully developed clinical prediction tools to identify ED syncope patients at risk of serious outcomes after ED disposition. Once successfully validated, the prediction model can aid physicians to confidently identify high-risk patients for appropriate treatment, while allowing safe disposition of low-risk patients. We believe that our syncope prediction models have promising potential to validate externally given that it displayed high internal validity and good statistical performance during the derivation phase.

In this study, we were also able to empirically compare some of the methodological considerations used in developing clinical prediction models by developing two parallel syncope prediction models using different methodological approaches. In the future, we will conduct a validation study to evaluate and compare the true performance of these prediction models.

References

1. European Society of Cardiology. Guidelines for the diagnosis and management of syncope (version 2009). *European Heart Journal*. 2009(30):2631-2671.

2. Thiruganasambandamoorthy V, Stiell I, Wells G. Frequency and outcomes of syncope in the emergency department. *Can J Emerg Med Care*. 2008;10:255-295.

3. Day SC, Cook EF, Funkenstein H, Goldman L. Evaluation and outcome of emergency room patients with transient loss of consciousness. *Am J Med*. 1982;73(1):15-23.

4. Kapoor WN, Karpf M, Wieand S, Peterson JR, Levey GS. A prospective evaluation and follow-up of patients with syncope. *N Engl J Med*. 1983;309(4):197-204.

5. Martin GJ, Adams SL, Martin HG, Mathews J, Zull D, Scanlon PJ. Prospective evaluation of syncope. *Ann Emerg Med*. 1984;13(7):499-504.

6. Puppala VK, Dickinson O, Benditt DG. Syncope: Classification and risk stratification. *J Cardiol*. 2014;63(3):171-177.

7. Task Force for the Diagnosis and Management of Syncope, European Society of Cardiology (ESC), European Heart Rhythm Association (EHRA), et al. Guidelines for the diagnosis and management of syncope (version 2009). *Eur Heart J*. 2009;30(21):2631-2671.

8. Sarasin FP, Hanusa BH, Perneger T, Louis-Simonet M, Rajeswaran A, Kapoor WN. A risk score to predict arrhythmias in patients with unexplained syncope. *Acad Emerg Med*. 2003;10(12):1312-1317.

9. Linzer M, Yang EH, Estes NM, Wang P, Vorperian VR, Kapoor WN. Clinical guideline: Diagnosing syncope: Part 1: Value of history, physical examination, and electrocardiography. *Ann Intern Med*. 1997;126(12):989-996.

10. Quinn JV, Stiell IG, McDermott DA, Sellers KL, Kohn MA, Wells GA. Derivation of the san francisco syncope rule to predict patients with short-term serious outcomes. *Ann Emerg Med*. 2004;43(2):224-232.

11. Quinn J, McDermott D, Stiell I, Kohn M, Wells G. Prospective validation of the san francisco syncope rule to predict patients with serious outcomes. *Ann Emerg Med*. 2006;47(5):448-454.

12. Birnbaum A, Esses D, Bijur P, Wollowitz A, Gallagher EJ. Failure to validate the san francisco syncope rule in an independent emergency department population. *Ann Emerg Med*. 2008;52(2):151-159.

13. Cosgriff TM, Kelly AM, Kerr D. External validation of the san francisco syncope rule in the australian context. *CJEM*. 2007;9(3):157-161.

14. Thiruganasambandamoorthy V, Hess EP, Alreesi A, Perry JJ, Wells GA, Stiell IG. External validation of the san francisco syncope rule in the canadian setting. *Ann Emerg Med*. 2010;55(5):464-472.

15. Sun BC, Mangione CM, Merchant G, et al. External validation of the san francisco syncope rule. *Ann Emerg Med*. 2007;49(4):420-427. e4.

16. Reed MJ, Newby DE, Coull AJ, Prescott RJ, Jacques KG, Gray AJ. The ROSE (risk stratification of syncope in the emergency department) study. *J Am Coll Cardiol*. 2010;55(8):713-721.

17. Grossman SA, Fischer C, Lipsitz LA, et al. Predicting adverse outcomes in syncope. *J Emerg Med*. 2007;33(3):233-239.

18. Thiruganasambandamoorthy V, Wells GA, Hess EP, Turko E, Perry JJ, Stiell IG. Derivation of a risk scale and quantification of risk factors for serious adverse events in adult emergency department syncope patients. *CJEM Canadian Journal of Emergency Medical Care*. 2014;16(2):120-130.

19. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med*. 1999;33(4):437-447.

20. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC medical research methodology*. 2014;14(1):40.

21. Kattan MW, Yu C, Stephenson AJ, Sartor O, Tombal B. Clinicians versus nomogram: Predicting future technetium-99m bone scan positivity in patients with rising prostate-specific antigen after radical prostatectomy for prostate cancer. *Urology*. 2013;81(5):956-961.

22. Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol*. 2002;20(2):82-88.

23. Toll D, Janssen K, Vergouwe Y, Moons K. Validation, updating and impact of clinical prediction rules: A review. *J Clin Epidemiol*. 2008;61(11):1085-1094.

24. Adams ST, Leveson SH. Clinical prediction rules. *BMJ*. 2012;344(7842):21-23.

25. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: How to use articles about clinical decision rules. *JAMA*. 2000;284(1):79-84.

26. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.

27. Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: A systematic review. *PLoS medicine*. 2012;9(5):e1001221.

28. Harrell FE. *Regression modeling strategies.* Springer Science & Business Media; 2001.

29. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338:b604.

