Injury

External validation of the Dutch prediction model for prehospital triage of trauma patients in South West region of England, United Kingdom

--Manuscript Draft--

Manuscript Number:	JINJ-D-20-02797R1
Article Type:	Full length article
Keywords:	emergency medicine; Emergency Care; prediction models; Trauma systems; Major Trauma; triage; Prehospital; undertriage; overtriage
Corresponding Author:	Thomas Shanahan, BA PGCert MA MBChB University of Manchester School of Medicine: The University of Manchester Faculty of Biology Medicine and Health Manchester, UNITED KINGDOM
First Author:	Thomas Shanahan, BA PGCert MA MBChB
Order of Authors:	Thomas Shanahan, BA PGCert MA MBChB
	Dr Gordon Ward Fuller, BSc, MPH, PhD, MBChB, MRCS
	Prof Trevor Sheldon, DSc, MSc, FMedSci
	Emily Turton, BA, MA
	Fionn Michael Appelton Quilty, BSc
	Dr Carl Marincowitz, MB BChir, PhD, MSc, BA(hons)
Abstract:	Importance: This paper investigates the use of a major trauma prediction model in the UK setting. We demonstrate that application of this model could reduce the number of patients with major trauma being incorrectly sent to non-specialist hospitals. However, more research is needed to reduce over-triage and unnecessary transfer to Major Trauma Centres. Objective: To externally validate the Dutch prediction model for identifying major trauma in a large unselected prehospital population of injured patients in England. Design: External validation using a retrospective cohort of injured patients who ambulance crews transported to hospitals. Setting: South West region of England. Participants: All patients 16 years with a suspected injury and transported by ambulance in the year from February 1, 2017. Exclusion criteria: 1) Patients aged ≤15 years; 2) Non-ambulance attendance at hospital with injuries; 3) Death at the scene and; 4) Patients conveyed by helicopter. This study had a census sample of cases available to us over a one year period. Interventions or exposures: Tested the accuracy of the prediction model in terms of discrimination, calibration, clinical usefulness, sensitivity and specificity and under- and over triage rates compared to usual triage practices in the South West region. Main outcome measure: Major trauma defined as an Injury Severity Score>15. Results: A total of 68799 adult patients were included in the external validation cohort. The median age of patients was 72 (i.q.r. 46-84); 55.5% were female; and 524 (0.8%) had an Injury Severity Score>15. The model achieved good discrimination with a C-Statistic 0.75 (95% CI, 0.73 – 0.78). The maximal specificity of 50% and sensitivity of 83% suggests the model could improve undertriage rates at the expense of increased overtriage rates compared with routine trauma triage methods used in the South West, England. Conclusions and relevance: The Dutch prediction model for identifying major trauma could lower the undertriage rate to 17%, however it would increase the overtriage rate

	to 50% in this United Kingdom cohort. Further prospective research is needed to determine whether the model can be practically implemented by paramedics and is cost-effective.
Suggested Reviewers:	

Title Page

Manuscript Title: External validation of the Dutch prediction model for prehospital triage of trauma patients in South West region of England, United Kingdom

Authors:

Corresponding author: Dr Thomas Alexander Gerrard Shanahan, BA(Hons), PGCert, MA, MBChB. Affiliation: University of Manchester, Division of Cardiovascular Sciences, Oxford Road, Manchester, M13 9PL; Manchester University NHS Foundation Trust and National Institute for Health Research. Email: Thomas.shanahan1@nhs.net

Dr Gordon Ward Fuller, BSc(Hons), MPH, PhD, MBChB, MRCS. Affiliation: Centre for Urgent and Emergency Care Research, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. G.fuller@sheffield.ac.uk

Professor Trevor Sheldon, DSc, MSc, FMedSci. Affiliation: Institute of Population Health Sciences, Barts and the London School of Medicine and Dentistry, Queen Mary University of London. Trevor.sheldo@york.ac.uk

Emily Turton BA (Dunelm) MA. Affiliation: School of Health and Related Research (ScHARR), The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA. E.j.turton@sheffield.ac.uk

Mr Fionn Michael Appleton Quilty, BSc(Hons). Affiliation: Hull York Medical School, University of York, Heslington, York, YO10 5DD. Hyfq1@hyms.ac.uk

Dr Carl Marincowitz, MB BChir, PhD, MSc, BA(hons). Affiliation: Centre for Urgent and Emergency Care Research, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. Cm444@doctors.org.uk

Keywords: emergency medicine, emergency care, prediction models, trauma systems, major trauma, triage, prehospital, undertriage, overtriage

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

This work was supported by the National Institute for Health Research (NIHR). CM & GF are NIHR Clinical Lecturers and TS is a NIHR Academic Clinical Fellow. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Acknowledgments: We would like to sincerely thank Professor van Heijl and colleagues from University Medical Center Utrecht for their assistance with specifying the major trauma prediction model. We are also very grateful to the South West Ambulance Service Foundation Truest and the Trauma Audit and Research Network for supplying anonymised data. Finally we would like to acknowledge the help of Simon Waterhouse (Data management, Clinical Trials and Research Unit, University of Sheffield) for his expertise and assistance in performing deterministic data linkage.

Data sharing statement

Our data sharing agreements with TARN and the South west ambulance service prohibit sharing our original data, but other researchers can request the data we used from these bodies.

Title Page

Manuscript Title: External validation of the Dutch prediction model for prehospital triage of trauma patients in South West region of England, United Kingdom

Word Count: 4074

Keywords: emergency medicine, emergency care, prediction models, trauma systems, major trauma, triage, prehospital, undertriage, overtriage

Abstract

Importance: This paper investigates the use of a major trauma prediction model in the UK setting. We demonstrate that application of this model could reduce the number of patients with major trauma being incorrectly sent to non-specialist hospitals. However, more research is needed to reduce over-triage and unnecessary transfer to Major Trauma Centres.

Objective: To externally validate the Dutch prediction model for identifying major trauma in a large unselected prehospital population of injured patients in England.

Design: External validation using a retrospective cohort of injured patients who ambulance crews transported to hospitals.

Setting: South West region of England.

Participants: All patients ≥16 years with a suspected injury and transported by ambulance in the year from February 1, 2017. Exclusion criteria: 1) Patients aged ≤15 years; 2) Non-ambulance attendance at hospital with injuries; 3) Death at the scene and; 4) Patients conveyed by helicopter. This study had a census sample of cases available to us over a one year period.

Interventions or exposures: Tested the accuracy of the prediction model in terms of discrimination, calibration, clinical usefulness, sensitivity and specificity and under- and over triage rates compared to usual triage practices in the South West region.

Main outcome measure: Major trauma defined as an Injury Severity Score>15.

Results: A total of 68799 adult patients were included in the external validation cohort. The median age of patients was 72 (i.q.r. 46-84); 55.5% were female; and 524 (0.8%) had an Injury Severity Score>15. The model achieved good discrimination with a C-Statistic 0.75 (95% CI, 0.73 – 0.78). The maximal specificity of 50% and sensitivity of 83% suggests the model could improve undertriage rates at the expense of increased overtriage rates compared with routine trauma triage methods used in the South West, England.

