# Hur påverkar överföring och uppdatering av kliniska prediktionsmodeller inom traumatriage feltriageringsgraden?

Bakgrund: Miljontals människor dör varje år till följd av trauma. Traumatriagering är ett sätt att dela in patienter i olika klasser för att bedöma graden av vård som behövs. Forskning inom området har tidigare visat att överföring av modeller kan ha ogynnsamma resultat vad gäller triagering. Syfte: Att granska hur överföring av prediktionsmodeller påverkar feltriageringsgraden när det överförs mellan länder. Även att utvärdera hur uppdatering påverkar feltriagering jämfört med ingen uppdatering. Material och metoder: Patientpopulationen består av patienter över 15 år i tre traumaregister från tre olika länder, Indien, Sverige och USA. En prediktionsmodell skapades genom logistisk regression som beräknar modellprestandan i feltriagering, övertriagering och undertriagering. Värden som saknas utelämnas helt från studien. Resultatet uppvisas som medianer med 95% osäkerhetsintervall(95% UI). Resultat: Feltriageringsgraden varierade mellan modellerna. Sämsta modellprestandan noterades i det amerikanska(NTDB) registret och validerades i det indiska(TITCO) där feltriagering räknat i medianskillnad uppmättes till 0.04 (95% UI -0.221, 0.189). Uppdatering ökade i vissa fall modellprestandan men försämrade i andra. Slutsats: Överföring och uppdatering av modeller kan leda till ökad feltriagering. Amerikanska modeller som används på indiska patienter leder till en större feltriagering. Uppdatering förbättrar i vissa fall modellprestandan men behöver nog anpassas specifikt efter valideringsdatan. Mer studier behövs inom detta område.

**How do transfers and updating of clinical prediction models for trauma triage affect mistriage rates?**

Background: Millions of people die each year as a result of trauma. Triage is a way of dividing trauma patients into different classes to asses the degree of care needed. Research in this area has previously shown that transferring models can have unfavorable results in terms of mistriage. Aim: To asses how transfers of prediction models for trauma triage between different settings affect mistriage rates and to assess how model updating affect these rates compared with no updating. Material and methods: The patient population consist of patients over 15 years of age in three trauma registries from three different countries, India, Sweden and the USA. A prediction model was created with logistic regression which estimates the model performance in mistriage, over- and undertriage rates. Missing values are completely excluded from this study. The results are presented as medians with a 95% uncertainty interval (95% UI). Result: The degree of mistriage varied between the models. The worst model performance was the model developed in the US(NTDB) register and validated in the Indian(TITCO) where the mistriage median difference was 0.04 (95% UI -0.221, 0.189). Updating in some cases increased model performance but deteriorated in others. Conclusion: Transfers and updating of clinical prediction models can lead to increased mistriage rates. US model used on Indian patients lead to greater mistriage rates. Updating in some cases improves model performance but probably needs to be calibrated specifically to the validation data in which it is used. Further studies are warranted in this area.

# Final report

### 2019-12-24

### Salar Mohammed

### 073-920 67 71

### [salar.mohammed@stud.ki.se](mailto:salar.mohammed@stud.ki.se)

Supervisor: Martin Gerdin Wärnberg

[Martin.gerdin@ki.se](mailto:Martin.gerdin@ki.se)

+46708539598

**How do transfers and updating of clinical prediction models for trauma triage affect mistriage rates?**

# Introduction

Trauma accounts for 10% of global mortality (1). More than 4.6 million people die annually due to trauma and it is one of the leading causes of mortality in individuals under 44 years old. Tens of millions of individuals go trough trauma that is non-fatal which require treatment (1,2). One of the tools used in trauma triage is clinical prediction models which will be the main focus targeting to improve and custom in different ways in this study.

Hospital- and prehospital care

The initial contact with the hospital when sustaining a trauma is through the emergency department. If the person is not able to travel to the emergency department, an emergency number is contacted and the ambulance will take the person to hospital to meet healthcare providers. In the ambulance there is instruments and methods to see how well the patient is doing. For example, pulse oximeters and sphygmomanometer. Also, the nurses working in the ambulance are trained in measuring Glasgow Coma Scale(GCS) and to record other vital signs. This is termed prehospital care and it is usually here the care of the patient start after trauma. The health service workers have predetermined protocols on how to manage the patient’s condition and where to take him or her based on the severity of the injury.

When taken to the hospital or when the first contact is through the emergency care department, the patient based on the severity can either see a nurse or a doctor immediately. The doctor or nurse evaluates the patient’s condition and records the parameters in predetermined trauma protocols which determines the level of healthcare given to the patient. The level is scored differently depending on which system is used to classify the patient. In this study, we will split it in major and minor traumas.

Injury severity

The Injury Severity Score (ISS) is an internationally recognized medical scoring system that assesses trauma severity. It correlates with mortality, morbidity, hospitalization time and other measures of severity. It is based on six body regions and is scored from 1 to 75(3). Values >15 has been used as a threshold marker for major trauma (4).

Mistriage

Adequate triage is essential to ensure that trauma patients receive correct administration and hospital care. When a patient is wrongly triaged, this is referred to as mistriage. Mistriage rates is measured as either over- or undertriage. Overtriage is overestimating the urgency of care in terms of resources and workers. When a patient is overtriaged a trauma protocol is being activated in a faulty way which result in the waste of hospital resources. When undertriaged however, a trauma protocol is being activated in a way resulting in an inadequate healthcare-service. The level of undertriage can be seen as a marker of the sensitivity in the trauma system. According to the American College of Surgeons, the degree of undertriage should be less than 5% and overtriage of 25-30% to be satisfactory (5).