30. Steyerberg EW. *Clinical prediction models.* Springer; 2009.

31. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ*. 2009;338:b375.

32. Brotman DJ, Walker E, Lauer MS, O'Brien RG. In search of fewer independent risk factors. *Arch Intern Med*. 2005;165(2):138-145.

33. Steyerberg EW, Moons KG, van der Windt, Danielle A, et al. Prognosis research strategy (PROGRESS) 3: Prognostic model research. *PLoS medicine*. 2013;10(2):e1001381.

34. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*. 2009;338:b2393.

35. Little RJ, Rubin DB. *Statistical analysis with missing data.* John Wiley & Sons; 2014.

36. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat Med*. 2006;25(1):127-141.

37. Perry JJ, Sharma M, Sivilotti ML, et al. A prospective cohort study of patients with transient ischemic attack to identify high-risk clinical characteristics. *Stroke*. 2014;45(1):92-100.

38. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med*. 1992;21(4):384-390.

39. Stiell IG, Clement CM, Aaron SD, et al. Clinical characteristics associated with adverse events in patients with exacerbation of chronic obstructive pulmonary disease: A prospective cohort study. *CMAJ*. 2014;186(6):E193-204.

40. Stiell IG, Lesiuk H, Wells GA, et al. The canadian CT head rule study for patients with minor head injury: Rationale, objectives, and methodology for phase I (derivation). *Ann Emerg Med*. 2001;38(2):160-169.

41. Schellingerhout JM, Heymans MW, de Vet HC, Koes BW, Verhagen AP. Categorizing continuous variables resulted in different predictors in a prognostic model for nonspecific neck pain. *J Clin Epidemiol*. 2009;62(8):868-874.

42. Lagakos S. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. *Stat Med*. 1988;7(1-2):257-274.

43. Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, Eron JJ,Jr. Splines for trend analysis and continuous confounder control. *Epidemiology*. 2011;22(6):874-875.

44. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med*. 2010;29(9):1037-1057.

45. Rosenberg PS, Katki H, Swanson CA, Brown LM, Wacholder S, Hoover RN. Quantifying epidemiologic risk factors using non-parametric regression: Model selection remains the greatest challenge. *Stat Med*. 2003;22(21):3369-3381.

46. Boucher KM, Slattery ML, Berry TD, Quesenberry C, Anderson K. Statistical methods in epidemiology: A comparison of statistical methods to analyze dose–response and trend analysis in epidemiologic studies. *J Clin Epidemiol*. 1998;51(12):1223-1233.

47. Sun G, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. 1996;49(8):907-916.

48. Wiegand RE. Performance of using multiple stepwise algorithms for variable selection. *Stat Med*. 2010;29(15):1647-1659.

49. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol*. 2004;57(11):1138-1146.

50. Steyerberg EW, Eijkemans MJ, Habbema JDF. Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. *J Clin Epidemiol*. 1999;52(10):935-942.

51. Taljaard M, Tuna M, Bennett C, et al. Cardiovascular disease population risk tool (CVDPoRT): Predictive algorithm for assessing CVD risk in the community setting. A study protocol. *BMJ Open*. 2014;4(10):e006701-2014-006701.

52. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: Validating a prognostic model. *BMJ*. 2009;338:b605.

53. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMC medicine*. 2015;13(1):1.

54. Martin TP, Hanusa BH, Kapoor WN. Risk stratification of patients with syncope. *Ann Emerg Med*. 1997;29(4):459-466.

55. Colivicchi F, Ammirati F, Melina D, et al. Development and prospective validation of a risk stratification system for patients with syncope in the emergency department: The OESIL risk score. *Eur Heart J*. 2003;24(9):811-819.

56. Costantino G, Perego F, Dipaola F, et al. 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*. 2008;51(3):276-283.

57. Del Rosso A, Ungar A, Maggi R, et al. Clinical predictors of cardiac syncope at initial evaluation in patients referred urgently to a general hospital: The EGSYS score. *Heart*. 2008;94(12):1620-1626.

58. Sun BC, Derose SF, Liang L, et al. Predictors of 30-day serious events in older patients with syncope. *Ann Emerg Med*. 2009;54(6):769-778. e5.

59. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting. *BMC Med*. 2011;9:103-7015-9-103.

60. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: A review. *BMC Med*. 2010;8:20-7015-8-20.

61. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: A review. *BMC Med*. 2010;8:21-7015-8-21.

62. Paul P, Pennell ML, Lemeshow S. Standardizing the power of the hosmer?lemeshow goodness of fit test in large data sets. *Stat Med*. 2013;32(1):67-80.

63. Sun BC, Costantino G, Barbic F, et al. Priorities for emergency department syncope research. *Ann Emerg Med*. 2014;64(6):649-655. e2.

64. Thiruganasambandamoorthy V, Stiell IG, Sivilotti ML, et al. Risk stratification of adult emergency department syncope patients to predict short-term serious outcomes after discharge (RiSEDS) study. *BMC emergency medicine*. 2014;14(1):8.

65. Sun BC, Thiruganasambandamoorthy V, Cruz JD. Standardized reporting guidelines for emergency department syncope risk-stratification research. *Acad Emerg Med*. 2012;19(6):694-702.

66. D'Agostino RB, Sullivan LM, Beiser AS. *Introductory applied biostatistics.* Thomson Brooks/Cole; 2006.

67. Kuhn M, Johnson K. *Applied predictive modeling.* ; 2013:600.

68. Statistical Analysis Systems (SAS). SAS user's guide, version 9.2. . 2007.

69. Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism-increasing the models utility with the SimpliRED D-dimer. *THROMBOSIS AND HAEMOSTASIS-STUTTGART-*. 2000;83(3):416-420.