Conclusions and relevance: The Dutch prediction model for identifying major trauma could lower the undertriage rate to 17%, however it would increase the overtriage rate to 50% in this United Kingdom cohort. Further prospective research is needed to determine whether the model can be practically implemented by paramedics and is cost-effective.

Highlights

- First study to externally validate the only empirically derived European prediction model for identifying major trauma in undifferentiated pre-hospital patients.
- Implementing the Dutch model could lower the undertriage rate from 56% to 17%, however it would increase the overtriage rate from 16% to 50%.
- Further research is needed to reduce overtriage and determine whether the model can be practically implemented by paramedics and is cost-effective.

Introduction

Trauma is a major cause of mortality and disability globally(1). It is the leading cause of death in those aged under 40 and increasing in the ageing population(1,2). Treating patients with serious injuries in specialist Major Trauma Centres (MTCs) can increase survival(1,3–6). In England, introduction of regional trauma systems resulted in a 20% increase in adjusted odds of survival from severe injury(1). In the United States, a 9% lower crude mortality rate was found in states with organised trauma systems(7), and in Victoria, Australia, there was a relative mortality reduction of 38% over five years following introduction of a trauma system(8).

Accurate pre-hospital identification of patients with severe injuries and their triage to specialised centres is required in order to achieve these benefits. Undertriage, whereby patients with major trauma are incorrectly taken to the nearest non-specialist hospital instead of a MTC, results in avoidable morbidity and mortality(1,3,9–11). Overtriage, whereby a patient without significant injury is inappropriately triaged to an MTC, can result in scarce resources being wasted and patients inconvenienced(12,13). The American College of Surgeons Committee on Trauma (ACS-COT) recommends limits on undertriage rates at 5% and up to 50% for overtriage(13).

Existing triage methods in the UK use a step by step process which looks at the vital signs and conscious level, anatomy of injury, mechanism of injury and special considerations, such as length of time it will take to get to an MTC, airway compromise and catastrophic haemorrhage (14). However, they are consensus based and inaccurate(3,15). As in other areas of clinical decision making(16), triage decisions may be improved by using empirically derived prediction models(17).

Recently, a Dutch major trauma triage prediction model was empirically derived to guide prehospital transport decisions. It was tested in the Central Netherlands and validated in the Brabant region, with an undertriage rate of 11.2% and overtriage rate of 50%. The 50% overtriage rate was pre-set by the Dutch researchers(18), as the limit of the ACS-COT guidance. The Dutch researchers recommended external validation to account for differences in the incidence of severe injuries in other populations and help determine the acceptable baseline risks in different settings(18).

The aim of this study was to externally validate the Dutch prediction model in a large and unselected pre-hospital population of injured patients in England.

Methods

We externally validated the Dutch prediction model in a retrospective cohort of injured patients attended by ambulance crews in the South West region of England. We tested the accuracy of the model in terms of discrimination, calibration, clinical usefulness and reported under- and overtriage rates compared to usual practice. The study adhered to, and is reported using, international guidelines (TRIPOD) for prognostic model validation(19).

Setting

Care for injured patients in England is provided by regional major trauma networks. Patients with suspected major trauma in the field are transported directly to MTCs (equivalent to ACS

Level 1 or 2 trauma centres). Injured patients who do not meet prehospital major trauma triage criteria, or with uncontrolled haemorrhage or unstable airways, are transported to their nearest hospital (‘trauma units’, equivalent to ACS level 3 trauma centres).

Sources of data

Data routinely collected in the prehospital environment on injured patients were provided by the South Western Ambulance Service NHS Foundation Trust (SWASFT) for the period February 1, 2017 through to February 1, 2018. Ambulance personnel routinely record demographic information, vital signs and other physiological parameters at scene, mechanism of injury, type of injury, frailty score, name of the receiving hospital and a free text summary of the incident in an electronic patient record form (ePRF).

Prehospital data was matched by deterministic data linkage with data submitted to the Trauma Audit and Research Network (TARN) on eligible patients (using incident number, age and sex) by MTCs and Trauma Units (TUs) to provide information on the Injury Severity Score (ISS)(20). TARN is the registry for collating data on trauma patients in the UK. TARN includes seriously injured patients admitted to MTCs and TU including those who die during admission, are admitted for ≥3 days, undergo transfer for specialist care, or require critical care. Full eligibility criteria have been published previously(21).

Participants

All patients 16 years and older with a suspected injury and transported by SWASFT from February 1, 2017 to February 1, 2018. The study had the following exclusion criteria:

- Patients aged ≤15 years.

- Non-ambulance attendance at hospital with injuries.
- Death at the scene.
- Patients conveyed by helicopter emergency medical services.

Predictors

The Dutch prediction model found eight independent predictor variables significant for identifying major trauma in patients(18). Data on these variables were available in the South West ambulance service ePRF. The coding ‘multiple trauma’ was used as a proxy for suspected injury in ≥2 AIS regions. An algorithm using free text in the ambulance dataset was developed to identify mechanism criteria (fall of greater than 2 metres, motor vehicle collision, or any type of entrapment), thorax injury, penetrating injury to head, thorax or abdomen, and head or neck injury. The algorithm identified these variables by the presence or absence of combinations of keywords in the ePRF, chosen to maximise the sensitivity and specificity of the algorithm. The algorithm has a very high accuracy found by comparing its output to that of 1000 randomly reviewed entries. Checking with hand coding of 1000 randomly chosen entries we found the algorithm had a high sensitivity and specificity for the following variables: mechanism criteria (99% and 98%), head and neck injury (98% and 97%), thorax injury (97% and 100%) and penetrating injury (98% and 100%).

Outcome

The outcome was major trauma, defined as a patient with Injury Severity Score (ISS)>15. The ISS is a retrospective post imaging anatomical and clinical scoring system. ISS is the current reference standard for major trauma in the UK and is widely used globally to classify patients with multiple injuries(22,23). The ISS score ranges from 1 to 75. ISS is recorded in the Trauma Audit and Research Network (TARN) registry and all patients with ISS>15 injuries are TARN eligible, therefore an ISS≤15 was assumed for all patients that were not linked to data in TARN.

Sample size

Experts recommend a minimum of 100 events and 100 non-events, (24) or 200 events and 200 non-events (25) for external validation samples using logistic regression models. This study adheres to the latter recommendation.

Missing data

Missing data were multiply imputed using chained equations on the basis of a missing at random assumption. Continuous variables were imputed using predictive mean matching when evaluation using diagplots(26) indicated imputations were implausible due to the variable having a non-normal distribution(27,28). The number of imputed datasets was determined by the fraction of missing information. The imputational model included the predictor variables, other relevant variables, such as gender, diastolic BP, respiratory rate, heart rate, oxygen saturation and destination hospital and the outcome. Model performance was averaged for missing data using robust methods(28–30).