Clinical prediction models, transfers and update, mistriage rates and expectation

A component in preventing death and being cost-effective due to trauma is a clinical prediction model. A clinical prediction model can provide information to healthcare providers regarding the probability of different outcomes with help of parameters, when it comes to aiding patients. This can be used as a tool to assist healthcare providers with decision-making and lead to earlier interventions. Prediction models used in trauma care seek to facilitate when prioritizing patients but also to guide treatment decisions, for example massive transfusion (6). Models have and are still being developed to predict death or survival rates in patients. The clinical prediction models have shown to be useful but may decrease in performance when transformed to other settings than the one they were originally developed in. Many models used in trauma triage are built on vital-parameters such as systolic blood pressure (SBP) and respiratory rate (RR) and other variables such as GCS. The variables are later put in a system to determine the level of trauma. One model that is being used as a triage tool is the Revised Trauma Score (RTS) which uses GCS, RR and SBP, often the initial parameters obtained from the patient before they arrive at the emergency care (7). Development of these models are in several cases being made limited to a specified location or setting and are later being used in other circumstances. Also, they are developed on a national level using databases for that specific country and is being used in other parts of the world. nn previous studies it has been shown that Norwegian clinical prediction models are good at predicting survival even in other countries from which the models were not originally created from (8). What has not been heavily studied is the grade of mistriage when using prediction models developed in one country, and applying in another. This study will hopefully cover the knowledge gap and answer how transferring prediction models from a country and applying it in another country affects mistriage rates using trauma registry data from India (TITCO), Sweden (Swetrau) and USA (NTDB).

Updating the prediction models may have an impact on mistriage rates. As mentioned before, the models are being developed in one country and used in another. Mistriage rates may improve if the models are updated(calibrated) using new data from the same setting in which it will be used in. Model updating can be done in various ways. One way is to update a new clinical prediction model for every population and setting with regards to time. A different approach is to update a model on an existing model. The reasoning behind updating the clinical prediction models is that they become miscalibrated over time (9).

We hypothesize that the clinical prediction models used in the same country which they are created from will perform better locally with data from the same country than when they are transferred between countries. Also, we believe that they will increase in performance once they get updated with new data from the same country. In a previous study it was shown that models created in one context and then transferred and used in another context with the same setting perform worse compared to model created and used in the same context (10). This study will hopefully contribute to healthcare providers and others seeking information regarding prediction models and how they perform when transferred and updated so they can make a decision on how to implement these models in a more efficient way.

# Aim

To asses how transfers of prediction models for trauma triage between different settings affect mistriage rates and to assess how model updating affect these rates compared with no updating.

# Material and methods

## Study design

This is a registry-based cohort study with data from the Swedish trauma registry (SweTrau), the US national trauma data bank (NTDB) and the Towards Improved Trauma Care Outcomes in India cohort (TITCO). Each dataset will be divided into samples of three; development, updating and validation samples. Logistic regression will be used to develop the models in the development samples. An estimation will be made of the mistriage rates in the validation samples models and will be compared to it self and to the other validation sample from the other databases. The updating samples will be tested in different settings and compared to see how model updating affect the mistriage rates.

Settings

95.5% of all Swedish hospitals are connected to SweTrau, making it 52 out of 55 hospitals. It holds approximately 55 000 cases. The NTDB holds data from pediatric and adult patients from different levels of trauma centers, including undesignated trauma centers from 2007 to 2017. TITCO collects information from designated trauma centers in India from four centers. These centers are based in large cities in urban India. Kolkata, Mumbai (2-centres) and Delhi (11-14)

## Participants

The **inclusion** criteria in patients registered in SweTrau are:

* All patient that have experienced a traumatic event in which a trauma protocol has been activated in a hospital
* Admitted patients with or without trauma protocol activation
* Patients transferred to the hospital <7 days after the traumatic event and with a NISS score of >15.

**Exclusion** criteria for SweTrau:

* Patients whose only traumatic injury is chronic subdural hematoma
* Trauma protocol activation without an underlying traumatic event(11)

**Inclusion and exclusion criteria** for NTDB (14):

Step #1

Patient NOT INCLUDED in the NTDB

Did the patient sustain one or more traumatic injuries within 14 days of initial hospital encounter?

No ⇒

⇓Yes

Patient NOT INCLUDED in the NTDB

Is the diagnostic code for any injury included in the following ICD-10-CM range?

S00-S99, T07, T14, T20-T28, T30-T32, T79.A1-T79.A9

⇒

⇓

Did the patient sustain at least one injury with a diagnosis code outside the ranges of ICD-10-CM codes below? S00, S10, S20, S30, S40, S50, S60, S70, S80, S90

Patient NOT INCLUDED in the NTDB

⇒

⇓

Continue to step #2

Did the patient’s injury result in death?

Patient included in NTDB

⇒

⇓

Was the patient transferred from one acute care hostpital to another acute care hospital?

Patient included in NTDB

⇒

⇓

Was the patient directly admitted to your hospital(exclude patients with isolated injuries admitted for elective and/or planned surgical intervention)?

Patient included in NTDB

⇒

⇓

Was the patient an in-patient admission and/or observed?

Patient included in NTDB

⇒

⇓

Patient included in NTDB

Was the patient a trauma consult or any level of trauma activation?

⇒

⇓ ⇒

Patient NOT INCLUDED in the NTDB

**Inclusion** criteria for TITCO:

* Patients presenting to the casualty department with injury sustained from road traffic, railway, assault or burns admitted to hospital for treatment
* Patients who died after arrival but before admission

**Exclusion** criteria for TITCO:

* Patients who were dead on arrival (12)

The Convention on the rights of the Child defines a child as a human being below the age of eighteen (15). Individuals 15 years or older registered in all three registers, SweTrau, NTDB and TITCO cohort will be included. We will use >15 years as the cut off because everything below that activates trauma team for pediatrics in many cases but mainly it is because of the difference in physiology amongst children and adults. (16).

## Data sources/measurement

All parameter will be obtained from NTDB, SweTrau and TITCO. SBP, RR, GCS and ISS sex and age have been registered or calculated by workers in hospitals (physicians, nurses, assistant nurse).

## Bias

Possible when coding. The programming will be double checked by an experienced colleague. Confirmation bias is also a risk when conducting data to the analysis program.