70. Allison PD. *Missing data.* Vol 136. Sage publications; 2001.

71. Lambert J, Lipkovich I. A macro for getting more out of your ROC curve. . 2008;231.

72. Harrell FEJ. Regression modelling strategies. *Course Notes*. June 2015.

73. Bollen KA, Jackman RW. Regression diagnostics an expository treatment of outliers and influential cases. *Sociological Methods & Research*. 1985;13(4):510-542.

74. Harrell Jr. FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.

75. Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*. 1983;78(382):316-331.

76. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The framingham study risk score functions. *Stat Med*. 2004;23(10):1631-1660.

77. Soteriades ES, Evans JC, Larson MG, et al. Incidence and prognosis of syncope. *N Engl J Med*. 2002;347(12):878-885.

78. UCLA: Statisitcal Consulting Group. What are pseudo R-squareds? http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm. Updated October 20, 2011.

79. Quinn J. Risk stratification of patients with syncope. *CJEM*. 2007;9(03):174-175.

80. Costantino G, Casazza G, Reed M, et al. Syncope risk stratification tools vs clinical judgment: An individual patient data meta-analysis. *Am J Med*. 2014;127(11):1126. e13-1126. e25.

81. Wells P, Hirsh J, Anderson D, et al. Accuracy of clinical assessment of deep-vein thrombosis. *The Lancet*. 1995;345(8961):1326-1330.

Appendices

Appendix 1. Survey administered to physicians to identify the most important variables for syncope prediction model

Predictors for Syncope Risk Tool

In this study, we are attempting to identify the most important clinical predictors to enter into the multivariable regression analysis to derive the risk tool. Our methodological approach requires avoiding data-driven variable selection in favour of pre-specifying all predictors for the risk tool prior to analysis.

The following contains a list of 26 Candidate Predictors (grouped by category) that we must choose from. Any predictors that are not selected in Question 1 or 2 in this survey will not be offered to the tool.

Demographics
-Sex
-Age

Visit History
-Systolic Blood Pressure in the ED
-Diastolic Blood Pressure in the ED
-Heart Rate in the ED
-Respiratory Rate at Triage
-Oxygen Saturation at Triage

Medical History
-History of Vascular disease (defined by coronary artery disease, TIA, or CVA)
-History of Heart disease (defined by valvular heart disease, cardiomyopathy, or congestive heart failure)
-History of Non-sinus cardiac rhythm (defined by ventricular arrhythmia, pacemaker, ICD, atrial fibrillation/Flutter, or non-sinus ECG rhythm observed in the ED)

Investigation
-Hemoglobin/ Hematocrit
-Creatinine/ BUN
-Troponin

Syncope Event details
-Most likely cause of syncope as per ED physician at the end of the ED visit (ie: vasovagal, orthostatic, cardiac, NYD)
-Syncope witnessed by someone else
-Palpitations prior to syncope
-Physical exertion prior to syncope
-Predisposition for vasovagal syncope
-Prodrome prior to syncope
-Orthostatic symptoms

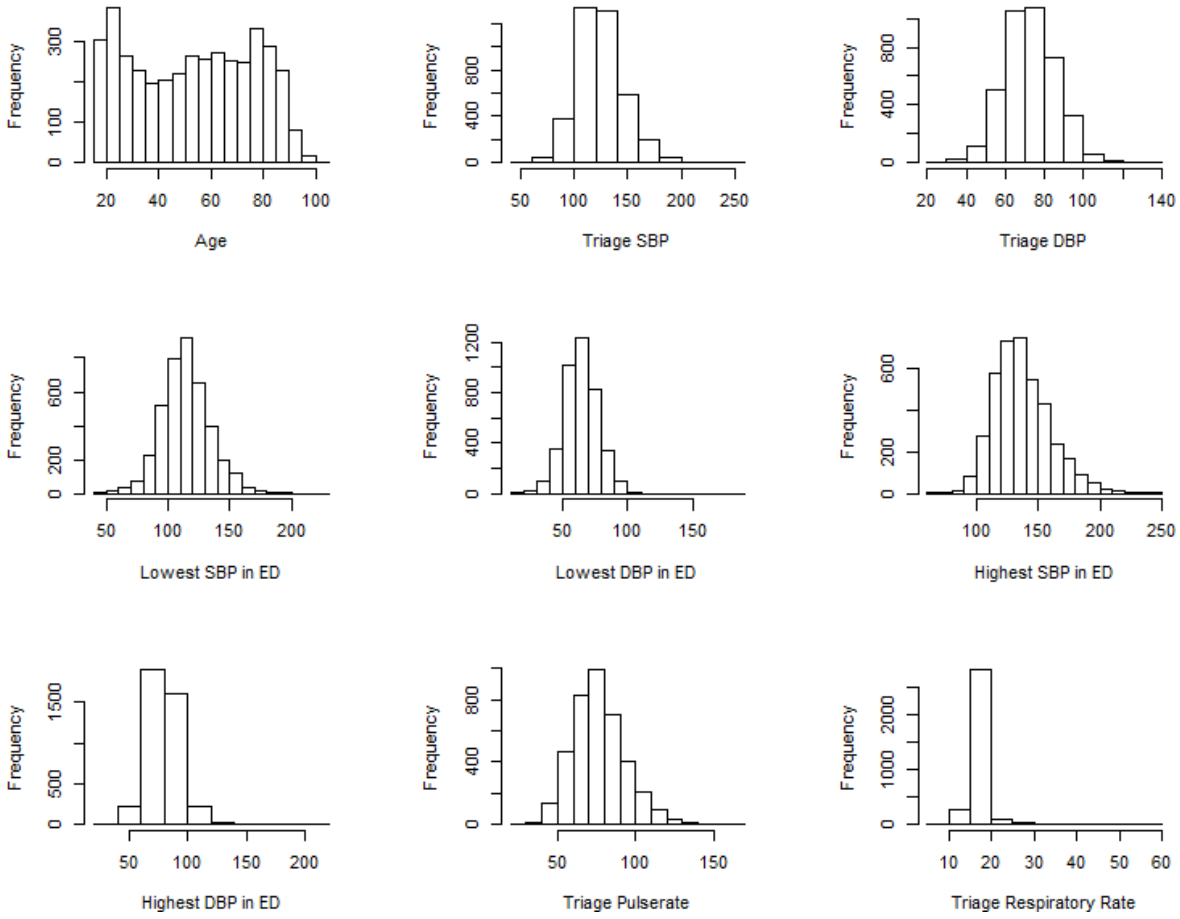
ECG characteristics
-Left Bundle Branch Block
-QRS duration