Data analysis

Firstly, baseline characteristics of the UK validation cohort compared with the original Dutch prediction cohort are presented as median and interquartile range for continuous variables and number and percentages for categorical variables. The full linear programming describing the published model, including the coefficients and intercept, was provided by the Dutch research team (eFigure 1).

Secondly, the diagnostic accuracy of the South West major triage tool was examined. Current performance in triaging patients with an ISS>15 to MTCs were estimated in the South West region. Undertriage was defined as the proportion of patients with major trauma (ISS>15) not identified, divided by the total number of major trauma patients. Overtriage was defined as the proportion patients without major trauma (ISS<15) identified as having major trauma, divided by the total number of patients without major trauma.

The triage tool in the South West has a number of stages and criteria and is similar to tools used across England(14). First, do serious injuries include certain physiological or anatomical criteria? If yes can airway and catastrophic haemorrhage be safely managed? If yes, transfer to an MTC, if it is less than 60 minutes away. If the injury does not meet the anatomical or physiological criteria, but the clinician remains concerned and it meets above time criteria and the airway or haemorrhage can be managed safely, then transfer to the MTC(14). Triage status was determined for each patient according to whether their recorded data met South West triage criteria; independent of whether they were transported to a trauma centre or trauma unit.

Thirdly, the sensitivity and specificity of the Dutch model was then compared to the current South West triage tool. A maximum specificity of 50% was pre-determined as per the limits recommended by ACS-COT and in line with the approach used by the Dutch researchers.

Fourthly, statistical performance of the Dutch prediction model was assessed by discrimination and calibration(31). Discrimination evaluates whether patients who have the outcome (major trauma) have higher risk predictions calculated by the model than those who do not. For the binary outcome of ISS>15, the C-Statistic, equivalent to the area under the Receiver Operating Characteristic (ROC) curve, was calculated as a measure of how well the model discriminates between those with major trauma and those without major trauma(31). The C statistic is interpreted as the probability that a randomly selected patient who had the outcome will have a higher predicted probability of major trauma than a randomly selected subject who did not experience the outcome. The minimum value of C is 0.0 and the maximum is 1.0. Conventionally, C-values of 0.7 to 0.8 to show acceptable discrimination, values of 0.8 to 0.9 to indicate excellent discrimination, and values of ≥0.9 to show outstanding discrimination. (31).

Calibration measures how closely predictions made by the model match observed outcomes. Calibration was assessed by multiple methods comprising: calculating the ratio of expected versus observed numbers of events (should be close to one if the model calibrates well in the validation dataset); computing the difference between the mean number of predicted outcomes and the mean number of observed outcomes ('calibration-in-the-large', should be close to zero for a well-calibrated model); visual inspection of a calibration plot (a scatter plot of predicted versus observed outcome probabilities, good predictions will be close to the 45° line); and measurement of calibration slope (the overall prognostic effects of predictors in the model, should be close to 1) (31). LOWESS (Locally Weighted Scatterplot Smoothing), a regression analysis, was used to create a smooth line through the calibration scatter plot to help visualisation of relationships or trends.(32) The original model was initially evaluated. Then, to account for the UK validation cohort having a different prevalence of major trauma than in the Dutch studies, resulting in a difference in the baseline risk, the intercept was reestimated(33). The intercept is the expected proportion of major trauma when all the model predictors are zero/negative. This will be lower when the overall cohort prevalence of major trauma is less as observed in the UK cohort. The content and weighting of variables (coefficients) in the Dutch model were not changed.

Decision curve analysis was then used to estimate the net benefit of using the Dutch prediction model to triage patients to an MTC. Net benefit analysis is a statistical technique recommended by the TRIPOD guidelines to help evaluate whether application of prognostic models would aid clinical decision making given a specific patient's risk of an adverse outcome and the clinical consequences of the adverse outcome not being predicted.(19,34) Here we use net benefit analysis to show visually the range of clinically relevant probability thresholds where application of the Dutch prediction model would show benefit to triaging all, or no, patients to an MTC.(35). Net benefit in decision curve analysis is defined as true positives minus false positives x weight(35). In this study, weight was pre-defined as 0.02 (1:50). This was the same as the approach used by the team that developed the Dutch prediction model. It also corresponds with the upper limit of the ACS-COT recommendation for over-triage of 50%.

We also conducted a sub-group analysis of TARN linked patients. We chose to do this to see if the Dutch prediction model would work better than existing triage methods in a high prevalence major trauma cohort. We provide information on the baseline characteristics of the group and how it compares with the derivative cohort and validation cohort. We tested the accuracy of the Dutch prediction model in the TARN linked patients, in terms of discrimination, calibration, sensitivity and specificity, undertriage and overtriage rates and clinical usefulness.

Data were analysed using STATA 16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC).

Ethics

The study received ethical approval from the Yorkshire and The Humber - Bradford Leeds Research Ethics Committee (Reference: 19/YH/0197). As the study involved secondary analysis of routinely collected healthcare data consent was not required.

Results

Participants

74261 patients had injuries and were transported by the SWASFT to hospitals in the South West region over the year. 5462 injured patients aged 15 years or below were excluded. A total of 68799 adult patients with injuries and transported by the SWASFT were included in the final external validation cohort. 1624 patients of these patients were recorded on the TARN registry for this period and linked to the prehospital data set. 524 patients with an ISS>15 were derived from this subset of TARN registry linked patients (figure 1).

Table 1 shows the demographics of our cohort compared to the Dutch derivation cohort. The median age of patients was 72 (i.q.r. 46-84); 55.5% were female; and 524 (0.8%) had an ISS>15. Among patients who had major trauma (ISS>15) the median age was 66 (i.q.r. 43-83) years; 58% were male; the median ISS was 22 (i.q.r. 17-26). In comparison, the Dutch cohort was younger (45 years), more were male (58.3%) and more patients had an ISS>15. (8.8%)

13% of records had one or more predictor variables missing. Multiple imputation (m=13) was performed for gender, age, GCS, respiratory rate, heart rate, systolic blood pressure, diastolic blood pressure, oxygen saturation and destination hospital (table 2).

Analyses

The existing triage tool in the South West achieved a specificity of 84% (95% CI, 81%-87%) and a sensitivity of 44% (95% CI, 40-48%). Using a maximum specificity of 50% for the Dutch model (as per the limits recommended by ACS-COT) led to an improved sensitivity of 83% (95% CI, 80%-86%). The Dutch prediction model was thus able to improve under-triage by 39%, at the expense of increasing over-triage by 34%.

Table 3 and 4 provide further information on how well the model compares with existing methods at identifying those with (true positive) and without (true negative) major trauma and corresponding sensitivity and specificity with 95% confidence intervals.

Validation of the Dutch model in the UK cohort suggested good discrimination (C-Statistic 0.75, 95% CI, 0.73 – 0.78) (figure 2), but sub-optimal prediction of individual patient's probability of having major trauma (calibration) (eFigure 2). The expected to observed events ratio before re-calibration (2.47) and "calibration in the large," (-1.02) indicates the model under-predicted major trauma in the UK validation cohort (Table 5). Re-estimation of the intercept of the Dutch model to adjust for the lower prevalence of major trauma in the UK cohort improved measures of calibration.