## Study size

All patients in SweTrau, NTDB and TITCO over 15 years of age and that fits the inclusion criteria to be in the register.

## Variables

### Study outcome

We will use ISS >15 to define major trauma and ISS ≤15 to define minor trauma. Overtriage will be defined as the event when major trauma is ISS ≤15 calculated by the prediction model. Undertriage will be when ISS > 15 is considered minor trauma by the prediction model. The overtriage rate will be calculated by dividing the number of overtriaged patients with all patients. The undertriage rate will be the number of undertriaged patients divided by all patients. The mistriage rate will be the sum of the overtriage and undertriage rate.

### Model outcome

Mortality within 30 days of the traumatic event. For NTDB and TITCO mortality will be in-hospital mortality. For SweTrau mortality includes out of hospital mortality too.

### Model predictors

SBP, RR and GCS will be used as quantitive variables to develop the models. These are the same parameters used in the Triage Revised Trauma Score (T-RTS).

Respiratory rate

|  |  |
| --- | --- |
| Parameter | Score |
| 10-29 | 4 |
| >29 | 3 |
| 6-9 | 2 |
| 1-5 | 1 |
| 0 | 0 |

Systolic blood pressure

|  |  |
| --- | --- |
| Parameter | Score |
| >89 | 4 |
| 76-89 | 3 |
| 50-75 | 2 |
| 1-49 | 1 |
| 0 | 0 |

Glasgow Coma Scale

|  |  |
| --- | --- |
| Parameter | Score |
| 13-15 | 4 |
| 9-12 | 3 |
| 6-8 | 2 |
| 4-5 | 1 |
| 3 | 0 |

The higher the RTS value, the higher the chance for survival. A lower value is associated with death (17).

## Statistical methods

We will use data from three sources: SweTrau, NTDB and TITCO. Each dataset will be split randomly into three samples: one development, one updating, and one validation sample. Each sample was a random draw from the complete dataset, making sure that patients could only be present in one sample. The size of the samples was calculated to include 300 patients with the outcome, henceforth referred to as events, in the development sample, and 100 events in each of the updating and validation samples, while keeping the proportion of events in each sample approximately the same as in the complete dataset.

The basis for including 300 events in the development sample was that simulation studies have shown that as many as 25 events per parameter and at least as many non-events may be required to achieve stable coefficient estimates when using logistic regression (REF). Our models included 12 parameters, why the number of events needed was 300. Similarly, other simulations show that at least 100 events are required when validating clinical prediction models based on logistic regression (REF).

We will develop one model per development sample, resulting in three models. The performance of each model will be evaluated by testing the model in all three validation samples. The entire process will be repeated 1000 times and results presented as medians and values at the 2.5th and 97.5th percentiles, henceforth referred to as the uncertainty interval (UI). Observations with missing data will be excluded.

Model development

Models will be developed using logistic regression with GCS, RR, and SBP as independent variables and 30 day mortality as the dependent variable. Predictors will be treated as continuous variables with linear associations with mortality. A bootstrap procedure will be used to avoid overfitting and will generate an estimation of a linear shrinkage factor which will be applied to the model coefficients. The model is then going to be used to estimate probability of all cause 30-day mortality. Then, a grid search across estimated probabilities is going to identify the optimal cut-off value for the model.

Model validation

In this step, the model performance will be evaluated. The probability of all cause 30 day mortality will be estimated in each validation sample. Then, the model performance will be assessed using the cut-off value obtained from the development sample. Local performance is defined as a model’s performance in the validation sample from the same dataset as the development sample, for example the SweTrau model’s performance in the SweTrau validation sample. Transferred performance will be defined as a model’s performance in a validation sample from a different dataset compared to the sample in which the model was developed, for example the SweTrau model’s performance in the NTDB and TITCO validation samples.

To assess how model transfer affect mistriage rates the local and transferred performances will be compared by subtracting the local performance from the transferred performance in a pair-wise manner. For example, the SweTrau model’s performance in the NTDB validation sample will be subtracted from the SweTrau model’s performance in the SweTrau validation sample. A positive difference then means that the performance declined when the model was transferred.

Then, each model will be updated in updating samples from datasets in which the model was not developed, the SweTrau model will for example be updated in the NTDB and TITCO updating samples. Numerous methods can be used when updating a prediction model. Recalibration methods can be used and also revision methods (18). We will use the updating sample to estimate a calibration intercept and slope that will then be added to the original model, resulting in an updated model.

Updated performance will then be defined as an updated model’s performance in the validation sample from the same dataset as the updating sample, for example the SweTrau model’s performance in the NTDB validation sample after having been updated in the NTDB updating sample.

To assess how model updating affect mistriage rates compared to no updating the updated performance will be compared with the transferred performance by subtracting the transferred performance from the updated performance. A positive difference means that the updating decreased the performance compared to no updating.

**Ethical considerations**

## Autonomy-respect

The patients can withdraw from the register if they choose to do so. They are not in all cases informed that the information can be used in a study. In that case, we have a responsibility to treat the data with respect like we will do with all data used in this study.

## The principle of beneficence

The study will hopefully improve the management of trauma care and contribute to better healthcare for patients.

## The principle of nonmaleficence

No intervention is being made so there is no risk for physical harm. Data leakage will be the biggest risk for harm and integrity.

## The principle of justice

All patients are depersonalized and anonymous when the data is being obtained. The information gained from the registry will either way be treated equal.

## Ethical Permit

2015/426-31 and 2016/461-32

# Results

The models were developed using logistic regression with predictors treated as continuous variables together with linear associations with mortality. The entire process was repeated 1000 times in a bootstrap procedure and results presented as medians and values is presented as point estimates with 95% Uncertainty Intervals (95% UI).