-QRS axis
-Corrected QT interval
-Left Ventricular Hypertrophy
-Old ischemia (evidence of old MI or old ischemic changes)

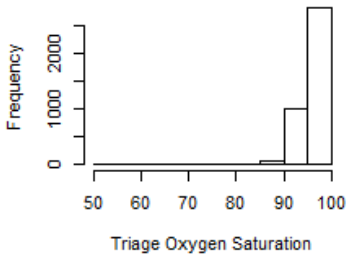
Question 1: Please identify the 5 most important variables that you feel must be included in the Syncope Risk Tool

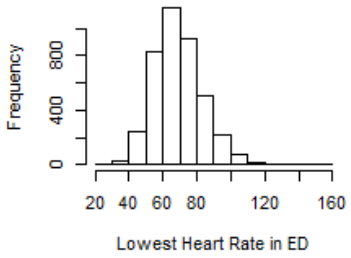
Question 2: Please identify 10 additional important variables (not identified in Question 1) that should be considered for the Tool from the remaining 21 predictors. Rank these additional predictors in order of importance (where 1= most important).

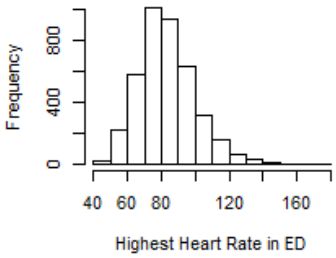
Appendix 2. Histograms describing the distributions of continuous potential predictors

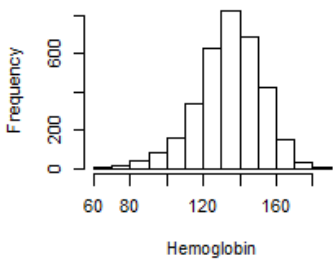


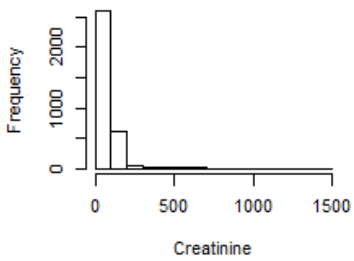
Appendix 2. Histograms describing the distributions of continuous potential predictors (cont'd)

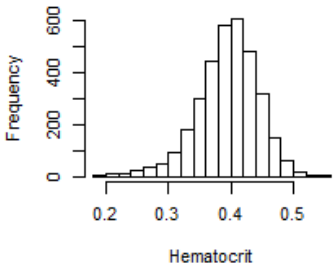


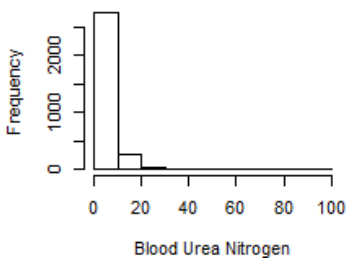


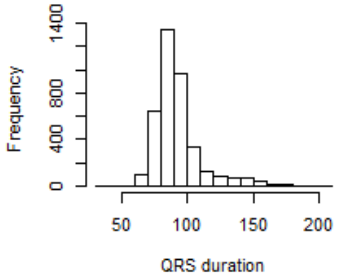


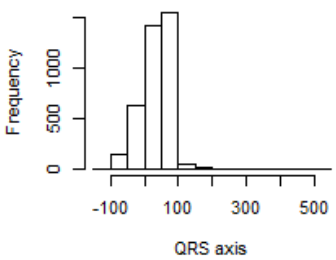


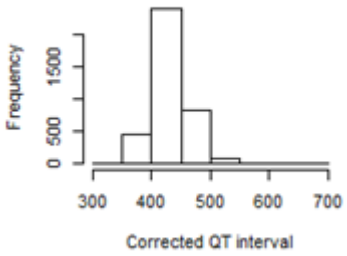












Appendix 3. Comparison of patient characteristics for patients who did and did not have a troponin assay completed in the ED