However, visually there remains poor agreement between the predicted and observed number of events in population groupings with the model underestimating risk in patients with higher probabilities of major trauma and over estimating risk in those with a very low probability of sustaining severe injuries (eFigure 2).(16)

Figure 3 shows the decision curves and net benefit of using the Dutch prediction model for identifying patients with major trauma (ISS≥15) compared to treating no, or all patients, at MTCs. The x-axis shows the threshold, defined as the ratio between true positives and false positive patients. In this case a threshold of 0.02 means one is willing to accept 49 wrongly classified as having major trauma for every one patient identified with major trauma. The y-axis shows net benefit which is defined as true positives – false positives x weight. Weight was pre-defined as 1:50 or 0.02, as explained in the methods section. Use of the Dutch model showed potential benefit over ‘triage all to an MTC’ and ‘triage none’ alternatives.

The 1624 patients in the sub-group of TARN linked cases showed a higher prevalence of major trauma (32.3%). 43.9% of major trauma patients in the sub-group were treated in a MTC. The Dutch prediction model had the same discriminative ability in the external validation cohort and sub-group with C-Statistic of 0.75 (95% CI, 0.72 – 0.77). The intercept was re-calibrated in the sub-group to account for the difference in prevalence of trauma with the derivative cohort. In the sub-group of TARN eligible patients there was better visual agreement between predicted and observed probability of major trauma across the whole range of risk of severe injury (eFigure 3). The Dutch prediction model in the sub-group had an undertriage rate of 12% and overtriage rate of 50%, which is comparable with the derivation cohort (11.2% and 50%, respectively). eFigure 4 decision curve analysis visually shows the Dutch prediction has potential benefit over ‘triage all to an MTC’ and ‘triage none’ alternatives at a threshold of 0.13. This means one is willing to accept 6.5 patients wrongly classified as having major trauma for every one patient identified with major trauma.

Discussion

Summary of results

The Dutch major trauma prediction model achieved good discrimination with a C-Statistic 0.75 (95% CI, 0.73 – 0.78). After adjusting the model intercept for difference in major trauma prevalence calibration was also good (calibration in the large = 0). However, the model is not as good at predicting higher probabilities of major trauma in our validation cohort. When aiming for a specificity of 50% the model reduced the undertriage rate compared with the existing South West trauma triage tool from 56% to 17%, at the expense of increasing overtriage from 16% to 50%. Overall performance in this external validation cohort was not as good as the derivation cohort and did not meet the ACS-COT under and over triage recommendations(13).

Interpretation of findings

As is often the case, the Dutch prediction model performed less well in external validation than in the original derivation study. There were differences in the inclusion criteria of traumatically injured patients used by the Dutch team to derive the model, compared to this UK validation cohort. The Dutch model only included patients that were determined to be highest priority by the dispatch centre and transported to a trauma centre (level 1-3)(18). In the UK cohort, however, all patients with injuries transported to any hospital in the region were included and so there is a lower prevalence of major trauma. In addition, the external validation cohort was older and had a higher proportion of females. Nevertheless, the Dutch model demonstrated good overall performance and in the sub-group analysis the model appears to perform better when applied to a population with a higher prevalence of trauma.

We were unable to study the effect of paramedic judgment in determining how injured patients are triaged in real-life practice. This could affect the under and over triage rates; however, the evidence is conflicting. One study suggests paramedic judgment is an independent predictor of serious injury(36); others suggest it may not be more accurate than triage tools(37–39). However, clinician gestalt should be factored into the development of new major trauma triage tools or prediction models as it may improve undertriage rates or quantify adherence to trauma triage protocols(18).

False negative assessments of major trauma (leading to undertriage) and false positive assessments (leading to overtriage) are unlikely to have the same impact and the optimal trade-off between sensitivity and specificity will vary according to different health care contexts and resource availability. Ideally, in order to optimise triage tool performance, clinical costs and values would be accounted for to achieve the best balance. Metrics such as the weighted comparison index or net benefit, allow transparent estimation of these trade-offs.

Although an important step in evaluating the accuracy of a triage tool, this external validation is limited by its retrospective design. From a health services perspective, the effectiveness of pre-hospital triage is more important than theoretical accuracy of triage tools. This would be evaluated by whether patients injured closest to a non-specialist hospital are appropriately sent to an MTC and whether the trauma team at the MTC is activated when using the triage tool. This would reflect the extent to which a triage tool is applied in practice, how clinical judgement is used to interpret observed triage tool variable values, and the influence of shared patient decision making or other contextual factors. These results should therefore be interpreted with caution prior to a large-scale implementation study.

Comparison to previous literature

Recent research has shown 59.8% of major trauma patients are treated in non-MTCs in the UK(2). Our study shows similar results with 56.1% of major trauma patients being treated at a non-MTC.

A systematic review of the accuracy of prehospital major trauma triage tools (using ISS>15 as the reference standard for major trauma) was recently conducted. Overall, the studies included had low sensitivity (high under-triage). As such, the reviewed tools were poor at identifying patients with major trauma. The studies included were low quality, which made it difficult to interpret the results.(40)

A recent study assessed the ability of six models to predict major trauma in a dataset from TARN.(22) This was undertaken before the Dutch prediction model had been derived. It found three models were best for predicting major trauma (ISS>15): Kampala Trauma Score (KTS); Physiologic Severity Score (PSS); Prehospital Index (PHI). KTS had an undertriage rate of 3.6% and overtriage rate of 82.8%(22). That validation study only included those patients that met the inclusion criteria for TARN, which means they are at higher risk of trauma than our cohort and may not reflect the wider trauma population that triage tools need to be applied to.

Limitations

The use of retrospective data has several limitations including: missing data and data potentially collected/recorded differently from the derivation cohort. The algorithm we used to provide information on variables from free text is different to how the Dutch emergency medical services recorded data, which could have resulted in misclassification. However, on checking with hand coding of 1000 randomly chosen entries we found the algorithm performed excellently.

Missing data were multiply imputed using recognised techniques and in accordance with international guidelines. Our imputation approach assumed physiological variables were likely missing at random and predicted by other physiological variables and age and sex. Imputation of missing data using these techniques is recommended in international guidelines for validation studies of prognostic models (TRIPOD).

We assumed all patients in the ambulance cohort not linked to the TARN database had an ISS≤15. Linking the SWASFT and TARN databases could have also resulted in information bias, if patients who had an ISS>15 were either not submitted to the TARN registry or not correctly linked. Included hospitals were found to have a data accreditation range of 88.1% to 96.6% and a mean case ascertainment of 86.1% of eligible patients submitted to TARN in this period. We used multiple factors to link patients on the SWAFST and TARN database.