In total, data was analyzed from [xxxx] patients counting across all three study cohorts NTDB, SweTrau and TITCO. The patients with missing values was completely left out of this study and the total number of excluded patients was [xxxx]. NTDB had the largest amount of patients (n) with 647856 people making it 96.4% of the total study population counting all three cohorts. In all three cohorts, men were overrepresented, most notably in the TITCO cohort with 83.4% of all cases. The age median varied with NTDB having the oldest patients (age median 50 years old) and TITCO the youngest patients (age median 32 years old). SweTrau had the lowest percentage (16.2%) of major trauma in contrast to NTDB (18.1%) and TITCO (23.9%).

*Table 1, Sample characteristics. Data is presented as medians with interquartile range (IQR) or counts with % as applicable*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Characteristic | Level | NTDB | SweTrau | TITCO |
| n (%) |  | 647856 (96.4) | 17284 (2.6) | 6879 (1.0) |
| Age, years (median [IQR]) |  | 50.0 [30.0, 68.0] | 41.0 [25.0, 59.0] | 32.0 [24.0, 45.0] |
| Sex (%) |  | 160 (0.0) | 0 (0.0) | 0 (0.0) |
|  | Female | 239163 (36.9) | 5777 (33.4) | 1141 (16.6) |
|  | Male | 408533 (63.1) | 11507 (66.6) | 5738 (83.4) |
| GCS (median [IQR]) |  | 15.0 [15.0, 15.0] | 15.0 [15.0, 15.0] | 15.0 [9.0, 15.0] |
| TRTS-GCS (%) | 3 | 28789 (4.4) | 210 (1.2) | 358 (5.2) |
|  | 4-5 | 2555 (0.4) | 99 (0.6) | 448 (6.5) |
|  | 6-8 | 8533 (1.3) | 244 (1.4) | 884 (12.9) |
|  | 9-12 | 14641 (2.3) | 579 (3.3) | 870 (12.6) |
|  | 13-15 | 593338 (91.6) | 16152 (93.5) | 4319 (62.8) |
| SBP (median [IQR]) |  | 138.0 [122.0, 154.0] | 138.0 [124.0, 154.0] | 120.0 [108.0, 128.0] |
| TRTS-SBP (%) | 0 | 4846 (0.7) | 67 (0.4) | 87 (1.3) |
|  | 1-49 | 658 (0.1) | 27 (0.2) | 7 (0.1) |
|  | 50-75 | 4292 (0.7) | 107 (0.6) | 176 (2.6) |
|  | 76-89 | 8094 (1.2) | 187 (1.1) | 256 (3.7) |
|  | >89 | 629966 (97.2) | 16896 (97.8) | 6353 (92.4) |
| RR (median [IQR]) |  | 18.0 [16.0, 20.0] | 18.0 [15.0, 21.0] | 18.0 [16.0, 22.0] |
| TRTS-RR (%) | 0 | 9704 (1.5) | 18 (0.1) | 134 (1.9) |
|  | 1-5 | 503 (0.1) | 11 (0.1) | 0 (0.0) |
|  | 6-9 | 2068 (0.3) | 157 (0.9) | 10 (0.1) |
|  | >29 | 11704 (1.8) | 1276 (7.4) | 108 (1.6) |
|  | 10-29 | 623877 (96.3) | 15822 (91.5) | 6627 (96.3) |
| ISS (median [IQR]) |  | 9.0 [4.0, 12.0] | 4.0 [1.0, 10.0] | 10.0 [9.0, 14.0] |
| Major trauma (%) | No | 530681 (81.9) | 14488 (83.8) | 5238 (76.1) |
|  | Yes | 117175 (18.1) | 2796 (16.2) | 1641 (23.9) |
| 30-day survival (%) |  | 95858 (14.8) | 0 (0.0) | 0 (0.0) |
|  | No | 528798 (81.6) | 16643 (96.3) | 5503 (80.0) |
|  | Yes | 23200 (3.6) | 641 (3.7) | 1376 (20.0) |

*Abbreviations and explanations: Age, years, Patient age, years; GCS, Glasgow coma scale; ISS, Injury severity score; RR, Respiratory rate; SBP, Systolic blood pressure; Sex, Patient sex; TRTS-GCS, Triage Revised Trauma Score - Glasgow coma scale; TRTS-RR, Triage Revised Trauma Score - Respiratory rate; TRTS-SBP, Triage Revised Trauma Score - Systolic blood pressure*

Validation and updating

In table 2, the models were first validated in the same sample in which the model was created (local performance) and followed by validation in the other two samples (transferred performance). Then, the models were updated with data from the other two cohorts.

Taking the NTDB model as an example the table shows the model first being validated in NTDB and then validated in SweTrau and after that in TITCO. Continuing the model is updated with data from SweTrau and lastly updated in TITCO. Higher number of the rate indicates bad performance rate. For example, a value of 0.3 implies that the model will wrongly classify 30% of the patients.

Observing local performance in all three cohorts, table 2 shows that SweTrau has the lowest local performance mistriage rate (0.152) while NTDB had 0.177 and TITCO 0.348. The transferred performance shows that both NTDB (0.387(0.373, 0.392)) and SweTrau(0.387(0.373, 0.391)) validated in the TITCO sample performed worse than the others. The best rate in transferred performance can be observed in the TITCO model operating in the SweTrau sample with a mistriage rate of 0.147.

Updating reduced the mistriage rate in the TITCO samples (both NTDB and SweTrau model) from 0.387 to 0.373(NTDB model (0.303, 0.391) and Swetrau(0.373(0.302, 0.391). It deteriorated the model performance most notably in the TITCO model operating in the NTDB sample with an initial score of 0.167(0.164, 0.185) that heightened to 0.178(0.165, 0.186).