Characteristic	Troponin Completed (n=1932)	No Troponin Completed (n=2102)	p-value
Age, Years (Mean, SD)	65.2 (18.8)	42.8 (21.2)	<0.0001
Range	16 -102	16-99	
Female	907 (47.0)	1334 (63.5)	<0.0001
Medical History			
Coronary artery disease	378 (19.6)	98 (4.7)	<0.0001
Valvular heart disease	94 (4.9)	43 (2.1)	<0.0001
Congestive heart failure	120 (6.2)	32 (1.5)	<0.0001
Atrial fibrillation/flutter	217 (11.2)	74 (3.5)	<0.0001
Ventricular arrhythmia	20 (1.0)	3 (0.1)	0.0002
Pacemaker/ICD insertion	66 (3.4)	15 (0.7)	<0.0001
Hypertension	928 (48.0)	366 (17.4)	<0.0001
Diabetes	286 (14.8)	116 (5.5)	<0.0001
Presumed cause of syncope at ED			
Vasovagal	788 (40.8)	1429 (68.0)	<0.0001
Cause unknown	443 (22.9)	204 (9.7)	<0.0001
Orthostatic hypotension	240 (12.4)	229 (11.0)	0.1128
Cardiac	198 (10.3)	40 (1.9)	<0.0001
Disposition/Outcomes			
Admitted to hospital	304 (15.7)	79 (3.8)	<0.0001
Serious Outcomes disposition	115 (6.0)	32 (1.5)	<0.0001

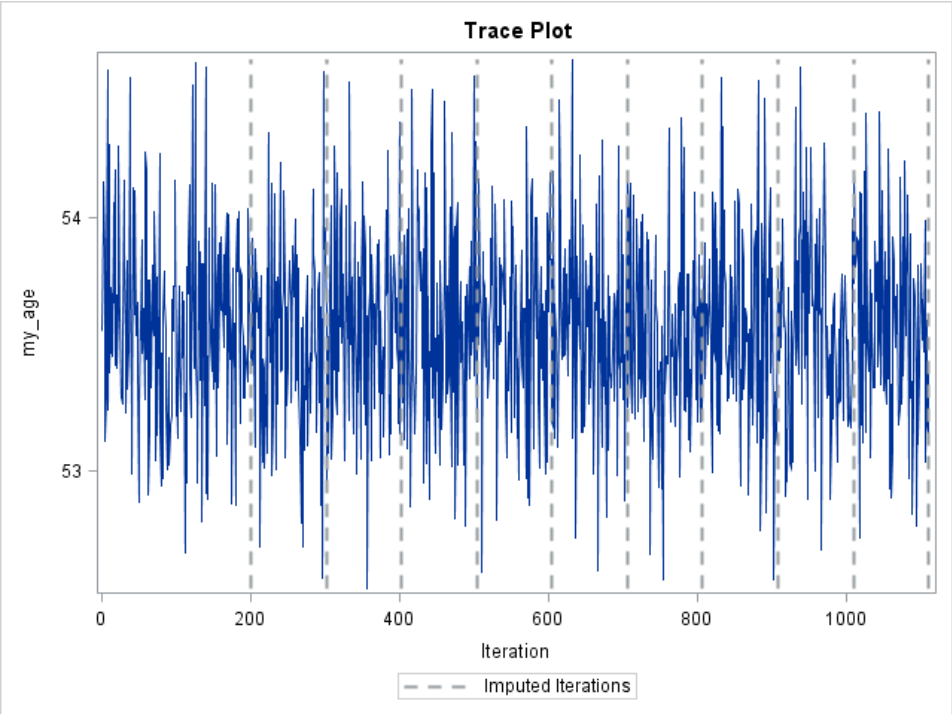
Appendix 4a. Investigation of missing data mechanism by comparing the values of continuous predictors between patients with complete data and missing values in at least one variable

	Complete cases (n=1981)	**Missing data (n=2052)**	
Variable	**Mean (SD)**	**Mean (SD)**	**p-value**
Demographics			
Age	56.4 (22.1)	50.8 (23.5)	<0.0001
Visit History			
Systolic blood pressure at triage	127.0 (22.3)	124.1 (21.6)	<0.0001
Lowest systolic blood pressure in the ED	114.9 (19.7)	115.2 (19.3)	0.61
Highest systolic blood pressure in the ED	139.6 (22.7)	135.5 (22.1)	<0.0001
Diastolic blood pressure at triage	74.0 (13.3)	73.4 (13.4)	0.15
Lowest diastolic blood pressure in the ED	64.5 (12.9)	65.9 (13.1)	0.0006
Highest diastolic blood pressure in the ED	81.8 (13.6)	79.2 (14.0)	<0.0001
Pulse rate at triage	77.3 (17.3)	77.7 (16.7)	0.43
Lowest heart rate in the ED	69.4 (14.4)	70.2 (13.9)	0.11
Highest heart rate in the ED	84.7 (17.1)	83.5 (16.5)	0.02
Respiratory rate at triage	17.3 (2.7)	17.2 (2.6)	0.56
Oxygen saturation at triage	96.3 (3.6)	96.8 (2.9)	<0.0001
Investigation			
Hemoglobin	134.9 (17.2)	134.0 (18.1)	0.15
Hematocrit	0.40 (0.05)	0.40 (0.05)	0.45
Creatinine (log-transformed)	4.4 (0.4)	4.4 (0.4)	0.22
Blood urea nitrogen (log-transformed)	1.7 (0.5)	1.8 (0.5)	0.06
ECG			
QRS duration	93.9 (19.2)	93.4 (17.9)	0.45
QRS axis	31.8 (42.0)	38.3 (44.3)	<0.0001
Corrected QT interval	433.7 (31.6)	431.9 (31.9)	0.08