The ISS>15 is contested as reference criteria for major trauma(3,18,22,41,42). Using ISS may overestimate undertriage, because some of the most unwell major trauma patients are taken to the nearest hospital for immediate stabilisation(43). Future external validation studies should compare performance against additional trauma scores such as the New Injury Severity Score (NISS) and whether it predicts mortality or other measures of patient benefit from MTC care, such as need for life saving interventions. Furthermore, future studies should evaluate the prediction model against different definitions of under- and overtriage, e.g. the Cribari method, which focuses on whether or not trauma teams were activated for patients with an ISS>15(44,45).

Implications

Research

New prediction tools need to be easily implemented by ambulance services. There is increasing use of mobile apps to assist with prehospital triage(46,47), and a mobile app based on the Dutch prediction model and which incorporates paramedic judgment is being used in the Netherlands to determine whether it improves trauma undertriage rates(18,48). This mobile app could be prospectively trialled in the UK.

Policy and practice

Our research shows that utilising the Dutch prediction model would result in more patients with an ISS≤15 being triaged to MTCs, in order for a greater number of severely injured patients to be correctly treated at a specialised centre (MTC). The increased bypassing of local hospitals and conveyance of patients to MTCs could result in ambulances being unavailable for other patients. Each regional trauma system would have to determine how much overtriage they are willing to tolerate clinically. The model might have better utility if only used by paramedics in the cohort of patients they suspect of having major trauma. Economic evaluation can help assess whether the benefits of improvements in undertriage rates outweigh the costs incurred by increasing overtriage rates to MTCs.

Conclusion

This external validation using a retrospective cohort showed theoretically the Dutch prediction model for identifying major trauma patients could lower the undertriage rate to 17%, however it would increase the overtriage rate to 50%. Further prospective research is needed to determine whether the model can be practically used by paramedics and whether the model's use is cost-effective.

References

1. Moran CG, Lecky F, Bouamra O, Lawrence T, Edwards A, Woodford M, et al. Changing the System - Major Trauma Patients and Their Outcomes in the NHS (England) 2008–17. EClinicalMedicine. 2018 Aug 1;2–3:13–21.
2. Dixon JR, Lecky F, Bouamra O, Dixon P, Wilson F, Edwards A, et al. Age and the distribution of major injury across a national trauma system. Age Ageing [Internet]. 2020 [cited 2020 May 15];49:218–26. Available from: https://academic.oup.com/ageing/article-abstract/49/2/218/5639746
3. Voskens FJ, Van Rein EAJ, Van Der Sluijs R, Houwert RM, Lichtveld RA, Verleisdonk EJ, et al. Accuracy of prehospital triage in selecting severely injured trauma patients. JAMA Surg. 2018 Apr 1;153(4):322–7.
4. Newgard CD, Zive D, Holmes JF, Bulger EM, Staudenmayer K, Liao M, et al. A multisite assessment of the American College of Surgeons Committee on trauma field triage decision scheme for identifying seriously injured children and adults. J Am Coll Surg. 2011 Dec;213(6):709–21.
5. Celso B, Tepas J, Langland-Orban B, Pracht E, Papa L, Lottenberg L, et al. A systematic review and meta-analysis comparing outcome of severely injured patients treated in trauma centers following the establishment of trauma systems [Internet]. Vol. 60, Journal of Trauma - Injury, Infection and Critical Care. 2006 [cited 2020 Mar 12]. p. 371–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16508498
6. MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, et al. A national evaluation of the effect of trauma-center care on mortality. N Engl J Med. 2006 Jan 26;354(4):366–78.
7. Nathens AB, Jurkovich GJ, Rivara FP, Maier R V. Effectiveness of state trauma

systems in reducing injury-related mortality: A national evaluation. In: Journal of Trauma - Injury, Infection and Critical Care. 2000. p. 25–31.

8. Cameron PA, Gabbe BJ, Cooper DJ, Walker T, Judson R, McNeil J. A statewide system of trauma care in Victoria: Effect on patient survival. Med J Aust. 2008 Nov 17;189(10):546–50.

9. Haut ER, Chang DC, Hayanga AJ, Efron DT, Haider AH, Cornwell EE, et al. Surgeon- and System-Based Influences on Trauma Mortality. Arch Surg [Internet]. 2009;144(8):495–502. Available from: http://archsurg.jamanetwork.com/article.aspx?doi=10.1001/archsurg.2009.100

10. Jeppesen E, Cuevas-Østrem M, Gram-Knutsen C, Uleberg O. Undertriage in trauma: an ignored quality indicator? Scand J Trauma Resusc Emerg Med [Internet]. 2020 Dec 6 [cited 2020 May 15];28(1):34. Available from: https://sjtrem.biomedcentral.com/articles/10.1186/s13049-020-00729-6

11. Newgard CD, Uribe-Leitz T, Haider AH. Undertriage remains a vexing problem for even the most highly developed trauma systems the need for innovations in field triage. Vol. 153, JAMA Surgery. American Medical Association; 2018. p. 328.

12. Uleberg O, Vinjevoll OP, Eriksson U, Aadahl P, Skogvoll E. Overtriage in trauma - What are the causes? Acta Anaesthesiol Scand [Internet]. 2007 Oct [cited 2020 Oct 23];51(9):1178–83. Available from: https://pubmed.ncbi.nlm.nih.gov/17714579/

13. Rotondo M, Cribari C, Smith R, Trauma AC of SC on. Resources for optimal care of the injured patient. American College of Surgeons: Chicago; 2014.

14. Network PT. Major Trauma Triage Tool [Internet]. 2017 [cited 2020 Jan 23]. Available from: http://www.peninsulatraumanetwork.nhs.uk/network-policies

15. van Rein EAJ, van der Sluijs R, Houwert RM, Gunning AC, Lichtveld RA, Leenen LPH, et al. Effectiveness of prehospital trauma triage systems in selecting severely

injured patients: Is comparative analysis possible? Am J Emerg Med [Internet]. 2018;36(6):1060–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29395772

16. Riley RD, Windt D van der, Croft P, Moons KGM. Prognosis research in healthcare : concepts, methods, and impact [Internet]. 2019 [cited 2020 Apr 15]. 354 p. Available from: https://global.oup.com/academic/product/prognosis-research-in-healthcare-9780198796619?cc=pt&lang=en&

17. Rehn M, Perel P, Blackhall K, Lossius HM. Prognostic models for the early care of trauma patients: A systematic review [Internet]. Vol. 19, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine. BioMed Central; 2011 [cited 2020 Apr 15]. p. 17. Available from: http://sjtrem.biomedcentral.com/articles/10.1186/1757-7241-19-17

18. van Rein EAJ, van der Sluijs R, Voskens FJ, Lansink KWW, Houwert RM, Lichtveld RA, et al. Development and Validation of a Prediction Model for Prehospital Triage of Trauma Patients. JAMA Surg [Internet]. 2019 May 1 [cited 2019 Oct 14];154(5):421–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30725101

19. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 2015 Jan 6;162(1):55–63.