Table 2, Under-, over-, and mistriage of models. Data is presented as point estimates with 95% Uncertainty Interval (95% UI)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Sample | Undertriage (95% UI) | Overtriage (95% UI) | Mistriage (95% UI) |
| NTDB | NTDB | 0.123 (0.115, 0.131) | 0.054 (0.050, 0.056) | 0.177 (0.165, 0.187) |
| NTDB | SweTrau | 0.084 (0.082, 0.087) | 0.067 (0.066, 0.072) | 0.153 (0.148, 0.159) |
| NTDB | TITCO | 0.107 (0.102, 0.115) | 0.277 (0.261, 0.289) | 0.387 (0.373, 0.392) |
| NTDB updated in SweTrau | SweTrau | 0.084 (0.082, 0.087) | 0.067 (0.066, 0.072) | 0.153 (0.148, 0.159) |
| NTDB updated in TITCO | TITCO | 0.115 (0.103, 0.131) | 0.260 (0.173, 0.287) | 0.373 (0.303, 0.391) |
| **SweTrau**  SweTrau | NTDB | 0.123 (0.115, 0.132) | 0.052 (0.049, 0.055) | 0.177 (0.165, 0.186) |
| SweTrau | SweTrau | 0.085 (0.083, 0.087) | 0.067 (0.061, 0.072) | 0.152 (0.145, 0.159) |
| SweTrau | TITCO | 0.113 (0.102, 0.115) | 0.275 (0.259, 0.284) | 0.387 (0.373, 0.391) |
| SweTrau updated in NTDB | NTDB | 0.123 (0.115, 0.131) | 0.054 (0.050, 0.055) | 0.177 (0.165, 0.186) |
| SweTrau updated in  **TITCO** | TITCO | 0.118 (0.103, 0.135) | 0.253 (0.172, 0.287) | 0.373 (0.302, 0.391) |
| TITCO | NTDB | 0.132 (0.116, 0.139) | 0.044 (0.030, 0.054) | 0.167 (0.164, 0.185) |
| TITCO | SweTrau | 0.093 (0.085, 0.130) | 0.054 (0.017, 0.067) | 0.147 (0.139, 0.154) |
| TITCO | TITCO | 0.115 (0.107, 0.149) | 0.202 (0.188, 0.266) | 0.348 (0.304, 0.373) |
| TITCO updated in NTDB | NTDB | 0.123 (0.115, 0.131) | 0.053 (0.050, 0.055) | 0.178 (0.165, 0.186) |
| TITCO updated in SweTrau | SweTrau | 0.084 (0.083, 0.087) | 0.067 (0.061, 0.077) | 0.153 (0.143, 0.164) |

Comparison

Table 3 contains information about the model comparison. Here, the local performance is subtracted from the transferred performance giving us either a negative or positive difference. A negative difference means the model improved in performance and a positive difference means it declined.

Table 3, Effects of transferring models between samples. Data is presented as point estimates with 95% Uncertainty Interval (95% UI)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Sample | Undertriage (95% UI) | Overtriage (95% UI) | Mistriage (95% UI) |
| NTDB | SweTrau | 0.035 (0.029, 0.049) | -0.018 (-0.019, -0.010) | 0.018 (0.011, 0.038) |
| NTDB | TITCO | 0.017 (0.003, 0.024) | -0.224 (-0.235, -0.205) | -0.209 (-0.221, -0.189) |
| SweTrau | NTDB | -0.035 (-0.048, -0.029) | 0.015 (0.007, 0.018) | -0.018 (-0.040, -0.011) |
| SweTrau | TITCO | -0.030 (-0.031, -0.015) | -0.210 (-0.213, -0.198) | -0.228 (-0.242, -0.221) |
| TITCO | NTDB | -0.008 (-0.021, 0.024) | 0.172 (0.154, 0.216) | 0.181 (0.137, 0.207) |
| TITCO | SweTrau | 0.021 (-0.014, 0.055) | 0.186 (0.146, 0.199) | 0.202 (0.151, 0.227) |

Updating

Table 4 covers information about the effect of updating the models. The mistriage rate is calculated by subtracting the performance after updating with the performance before updating. A negative difference infers an increase in performance and a positive value infers a decrement. The best performance is the TITCO model operating in the NTDB sample with a value of -0.011(-0.039, 0.003) and the poorest performance in the TITCO model operating in the SweTrau sample with a score of 0.024(0.012, 0.041).

Table 4, Effects of updating models. Data is presented as point estimates with 95% Uncertainty Interval (95% UI)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Sample | Undertriage (95% UI) | Overtriage (95% UI) | Mistriage (95% UI) |
| NTDB | SweTrau | 0.000 (0.000, 0.000) | 0.000 (0.000, 0.000) | 0.000 (0.000, 0.000) |
| NTDB | TITCO | 0.003 (0.000, 0.028) | -0.003 (-0.111, -0.000) | -0.003 (-0.084, -0.000) |
| SweTrau | NTDB | 0.000 (-0.004, 0.000) | 0.000 (0.000, 0.005) | 0.000 (0.000, 0.001) |
| SweTrau | TITCO | 0.012 (-0.010, 0.029) | -0.015 (-0.112, 0.013) | -0.003 (-0.083, 0.003) |
| TITCO | NTDB | -0.048 (-0.056, -0.030) | 0.028 (0.008, 0.044) | -0.011 (-0.039, -0.003) |
| TITCO | SweTrau | 0.025 (-0.007, 0.046) | -0.001 (-0.017, 0.037) | 0.024 (0.012, 0.041) |

Missing values

In total, 17% of the patients was excluded from this study due to missing values. The most common missing value counting all cohorts combined was GCS(gcs\_t\_1) with 10.66% of the missing values. Inspecting the cohorts separately, the most common missing value is RR(rr\_1) for SweTrau and TITCO scoring 33.20% and 41.48% respectively. For NTDB it is GCS 10.69% of the patients.