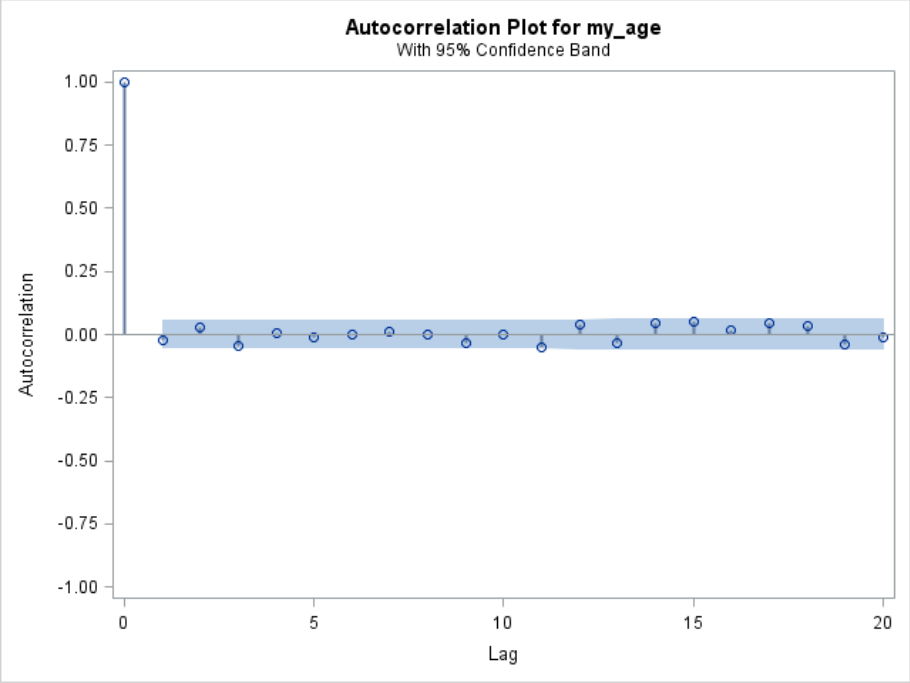
Appendix 4b. Investigation of missing data mechanism by comparing the values of categorical predictors between patients with complete data and missing values in at least one variable

	Complete cases (n=1981)	**Missing data (n=2053)**	
Variable	**n (column %)**	**n (column %)**	**p-value**
Demographics			
Female	1068 (53.9)	1137 (57.1)	0.04
Medical History			
History of cerebrovascular disease	147 (7.4)	108 (5.3)	0.01
History of heart disease	469 (23.7)	362 (20.1)	0.01
Investigation			
Troponin	86 (4.3)	83 (4.0)	0.93
MD Data Form			
MD suspected diagnosis			<0.01
Cardiac	137 (6.9)	108 (5.3)	
Vasovagal	952 (48.1)	1233 (60.1)	
Orthostatic/ unknown	892 (45.0)	710 (34.6)	
Witnessed by someone	1363 (68.8)	1469 (73.1)	<0.01
Palpitations	120 (6.0)	98 (5.0)	0.13
High physical exertion prior to event	785 (39.6)	793 (39.6)	0.96
Predisposition for vasovagal	680 (34.3)	981 (49.3)	<0.01
Prodrome	1423 (71.8)	1565 (78.9)	<0.01
Orthostatic symptoms	448 (22.6)	367 (19.5)	0.02
ECG			
Left bundle branch block	51 (2.6)	49 (2.7)	0.78
Left ventricular hypertrophy	121 (6.1)	91 (5.1)	0.16
Old ischemia	103 (5.2)	73 (4.1)	0.10
Serious Outcome			
Suffered Serious Outcome	83 (4.2)	64 (3.2)	0.07

Appendix 5. Trace plot displaying imputation coefficients throughout imputation iterations for age



Appendix 6. Autocorrelation plot displaying imputation coefficient correlations between various lag intervals of imputation iterations for age



Appendix 7a. Mean, range and standard deviation for continuous variables, comparing raw and imputation-completed datasets

			Multiple Imputation Datasets	
Variable	**Number missing (n,%)**	**Raw Dataset (n=4034)**	**Proc MI (n=4034)**	**aregImpute (n=4034)**
Analysis Variables				
Age	1 (0.0)	53.6 (23.0)	53.6 (23.0)	53.6 (23.0)
Systolic blood pressure at triage	119 (3.0)	125.6 (22.0)	125.6 (22.0)	126.6 (22.0)
Lowest systolic blood pressure in the ED	30 (0.7)	115.0 (19.5)	115.0 (19.5)	115.0 (19.5)
Highest systolic blood pressure in the ED	30 (0.7)	137.5 (22.5)	137.4 (22.5)	137.5 (22.5)
Diastolic blood pressure at triage	142 (3.5)	73.7 (13.3)	73.6 (13.4)	73.6 (13.4)
Lowest diastolic blood pressure in the ED	33 (0.8)	65.2 (13.0)	65.2 (13.0)	65.2 (13.0)
Highest diastolic blood pressure in the ED	33 (0.8)	80.5 (13.9)	80.5 (13.9)	80.4 (13.9)
Lowest heart rate in the ED	31 (0.8)	69.8 (14.2)	69.9 (14.2)	69.9 (14.2)
Highest heart rate in the ED	32 (0.8)	84.1 (16.8)	84.1 (16.8)	84.1 (16.8)
Pulse rate at triage	149 (3.7)	77.5 (17.0)	77.4 (17.0)	77.4 (17.0)
Respiratory rate at triage	812 (20.1)	17.2 (2.7)	17.2 (2.7)	17.2 (2.7)
Oxygen saturation at triage	82 (2.0)	96.6 (3.3)	96.6 (3.3)	96.6 (3.3)
Hemoglobin	662 (16.4)	134.5 (17.6)	135.3 (17.7)	134.9 (17.3)
Hematocrit	661 (16.4)	0.4 (0.05)	0.4 (0.05)	0.40 (0.05)
Creatinine (log-transformed)	729 (18.0)	4.4 (0.4)	4.3 (0.4)	4.4 (0.4)
Blood urea nitrogen (log-transformed)	958 (23.8)	1.7 (0.5)	1.7 (0.5)	1.7 (0.5)
QRS duration	221 (5.5)	93.7 (18.6)	93.5 (18.5)	93.6 (18.5)
QRS axis	219 (5.4)	34.9 (43.2)	35.4 (43.1)	35.6 (43.1)
Corrected QT interval	223 (5.5)	432.8 (31.7)	432.4 (31.8)	432.5 (31.6)