20. Newgard CD, Malveau S, Zive D, Lupton J, Lin A. Building A Longitudinal Cohort From 9-1-1 to 1-Year Using Existing Data Sources, Probabilistic Linkage, and Multiple Imputation: A Validation Study. Acad Emerg Med [Internet]. 2018 Nov 1 [cited 2020 Nov 13];25(11):1268–83. Available from: https://pubmed.ncbi.nlm.nih.gov/29969840/

21. TARN. The Trauma Audit & Research Network (TARN) Procedures Manual.

22. Sewalt CA, Venema E, Wiegers EJA, Lecky FE, Schuit SCE, den Hartog D, et al.

Trauma models to identify major trauma and mortality in the prehospital setting. Br J Surg. 2019;

23. Dillon B, Wang W, Bouamra O. Focus on Multiple Trauma A Comparison Study of the Injury Score Models. Eur J Trauma. 2006;32(6):538–85.
24. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol [Internet]. 2005 May [cited 2020 Feb 25];58(5):475–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15845334
25. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. Stat Med. 2016 Jan 30;35(2):214–26.
26. StataCorp. Diagnostic plots -Distributional diagnostic plots [Internet]. [cited 2020 Apr 22]. Available from: https://www.stata.com/manuals13/rdiagnosticplots.pdf
27. StataCorp. Stata: Release 16. Statistical software [Internet]. College Station, Texas; 2019 [cited 2020 Mar 12]. Available from: https://www.stata.com/manuals/mi.pdf
28. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. Vol. 339, BMJ (Online). British Medical Journal Publishing Group; 2009. p. 157–60.
29. Nguyen CD, Carlin JB, Lee KJ. Model checking in multiple imputation: an overview and case study. Emerg Themes Epidemiol 2017 141 [Internet]. 2017;14(1):1–12. Available from: http://ete-online.biomedcentral.com/articles/10.1186/s12982-017-0062-6
30. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. J Clin Epidemiol. 2006

Oct;59(10):1087–91.

31. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. Eur J Cardio-Thoracic Surg [Internet]. 2018;54(2):203–8. Available from: https://academic.oup.com/ejcts/article/54/2/203/4993384

32. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74(368):829–36.

33. Pirracchio R, Ranzani OT. Recalibrating our prediction models in the ICU: Time to move from the abacus to the computer. Vol. 40, Intensive Care Medicine. Springer Verlag; 2014. p. 438–41.

34. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation [Internet]. Vol. 35, European Heart Journal. Oxford University Press; 2014 [cited 2020 Nov 13]. p. 1925–31. Available from: https://pubmed.ncbi.nlm.nih.gov/24898551/

35. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016 Jan 25;352.

36. Newgard CD, Kampp M, Nelson M, Holmes JF, Zive D, Rea T, et al. Deciphering the use and predictive value of “emergency medical services provider judgment” in out-of-hospital trauma triage: A multisite, mixed methods assessment. J Trauma Acute Care Surg. 2012 May;72(5):1239–48.

37. Newgard CD, Kampp M, Nelson M, Holmes JF, Zive D, Rea T, et al. Deciphering the use and predictive value of “emergency medical services provider judgment” in out-of-hospital trauma triage: A multisite, mixed methods assessment. J Trauma Acute Care Surg. 2012 May;72(5):1239–48.

38. Mulholland SA, Gabbe BJ, Cameron P. Is paramedic judgement useful in prehospital trauma triage? Injury. 2005 Nov 1;36(11):1298–305.

39. Cameron PA, Gabbe BJ, Smith K, Mitra B. Triaging the right patient to the right place in the shortest time. Br J Anaesth. 2014;113(2):226–59.

40. van Rein EAJ, Houwert RM, Gunning AC, Lichtveld RA, Leenen LPH, van Heijl M. Accuracy of prehospital triage protocols in selecting severely injured patients. J Trauma Acute Care Surg [Internet]. 2017;83(2):328–39. Available from: http://insights.ovid.com/crossref?an=01586154-201708000-00018

41. Paffrath T, Lefering R, Flohé S. How to define severely injured patients? - An Injury Severity Score (ISS) based approach alone is not sufficient. Injury. 2014;45(Suppl 3):S64–9.

42. Lossius HM, Rehn M, Tjosevik KE, Eken T. Calculating trauma triage precision: effects of different definitions of major trauma. J Trauma Manag Outcomes [Internet]. 2012 Aug 17 [cited 2020 May 15];6(1):9. Available from: http://traumamanagement.biomedcentral.com/articles/10.1186/1752-2897-6-9

43. Centre NCG. Major trauma: service delivery. NICE Guideline NG40 Methods, evidence and recommendations [Internet]. 2016 [cited 2020 May 17]. Available from: https://www.nice.org.uk/guidance/ng40/evidence/full-guideline-pdf-2313258877

44. Peng J, Xiang H. Trauma undertriage and overtriage rates: are we using the wrong formulas? Vol. 34, American Journal of Emergency Medicine. W.B. Saunders; 2016. p. 2191–2.

45. Davis JW, Dirks RC, Sue LP, Kaups KL. Attempting to validate the overtriage/undertriage matrix at a Level i trauma center. In: Journal of Trauma and Acute Care Surgery. Lippincott Williams and Wilkins; 2017. p. 1173–8.

46. Nogueira RG, Silva GS, Lima FO, Yeh YC, Fleming C, Branco D, et al. The FAST-ED App: A Smartphone Platform for the Field Triage of Patients with Stroke. Stroke. 2017 May 1;48(5):1278–84.

47. Lewis TL, Fothergill RT, Karthikesalingam A. Ambulance smartphone tool for field triage of ruptured aortic aneurysms (FILTR): Study protocol for a prospective observational validation of diagnostic accuracy. BMJ Open. 2016 Oct 1;6(10).

48. van der Sluijs R, Fiddelers AAA, Waalwijk JF, Reitsma JB, Dirx MJ, den Hartog D, et al. The impact of the Trauma Triage App on pre-hospital trauma triage: design and protocol of the stepped-wedge, cluster-randomized TESLA trial. Diagnostic Progn Res [Internet]. 2020 Dec 18 [cited 2021 Jan 2];4(1):10. Available from: https://diagnprognres.biomedcentral.com/articles/10.1186/s41512-020-00076-1

Highlights

- First study to externally validate the only empirically derived European prediction model for identifying major trauma in undifferentiated pre-hospital patients.
- Implementing the Dutch model could lower the undertriage rate from 56% to 17%, however it would increase the overtriage rate from 16% to 50%.
- Further research is needed to reduce overtriage and determine whether the model can be practically implemented by paramedics and is cost-effective.