**SweTrau…Table 5**

|  |  |  |
| --- | --- | --- |
|  | number.of.na | percentage |
| age | 278 | 0.91 |
| sex | 142 | 0.46 |
| gcs\_t\_1 | 2465 | 8.06 |
| sbp\_1 | 3214 | 10.51 |
| rr\_1 | 10152 | 33.20 |
| m30d | 1376 | 4.50 |
| iss | 3 | 0.01 |
| dataset | 0 | 0.00 |

|  |  |  |
| --- | --- | --- |
|  | total.number.of.na | percent.NA |
| 1 | 12203 | 40 |

**TITCO...Table 6**

|  |  |  |
| --- | --- | --- |
|  | number.of.na | percentage |
| age | 21 | 0.13 |
| sex | 0 | 0.00 |
| gcs\_t\_1 | 2126 | 13.29 |
| sbp\_1 | 3255 | 20.34 |
| rr\_1 | 6637 | 41.48 |
| m30d | 22 | 0.14 |
| iss | 1189 | 7.43 |
| dataset | 0 | 0.00 |

|  |  |  |
| --- | --- | --- |
|  | total.number.of.na | percent.NA |
| 1 | 8020 | 50 |

**NTDB…Table 7**

|  |  |  |
| --- | --- | --- |
|  | number.of.na | percentage |
| age | 43748 | 5.69 |
| sex | 0 | 0.00 |
| gcs\_t\_1 | 82172 | 10.69 |
| sbp\_1 | 57926 | 7.54 |
| rr\_1 | 64149 | 8.35 |
| m30d | 0 | 0.00 |
| iss | 69107 | 8.99 |
| dataset | 0 | 0.00 |

|  |  |  |
| --- | --- | --- |
|  | total.number.of.na | percent.NA |
| 1 | 120837 | 16 |

**Combined.cohorts…Table 8**

|  |  |  |
| --- | --- | --- |
|  | number.of.na | percentage |
| age | 44047 | 5.44 |
| sex | 44141 | 5.45 |
| gcs\_t\_1 | 86275 | 10.66 |
| sbp\_1 | 62861 | 7.77 |
| rr\_1 | 78188 | 9.66 |
| m30d | 45312 | 5.60 |
| iss | 70372 | 8.69 |
| dataset | 44047 | 5.44 |

|  |  |  |
| --- | --- | --- |
|  | total.number.of.na | percent.NA |
| 1 | 137438 | 17 |

# Discussion

The objective with the project was to determine how transfers and updates of clinical prediction models would affect trauma triage. The models have been transferred between different countries and the performance was evaluated first locally, and then estimated when transferred. The models local and transferred performance was then calculated estimating over- and undertriage giving us mistriage together. Model comparison was then completed and gave us a mistriage rate with a positive or negative difference. After, the models were updated temporally with new data, developed, validated and compared again.

In all three cohorts, model transfer was performed resulting in uncertainty giving diverse results. The models developed in TITCO and used in NTDB and SweTra has the best rates in terms of mistriage score, the best model being TITCO model validated in SweTrau scoring 0.147. Furthermore, all the models working in the TITCO samples also have the worst rates in terms of mistriage. This includes the local performance of TITCO model. The most significant alteration due to transfer can be noticed in the NTDB and SweTrau model when transferred to TITCO sample. Local performance in TITCO model scored 0.348 locally and when validated by NTDB model and SweTrau model it scores 0.387. This implies the mistriage rate increases by approximately 0.04 when using the NTDB and SweTrau model in the TITCO setting.

One possible reason explaining the rates in the TITCO samples can possibly be the relationship between the definition for major trauma (ISS<15) and the model outcome (probability for 30-day survival). Previous studies have indicated that different injury severity scores that depend on in-depth information may not perform satisfactory when it comes to mortality prediction, especially with incomplete or lacking values (19)(20). The availability for CT-scan and other measurement method are likely not accessible to the same degree as it is in for example Sweden or the United States. The ISS-score is as mentioned before an anatomical-based scoring system and circumstances that complicates CT-scan availability for example can result in incomplete data and therefore an insufficient ISS score lowering the value for ISS and misleads when calculating the mistriage rates.

However, taking the NTDB model for example when transferred to TITCO, in clinical terms this would mean that the transfer would wrongly classify 38.7% of the patients instead of 34.8% if developed locally. Taking 100 trauma patients, this would erroneously classify approximately 4 additional patients. However, inspecting the values furthermore shows us that in both cases, the undertriage rate improved and it was instead the overtriage rate that declined. TITCOs local performance scored 0.202 in overtriage and when transferred to NTDB 0.277 an SweTrau 0.275. Clinically, this would mean that hospital resources would be the biggest loss following this model transfer. Yet, in other cases such as the TITCO model transfer to the NTDB sample, the undertriage rate deteriorated from 0.123 to 0.132 inferring patient safety would be threatened.

When using an algorithm that is established in a setting with a high disease incidence, it may overestimate risks when used in a context with lower incidence (21). Across different context such as regions and countries, patient characteristics and incidence rates differ (22). Tranfering models between different settings should therefore be done cautiously. Using models in different settings in which they were created have also shown to negatively affecting the performance rate of the clinical prediction models in India and in the United States when transferred (23), similarly to the NTDB and SweTrau model when transferred to the TITCO sample in this study. However, a study examining external validation in prediction models for coronary heart disease (CAD) did not show the strong decrease in performance after the model was transferred (24). This can be noticed in this study looking at the TITCO model operating in the NTDB sample with a score of 0.167 compared to the local performance scoring 0.177. Further studies in this area are warranted and transfers should therefore be managed with great caution.

Model updating shows varied results. The performance after updating varied evenly between improvement, no difference and decrement. Notably, the decrease in performance after updating can only be seen in the TITCO models (table 2). Transferred models over time perform inadequate due to bad calibration (23). Poor calibration can make the prediction misleading (25). Patient population change over time due to changes in for example healthcare and treatment policies and following that, the model can be enhanced by updating and re-calibrating it (26) (27). Yet, when developing an optimal prediction model, choice of calibration method should be completed cautiously as the method of calibration depends on availability and performance of validation data (28). In this study, the same updating(calibration) method were used across all the samples. Using specific calibration method for the samples may increase the performance.

Strength and limitations

*Strengths*

Development of clinical prediction models are usually developed in a country with a lot of resources, wealth and access to data and used in developing countries. This study sensibly mirrors how clinical prediction models are developed, transferred and used from one setting to another, realistically mirroring how it is practically implemented in trauma triage around the world. The results give us an indication of the degree of which transfers affect the mistriage rate when developed in a strong population sample for example and used in a smaller.