Appendix 7b. Frequency distribution for categorical variables, comparing original and imputation-completed datasets. Table entries are smallest frequency (%)

Variable	Number missing (n,%)	Raw Dataset (n=4034)	Multiple Imputation datasets	
			Proc MI (n=4034)	aregImpute (n=4034)
Analysis Variables				
Outcome	0 (0.0)	147 (3.6)	147 (3.6)	147 (3.6)
Sex	0 (0.0)	1793 (44.5)	1793 (44.5)	1793 (44.5)
Cerebrovascular disease	7 (0.2)	255 (6.3)	255 (6.3)	256 (6.3)
Heart disease	254 (6.3)	831 (22.0)	858 (21.3)	858 (21.3)
Troponin	0 (0.0)	169 (4.2)	169 (4.2)	169 (4)
MD suspected diagnosis	4 (0.1)	246 (6.1)	246 (6.1)	248 (6)
Witnessed by someone	44 (1.1)	1158 (29.0)	1172 (29.1)	1176 (29)
Palpitations	75 (1.9)	218 (5.5)	221 (5.4)	221 (5)
Physical exertion prior to event	48 (1.2)	1578 (39.6)	1599 (39.6)	1605 (39)
Predisposition for vasovagal	63 (1.6)	1661 (41.8)	1680 (41.7)	1682 (42)
Prodrome	69 (1.7)	977 (24.6)	1006 (24.9)	998 (25)
Orthostatic symptoms	166 (4.1)	815 (21.1)	868 (21.5)	848 (21)
Left bundle branch block	254 (6.3)	100 (2.7)	101 (2.5)	104 (3)
Left ventricular hypertrophy	253 (6.3)	212 (5.6)	217 (5.4)	227 (6)
Auxiliary variables				
Right axis deviation	252 (6.3)	75 (2.0)	75 (1.9)	83 (2)
Old ischemia	254 (6.3)	176 (4.7)	176 (4.4)	187 (5)
Symptoms post-event	356 (8.8)	889 (24.2)	1003 (24.9)	971 (24)
Left anterior fascicular block	253 (6.3)	111 (2.9)	111 (2.8)	113 (3)
Right bundle branch block	253 (6.3)	178 (4.7)	180 (4.5)	188 (5)
Left axis deviation	253 (6.3)	219 (5.8)	224 (5.6)	227 (6)
Diabetes	6 (0.2)	402 (10.0)	403 (10.0)	403 (10)
Hypertension	4 (0.1)	1294 (32.1)	1296 (32.1)	1295 (32)
Dementia	4 (0.1)	131 (3.3)	131 (3.3)	131 (3.3)

Appendix 8. Final multivariable logistic regression model derived in traditional approach estimated on original dataset to verify stability of multiple imputation procedure

Parameter	**Parameter Estimate**	**Odds Ratio (95% CI)**	**Standard Error**	**p-value**
History of heart disease	0.56	1.7 (1.2, 2.6)	0.21	0.012
Elevated troponin (>99%ile normal population)	1.24	3.4 (2.1, 5.7)	0.26	<0.0001
MD suspected diagnosis - Vasovagal	-1.08	0.3 (0.2, 0.6)	0.31	0.0002
MD suspected diagnosis - Cardiac	1.16	3.2 (2.0, 5.0)	0.23	<0.0001
Predisposition to vasovagal	-0.67	0.5 (0.3, 0.9)	0.30	0.045
Extreme systolic blood pressure (<90 or >180 mmHg)	0.81	2.2 (1.5, 3.5)	0.22	0.0004
QRS duration (>130 milliseconds)	0.71	2.0 (1.2, 3.5)	0.28	0.015
Abnormal QRS axis (<-30 or >110)	0.47	1.6 (1.0, 2.6)	0.25	0.031
Corrected QT interval (>480 milliseconds)	1.08	2.9 (1.8, 4.7)	0.24	<0.0001

c-statistic: 0.885
Nagelkerke's R^2: 0.28
Brier Score: 0.031

Appendix 9. Equation describing full model derived from the modern approach with non-shrunken regression coefficients