Characteristic	Central Netherlands region (*n* = 4950)	South West region of England (*n* = 68799)	South West region of England ISS>15 (major trauma) (*n* = 524)
Demographics			
Age, years*	45 (27-63)	72 (46-85)	66 (43-83)
Age, 16-64 years	3808 (76.9)	28419 (41.4)	255 (48.9)
Age, 65 years and above	1142 (23.1)	40279 (58.6)	267 (51.2)
Male	2887 (58.3)	30419 (44.5)	301 (58)
Alcohol use	531 (10.7)	6 301 (9.2)	46 (8.8)
Drug use	43 (0.9)	376 (0.6)	3 (0.6)
Physiological characteristics			
Systolic blood pressure, mm Hg*	136 (124-150)	140 (124-157)	138 (121-158)
Diastolic blood pressure, mm Hg*	N/A	80 (72-90)	82 (72-91)
Heart rate, beats per minute*	82 (74-92)	81 (71-93)	83 (73-95)
Respiratory rate, breaths/min*	16 (14-18)	18 (16-20)	18 (18-22)
Oxygen saturation level, %*	98 (96-99)	96 (95-98)	95 (92-98)
GCS score*	15 (15-15)	15 (15-15)	15 (14-15)
Mechanism of injury, No. (%)			
Mechanism criteria	819 (16.5)	8695 (12.6)	203 (38.7)
Fall < 2 meters[£]	N/A	43971 (63.9)	244 (46.6)
Fall > 2 meters	391 (8.0)	1121 (1.6)	47 (9.0)
Fall from stairs	474 (9.6)	4073 (5.9)	69 (13.2)
Motor vehicle collision	406 (8.2)	7618 (11.1)	158 (30.2)
Injury characteristics, No. (%)			
Penetrating injury to head, thorax or abdomen	90 (1.8)	5821 (8.5)	82 (15.6)
Head or neck injury	2635 (53.2)	30499 (44.3)	306 (58.4)
Expected injury in AIS region			
Thorax	719 (14.5)	2767 (4.0)	59 (11.3)
Expected injury in ≥2 AIS regions	1230 (24.8)	490 (0.7)	20 (3.8)
Clinical characteristics			
ISS* ISS+	2 (1-6) 5 (7)	1 (1-1) 1.3 (2.3)	22 (17-26) 23.8 (8.3)
ISS ≥15, No. (%)	435 (8.8)	524 (0.8)	524 (100)
Destination, No. (%)			
Major trauma centre	1724 (34.8)	11020 (16.2)	230 (43.9)
Non-MTC	3226 (65.2)	56812 (83.8)	294 (56.1)

Values before parentheses are number of and in parentheses are percentages unless indicated otherwise. *Values are median (i.q.r.). +values are mean and SD. £Dutch derivative sample had no data for fall >2m.

Table 1: Baseline characteristics

Baseline characteristics of the prediction model cohort (Central Netherlands region), validation cohort (South West region of England, United Kingdom) and validation cohort with the outcome: ISS≥15 (major trauma).

ISS, injury severity score; AIS, abbreviated injury score; MTC, major trauma centre; GCS, Glasgow Comma Score; mm Hg, millimetre of mercury; n, number; i.q.r, inter-quartile range; SD, standard deviation

Variable	Number of missing (n)	Percentage missing (%)
Gender	415	0.6
Age	101	0.1
GCS	2257	3.3
Respiratory rate	4826	7.0
Heart rate	1625	2.4
Oxygen saturation	2745	4.0
Systolic blood pressure	2854	4.1
Diastolic blood pressure	2962	4.3
Destination hospital	443	0.7

Table 2: Missing data for continuous and categorical variables

		Major trauma			
		Yes (ISS>15)	No (ISS<15)	Total	
Triaged to MTC	Yes	435 (TP)	34257 (FP)	34692	Sensitivity (0.83, 95% CI 0.80 – 0.86)
	No	89 (FN)	34018 (TN)	34107	Specificity (0.50, 95% CI 0.46 – 0.54)
	Total	524	68275	68799	

Table 3: 2X2 contingency table showing how the Dutch prediction model performed in identifying major trauma patients and triaging them to Major Trauma Centre (MTC) and identifying non-major trauma patients and triaging them to a Trauma Unit (TU). Sensitivity and specificity are provided with 95% CI.

TP, true positive; FN, false negative; FP, false positive; TN, true negative

		Major trauma			
		Yes (ISS>15)	No (ISS<15)	Total	
Triaged to MTC	Yes	230 (TP)	11005 (FP)	11235	Sensitivity (0.44, 95% CI 0.40 – 0.48)
	No	294 (FN)	56827 (TN)	57121	Specificity (0.84, 95% CI 0.81 – 0.87)
	Total	524	67832	68356	

Table 4: 2X2 contingency table showing existing triage methods in the South West and whether major trauma patients were triaged correctly to MTC or incorrectly to a TU and whether non-major trauma patient was correctly triaged to a TU or incorrectly to a MTC based on ISS score. Sensitivity and specificity are provided with 95% CI. There is missing data for destination hospital.

TP, true positive; FN, false negative; FP, false positive; TN, true negative

Model performance	**Value**	**Recalibrated model performance**	**Value**
C-Statistic	0.75 (95% CI: 0.73 – 0.78)	C-Statistic	0.75 (95% CI: 0.73 – 0.78)
Calibration in the large (CITL)	-1.02	Calibration in the large (CITL)	0.0
Calibration slope[1]	0.81	Calibration slope	0.81
Expected vs. observed patients with ISS>15 stratified by decile (E:O)	2.5	Expected vs. observed (E:O)	1.0

Table 5: Dutch model performance in external validation cohort before and after recalibration.

[1] Calibration slope of logistic regression model with Dutch linear prediction model as the sole predictor using data from our cohort

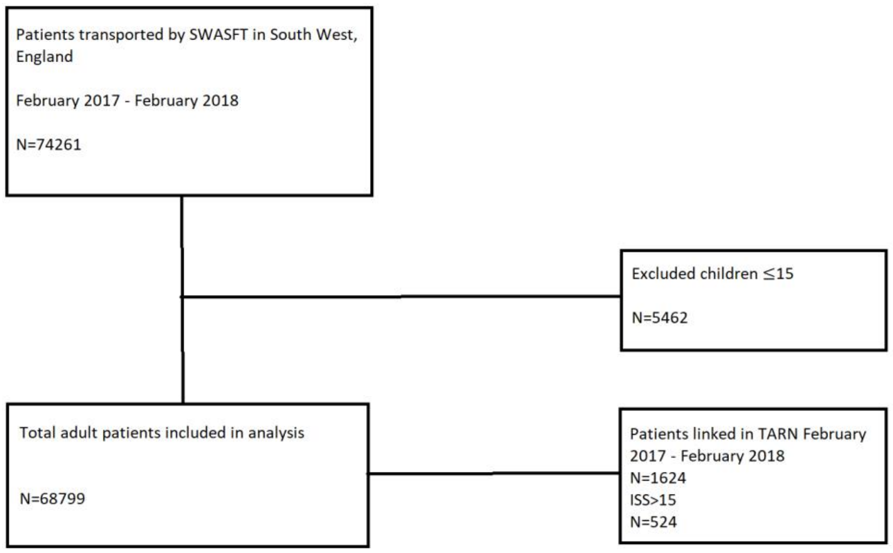

Figure 1: STROBE flow diagram of included and excluded patients.