The model predictors used in this study are established model predictors used in many emergency care units around the world. GCS, RR and SBP are all predictors in the Triage Revised Trauma Score (17).

In similar studies observing the transferability regarding clinical prediction models in different settings, one problem that can be prominent is the amount of patients. Dividing it into different settings can lead to shortage of patients in some settings. In this study, using three different countries and comparing them generated a large number of patients to analyze and work with. This was an advantage because the amount of patient did not become a problem, but rather a benefit.

The coding and statistical analysis is available online (GitHub) for anyone to inspect and review how it has been completed (29). We followed a step-by-step fashion.

*Limitations*

When developing the model, we used the T-RTS predictors as model developing variables. GCS, RR and SBP with coefficient for each variable was used as predictors for 30-day mortality. To enhance the model one could argue that adding predictors to the model might make it increase in performance. For example, age or mechanism of injury. Similarly, the outcome 30-day mortality was used with recognition that late mortality, time in hospital and morbidity also are of significant value in the context of trauma care. However, the purpose of this study was not to modify and create the optimal model for emergency care triage. Instead, it was to evaluate the effect of transferability between different contexts.

The statistical methods were completed with a program called RStudio and the code is available for anyone to analyze and see (10). However, as mentioned in strengths, we followed a step by step fashion and sometimes deviated from the time plan and changes in coding were made in retrospect. This was not optimal.

Clinical Applications

This study realistically reflects the course of action and process in which clinical prediction models are being developed. As mentioned before, clinical prediction models are regularly developed in a country with high-end resources and a big population. Countries like the Canada, USA, United Kingdom (UK), and other countries with a large population and well working health care service are often better at documenting trauma and the protocols are often more complete with less missing values. Healthcare workers from poorer countries then use the clinical prediction model which is both developed and internally validated in a high-resource country. In this study, we develop and transfer the model like it would be done clinically in real life only validating it internally and externally that generates rates for comparison. And the result of the study shows and indicate that transfers of models can lead to increases in mistriage rate which fortify the view that transfers between countries should be done with certain caution and prudence.

Understanding and acknowledging these risks with adequate actions such as validating optimally and calibrating correctly, the model will increase in precision. This modification can possibly reduce patient mortality and decrease the mistriage rate in trauma care.

Regarding equity in the subject area, studies in this subject supports the view that clinical prediction models should not be transferred without cautious thinking due to the fact that the performance might decline. If all patients setting-wise were treated equally in terms of objective parameters and model development, it would trump the idea behind the study. Also, age and sex was not model predictors in this study making the patients equal with the exception of geographical position when divided into samples.

Future studies

In this study, we looked at the outcome when transferring models between countries and how it affects the mistriage rate. A study in the future could potentially evaluate the importance of adding additional model predictors to modify the clinical prediction models. The model can be improved with more information such as age, mechanism of injury or sex.

To study model transfers between countries is one way to evaluate the effect model transfer has on mistriage rates. One could look at it from others settings to contribute to how transfers affect mistriage rates in a larger picture. A similar way is for example to transfer models between socioeconomic different contexts. First developing the model in a setting with high socioeconomic standards and then applying it to a lower socioeconomic setting. Ranking socioeconomic settings can be done by collecting address and insurance status or dividing it to domains containing information about education, wealth and occupation etc. (30) (31).

Successively while broadening the knowledge spectrum of model transfers, one could look at the connection and relationship between different types of model transfers, for example between countries or other settings that differ and see if there are additional factors contributing to model development. This provides with information on what type of model and what calibration method is most suited for a certain setting. Adding further information in this area will make it beneficially when developing a clinical prediction model in a setting or context where trauma data is lacking.

**Conclusion**

In conclusion, model transfer between countries according to this study lead to uncertain outcomes of varying degrees. In some cases, the performance decreased and in other cases increased. Potentially, it can lead to increased rates of mistriage with decline in performance in undertriage as well as overtriage making it unpredictable and not safe.

Model updating also had varying outcomes. However, the method of updating may suit one sample but not another. Further studies of this kind are therefore needed in order to to reduce patient mortality and resource waste following poor model transfer and poor calibration.