$$
\begin{aligned}
&-1.71 + 0.04 * \text{ Age} - 6.07 * 10^{-6} * \text{pmax}(\text{Age} - 22,0)^3 + 1.29 * 10^{-5} \\
&* \text{pmax}(\text{Age} - 55,0)^3 - 6.92 * 10^{-6} * \text{pmax}(\text{Age} - 84,0)^3 - 0.02 \\
&* \text{ Triage SBP} + 1.26 * 10^{-5} * \text{pmax}(\text{Triage SBP} - 100,0)^3 - 2.26 * 10^{-5} \\
&* \text{pmax}(\text{Triage SBP} - 124,0)^3 + 1.00 * 10^{-5} * \text{pmax}(\text{Triage SBP} - 154,0)^3 \\
&+ 0.58 * (\text{Heart Disease}) - 0.18 * (\text{Cerebrovascular Disease}) + 1.19 \\
&* (\text{Positive Troponin Test}) - 1.20 * (\text{MD suspected Diagnosis of Vasovagal}) \\
&+ 1.15 * (\text{MD suspected Diagnosis of Cardiac}) - 0.24 * (\text{Palpitations}) + 0.01 \\
&* (\text{Physical Exertion}) - 0.07 * (\text{Prodrome}) - 0.004 * \text{ QT interval} + 3.00 \\
&* 10^{-6} * \text{pmax}(\text{QT interval} - 398,0)^3 - 5.17 * 10^{-6} \\
&* \text{pmax}(\text{QT interval} - 429,0)^3 + 2.17 * 10^{-6} * \text{pmax}(\text{QT interval} - 472,0)^3
\end{aligned}
$$

The function "pmax" in the full equation indicates that one should use the maximum value from the terms in brackets. This function is required to describe the intervals defined by the various knots of the restricted cubic spline functions.

Appendix 10. Final multivariable logistic regression model derived in modern approach estimated on original dataset to verify stability of multiple imputation procedure

Parameter	**Parameter Estimate**	**Odds Ratio (95% CI)**	**Standard Error**	**p-value**
History of heart disease	0.64	1.90 (1.23 to 2.95)	0.22	<0.01
History of vascular disease	-0.25	0.77 (0.39 to 1.55)	0.35	0.47
Elevated troponin (>99%ile normal population)	1.15	3.18 (1.87 to 5.40)	0.27	<0.01
MD suspected diagnosis - vasovagal	-1.26	0.28 (0.15 to 0.52)	0.31	<0.01
MD suspected diagnosis - cardiac	1.12	3.06 (1.91 to 4.91)	0.24	<0.01
Palpitations	-0.16	0.85 (0.29 to 2.49)	0.55	0.76
Physical Exertion	-0.07	0.94 (0.62 to 1.42)	0.21	0.75
Prodrome	-0.16	1.17 (0.77 to 1.80)	0.21	0.45
Age (74 vs. 31)	0.97	2.63 (1.22 to 5.63)	0.39	0.03
Systolic Blood Pressure (138 vs. 110 mmHg)	-0.09	0.91 (0.72 to 1.15)	0.12	<0.01
Corrected QT interval (450 vs. 411 mmHg)	0.18	1.20 (0.86 to 1.66)	0.16	<0.01

c-statistic: 0.89
Nagelkerke's R^2: 0.289
Brier Score: 0.029

Appendix 11. Equation describing full model derived from the modern approach with shrunken regression coefficients

$$-1.75 + 0.04 * \text{Age} - 5.71 * 10^{-6} * \text{pmax}(\text{Age} - 22,0)^3 + 1.22 * 10^{-5} * \text{pmax}(\text{Age} - 55,0)^3 - 6.49 * 10^{-6} * \text{pmax}(\text{Age} - 84,0)^3 - 0.02 * \text{Triage SBP} + 1.18 * 10^{-5} * \text{pmax}(\text{Triage SBP} - 100,0)^3 - 2.12 * 10^{-5} * \text{pmax}(\text{Triage SBP} - 124,0)^3 + 9.44 * 10^{-6} * \text{pmax}(\text{Triage SBP} - 154,0)^3 + 0.54 * (\text{Heart Disease}) - 0.17 * (\text{Cerebrovascular Disease}) + 1.12 * (\text{Positive Troponin Test}) - 1.13 * (\text{MD suspected diagnosis of Vasovagal Syncope}) + 1.08 * (\text{MD suspected diagnosis of Cardiac Syncope}) - 0.22 * (\text{Palpitations}) + 0.01 * (\text{Physical Exertion}) - 0.06 * (\text{Prodrome}) - 0.003 * \text{QTinterval} + 2.82 * 10^{-6} * \text{pmax}(\text{QT interval} - 398,0)^3 - 4.86 * 10^{-6} * \text{pmax}(\text{QT interval} - 429,0)^3 + 2.04 * 10^{-6} * \text{pmax}(\text{QT interval} - 472,0)^3$$

Appendix 12. Equation describing reduced model derived from the modern approach with shrunken regression coefficients

$$-1.60 - 0.02 * \text{Triage SBP} + 1.24 * 10^{-5} * \text{pmax}(\text{Triage SBP} - 100,0)^3 - 2.22 * 10^{-5} * \text{pmax}(\text{Triage SBP} - 124,0)^3 + 9.89 * 10^{-6} * \text{pmax}(\text{Triage SBP} - 154,0)^3 + 0.67 * (\text{History of Heart Disease}) + 1.16 * (\text{Positive Troponin Test}) - 1.22 * (\text{MD suspected diagnosis of Vasovagal Syncope}) + 1.09 * (\text{MD suspected diagnosis of Cardiac Syncope}) + 0.0003 * \text{QTinterval} + 2.30 * 10^{-6} * \text{pmax}(\text{QT interval} - 398,0)^3 - 3.96 * 10^{-6} * \text{pmax}(\text{QTinterval} - 429,0)^3 + 1.66 * 10^{-6} * \text{pmax}(\text{QT interval} - 472,0)^3$$

Appendix 13. Overall comparison of the expected and predicted probabilities from the traditional and modern mode (probability scale)