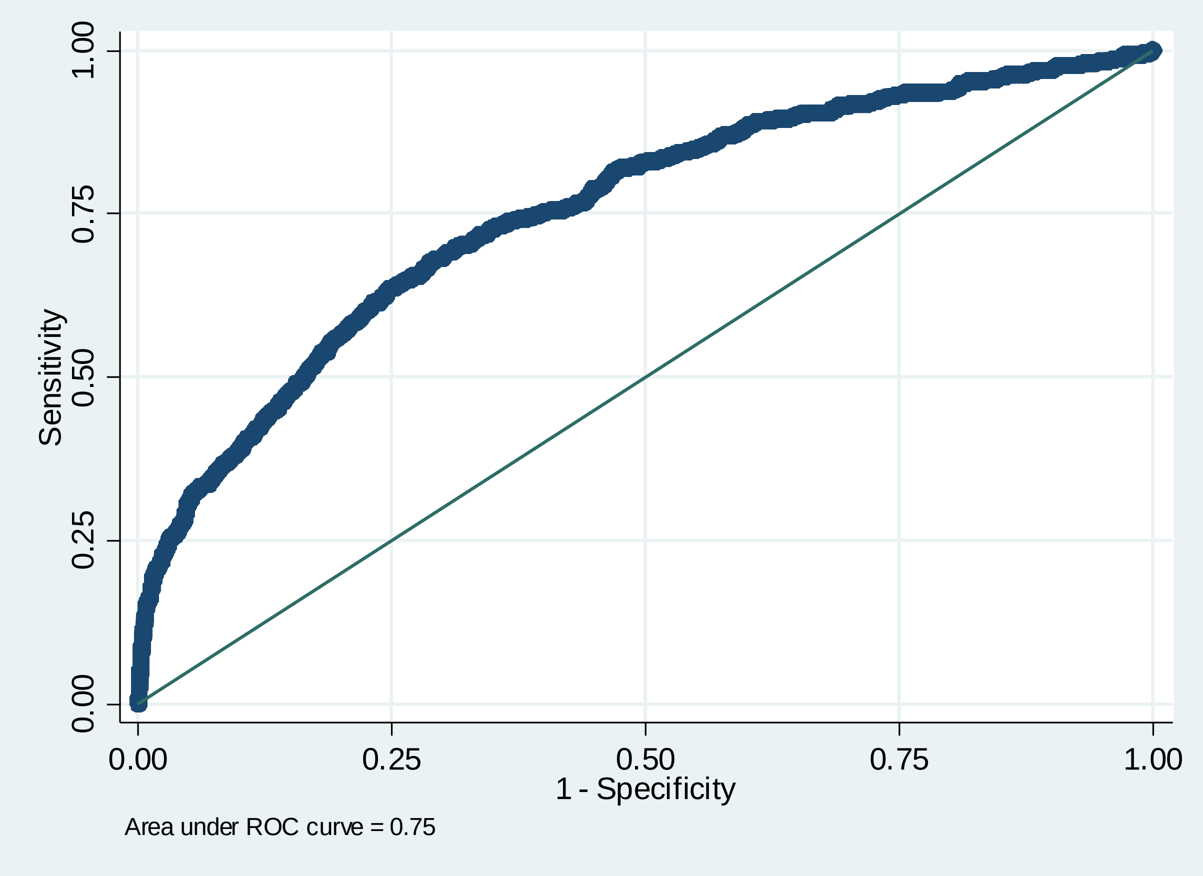

Figure 2: ROC curve for the Dutch prediction model

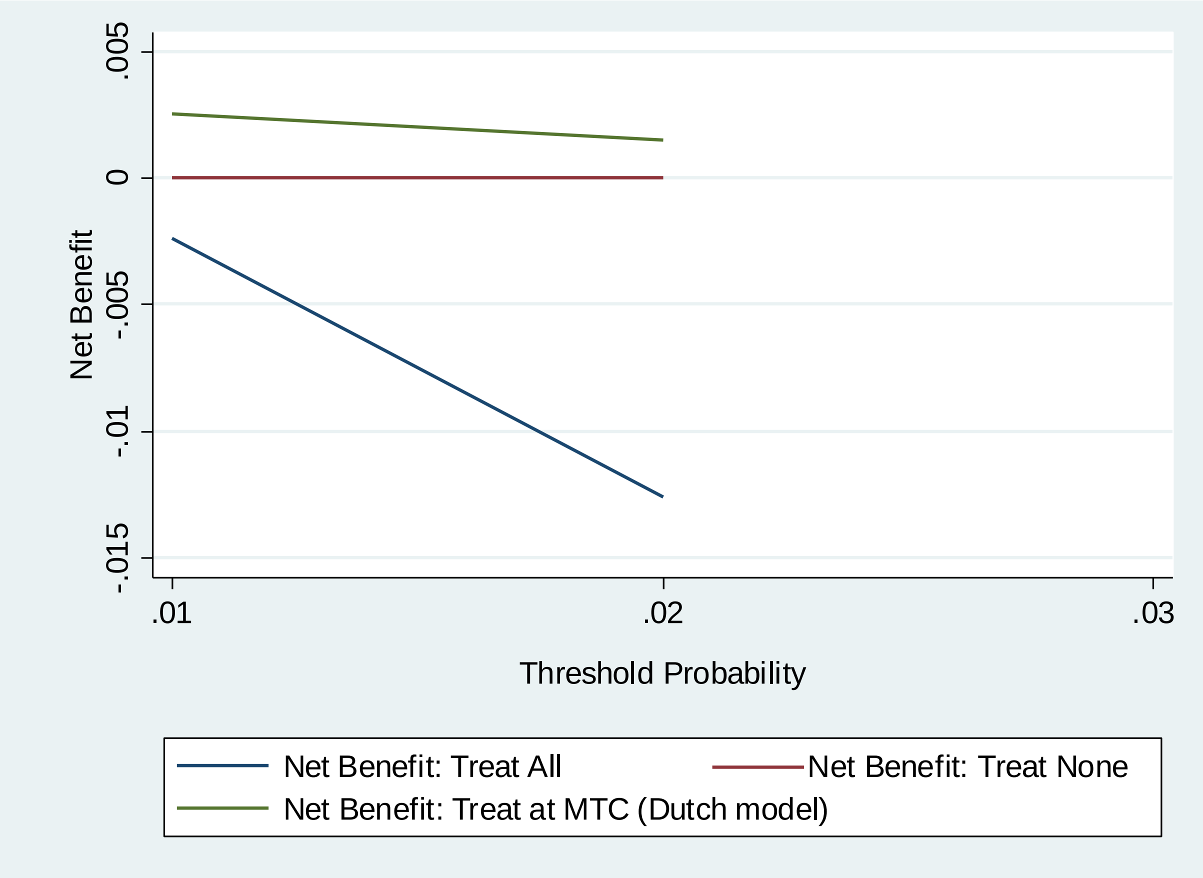

Figure 3: Net benefit curves for the Dutch prediction model for identifying patients with major trauma.