**References**

1. 1.World Health Organization. Injuries and violence: The facts. Geneva: World Health Organization; 2014
2. Naghavi, M. (2019). *Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016*. [online] The Lancet. Available at: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)32152-9/fulltext [Accessed 27 Oct. 2019].
3. Injury Severity Score | Institute of Trauma and Injury Management | ACI [Internet]. Aci.health.nsw.gov.au. 2019 [cited 25 October 2019]. Available from: https://www.aci.health.nsw.gov.au/get-involved/institute-of-trauma-and-injury-management/Data/injury-scoring/injury\_severity\_score
4. Ogilvie R. Incidence and outcomes of major trauma patients managed in the Australian Capital Territory. - PubMed - NCBI [Internet]. Ncbi.nlm.nih.gov. 2019 [cited 25 October 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24456378
5. Trauma, C.o. *Resources for Optimal Care of the Injured Patient* 2014. 2014; Available from: <https://www.facs.org/~/media/files/quality%20programs/trauma/vrc%20resources/resources%20for%20optimal%20care.ashx> [cited 25 october 2019]
6. Maegele, M. (2019). *Predictive Models and Algorithms for the Need of Transfusion Including Massive Transfusion in Severely Injured Patients*. [online] Pubmed l NCBI. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3364090/ [Accessed 27 Oct. 2019].
7. xJ.M.JONES. (2019). Norwegian survival prediction model in trauma: modelling effects of anatomic injury, acute physiology, age, and co-morbidity. Wiley Online Library [online] available at: https://onlinelibrary.wiley.com/doi/full/10.1111/aas.12256
8. . Ghorbani P e. Validation of the Norwegian survival prediction model in trauma (NORMIT) in Swedish trauma populations. - PubMed - NCBI [Internet]. Ncbi.nlm.nih.gov. 2019 [cited 25 September 2019]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Validation+of+the+Norwegian+survival+prediction+model+in+trauma+(NORMIT)+in+Swedish+trauma+populations>
9. Sperrin M. A review of statistical updating methods for clinical prediction models - Ting-Li Su, Thomas Jaki, Graeme L Hickey, Iain Buchan, Matthew Sperrin, 2018 [Internet]. SAGE Journals. 2019 [cited 25 October 2019]. Available from: https://journals.sagepub.com/doi/full/10.1177/0962280215626466?url\_ver=Z39.88-2003&rfr\_id=ori%3Arid%3Acrossref.org&rfr\_dat=cr\_pub%3Dpubmed&
10. Martin H. 613Martin/transfer-effect-mistriage [Internet]. GitHub. 2019 [cited 25 October 2019]. Available from: https://github.com/613Martin/transfer-effect-mistriage
11. .Rcsyd.se. (2019). Om SweTrau | SweTrau. [online] Available at: http://rcsyd.se/swetrau/om-rc-syd [Accessed 13 Sep. 2019]
12. Sites.google.com. (2019). About TITCO - India - TITCO-India. [online] Available at: https://www.sites.google.com/site/titcoindia/about-titco [Accessed 13 Sep. 2019]
13. Trauma Quality Programs Participant Use File [Internet]. American College of Surgeons. 2019 [cited 16 September 2019]. Available from: https://www.facs.org/quality-programs/trauma/tqp/center-programs/ntdb/datasets
14. [Internet]. Facs.org. 2019 page 5 [cited 16 September 2019]. Available from: https://www.facs.org/-/media/files/quality-programs/trauma/ntdb/ntds/data-dictionaries/ntds\_data\_dictionary\_2020.ashx?la=en
15. OHCHR | Convention on the Rights of the Child [Internet]. Ohchr.org. 2019 [cited 1 October 2019]. Available from: https://www.ohchr.org/en/professionalinterest/pages/crc.aspx
16. School, M. (2019). Trauma Team Activation for Pediatrics (age 15 years and below) | Department of Surgery | McGovern Medical School. [online] Med.uth.edu. Available at: https://med.uth.edu/surgery/trauma-team-activation-for-pediatrics-age-15-years-and-below/ [Accessed 13 Sep. 2019].
17. TRAUMA.ORG: Trauma Scoring: Revised Trauma Score [Internet]. Trauma.org. 2019 [cited 1 October 2019]. Available from: http://www.trauma.org/archive/scores/rts.html
18. Janssen KJ e. Updating methods improved the performance of a clinical prediction model in new patients. - PubMed - NCBI [Internet]. Ncbi.nlm.nih.gov. 2019 [cited 1 October 2019]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/18083464>
19. Roy N, Gerdin M, Schneider E, et al. Validation of international trauma scoring systems in urban trauma centres in India. Injury 2016;47:2459–64. 10.1016/j.injury.2016.09.027 [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/27667119)] [[CrossRef](https://dx.doi.org/10.1016%2Fj.injury.2016.09.027" \t "pmc_ext)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Injury&title=Validation+of+international+trauma+scoring+systems+in+urban+trauma+centres+in+India&author=N+Roy&author=M+Gerdin&author=E+Schneider&volume=47&publication_year=2016&pages=2459-64&pmid=27667119&doi=10.1016/j.injury.2016.09.027&)]
20. Hung YW, He H, Mehmood A, et al. Exploring injury severity measures and in-hospital mortality: A multi-hospital study in Kenya. Injury 2017;48:2112–8. 10.1016/j.injury.2017.07.001 [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/28716210)] [[CrossRef](https://dx.doi.org/10.1016%2Fj.injury.2017.07.001" \t "pmc_ext)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Injury&title=Exploring+injury+severity+measures+and+in-hospital+mortality:+A+multi-hospital+study+in+Kenya&author=YW+Hung&author=H+He&author=A+Mehmood&volume=48&publication_year=2017&pages=2112-8&pmid=28716210&doi=10.1016/j.injury.2017.07.001&)]
21. Testa A, Kaijser J, Wynants L, et al. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. Br J Cancer. 2014;111:680–8
22. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424
23. Gerdin M, Roy N, Fellander-Tsai L, Tomson G, von Schreeb J, Petzold M, et al. Traumatic transfers: calibration is adversely affected when prediction models are transferred between trauma care contexts in India and the United States. J Clin Epidemiol. 2016;74:177-86
24. Genders TSS, Coles A, Hoffmann U, Patel MR, Mark DB, Lee KL, et al. The External Validity of Prediction Models for the Diagnosis of Obstructive Coronary Artery Disease in Patients With Stable Chest Pain: Insights From the PROMISE Trial. JACC Cardiovasc Imaging. 2018;11(3):437-46.
25. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Mak. 2015;35:162–9
26. Van Calster, B., McLernon, D.J., van Smeden, M. *et al.* Calibration: the Achilles heel of predictive analytics. *BMC Med* **17,**230 (2019) doi:10.1186/s12916-019-1466-7
27. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc. 2017;24:1052–61
28. 5. Walsh C. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk [Internet]. Sciencedirect. 2017 [cited 24 December 2019]. Available from: https://www.sciencedirect.com/science/article/pii/S1532046417302277
29. Salar M. Sal999/transfer-between-settings [Internet]. GitHub. 2019 [cited 24 December 2019]. Available from: https://github.com/Sal999/transfer-between-settings
30. Bhavsar N, Aijing G. Value of Neighborhood Socioeconomic Status in Predicting Risk of Outcomes in Studies That Use Electronic Health Record Data [Internet]. Pubmed - NCBI. 2018 [cited 18 December 2019]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324505/
31. . Bilal U. Association of neighbourhood socioeconomic status and diabetes burden using electronic health records in Madrid (Spain): the HeartHealthyHoods study [Internet]. BMJ open. 2018 [cited 18 December 2019]. Available from: https://bmjopen.bmj.com/content/8/9/e021143