Transferability of clinical prediction models for early trauma care

Martin Henriksson

Dell D Saulnier

Johanna Berg

Martin Gerdin Wärnberg

September 28, 2019

**Abstract**

*Background:* Trauma results in approximately four million deaths annually. Clinical prediction models are often used for triage in early trauma care. Previous research has indicated that transferring these models between different settings might adversely affect model performance.

*Methods:* Patients aged 15 years or older who were registered between 2011 and 2016 in the Swedish national trauma registry, SweTrau, were included. Three sets of data were created: high- and low-volume centres, metropolitan and non-metropolitan centres, and multiple and single centres. Clinical prediction models were developed using logistic regression in each data set and transferred within sets. Model performance was evaluated in all data sets using the mistriage rate, undertriage rate and overtriage rate. Multiple imputation using chained equations was used to handle missing data. Model performance was reported as medians with the 95% uncertainty interval across imputed data sets.

*Results:* A total of 26965 patients were included. Model validation mistriage rate ranged from 0.18 (0.18 - 0.44) to 0.64 (0.47 - 0.73), and the transferred model mistriage rate ranged from 0.22 (0.24 - 0.42) to 0.67 (0.46 - 0.79), across all samples. The worst performance after transfer was the low-volume model transferred to the high-volume sample, resulting in a significant mistriage rate increase of 0.44 (0.17 - 0.60), primarily due to increased overtriage.

*Conclusions:* In all samples, model transfer was associated with some degree of uncertainty. Models developed in samples with high numbers of patients performed best after transfer, and the model developed in the low-volume sample performed worst after transfer.

**Keywords:** Trauma, Clinical prediction model, Transferability, Mistriage, Undertriage, Overtriage

# Introduction

Trauma, defined as physical injuries to a host by outside objects [1], accounts for approximately 458.5 million cases annually across the globe [2] and around four million deaths [3, 4]. Each year, 9% of global deaths are the result of trauma, with the leading causes being road traffic accidents, suicide, and homicide. Predictions indicate that the incidence of these causes is likely to increase by 2030 [5].

In a typical high-resource setting, the initial management of trauma is performed on the scene by Emergency Medical Services (EMS). Patient data and vital signs are transferred to the receiving hospital. This information is then evaluated using a system to determine the level of trauma, prepare adequate resources [6] and dictate whether a full or limited trauma team is activated [7].

Systems that determine the level of trauma during early trauma care can be based on clinical prediction models. Models differ in quality and characteristics but generally perform well at predicting survival [8]. Many models are developed in a single, standardized context, such as a major trauma centre, but are then implemented in different contexts, such as non-trauma centres.

What is not fully understood is how this transfer affects model performance. Previous research has shown that model performance in terms of calibration can be adversely affected [9]. However, the aforementioned research studied model transfers between substantially different settings and did not assess more clinically relevant performance measures, such as misclassification.

In trauma, misclassification is often referred to as mistriage. Triage refers to the classification of trauma severity as minor or major. Mistriage can be subdivided into overtriage, which is the incorrect classification of a patient with minor trauma as one with major trauma, or undertriage, which is the incorrect classification of a patient with major trauma as one with minor trauma. Mistriage can ultimately lead to decreased patient survival and is also detrimental to patient care and the distribution of resources [7].

Thus, the effect of model transfers between care contexts within a single healthcare system, as well as the effect of such transfers on mistriage, have not been studied and represent substantial knowledge gaps. The aim of this study was to assess how transfers of clinical prediction models for early trauma care between different care contexts within a single health system affect mistriage rates.

# Materials and methods

## Study design

A registry-based cohort study was conducted. Sweden has a nationally encompassing trauma registry called SweTrau. SweTrau data were used to create clinical prediction models, which were then transferred between different care contexts to study the effects on mistriage. The study and analysis plans were made publicly available before the research was undertaken [10].

## Setting

In Sweden, trauma is the most common cause of death before 44 years of age, and 10% of major trauma patients suffer some form of disability. The societal effects of trauma are substantial, with the loss of days at work comparable to those lost from cardiovascular and malignant diseases combined [11].

Trauma care in Sweden is usually clearly defined in major metropolitan areas: trauma patients are transported directly to a predesignated hospital with specific trauma competency. In more rural parts of the country, trauma patients are instead transported to the nearest hospital for stabilization. Once stabilized, they may then be transported to a hospital with more advanced trauma care capacity [12]. Most hospitals in Sweden use some system to categorize the level of trauma, usually based on patient vital signs and the mechanism of injury, as reported by the EMS or as registered in the emergency department [13].

Today, 52 of the 55 Swedish hospitals record trauma cases in SweTrau. Currently, the registry includes 55 000 cases, and its use is encouraged for both academic research and local quality improvement initiatives [14].

## Participants

The eligibility criteria were adult patients aged 15 years or above registered in SweTrau between 2011 and 2016. The study aimed to assess adult trauma and not paediatric trauma, which differs in physiology, triage and initial care [15]. The Swedish guidelines for trauma activation define children as individuals aged <15 years [13]. Further, patients with missing date and time of trauma were excluded.

## Variables

### Participant characteristics

To describe the patient cohort, we reported age, sex, Injury Severity Score (ISS) and New ISS (NISS).

### Model predictors

The clinical prediction models used the following predictors on arrival at the hospital: systolic blood pressure (SBP), respiratory rate (RR) and Glasgow coma scale (GCS). The rationale for using these three predictors is that they are part of multiple established clinical prediction models for early trauma care, such as the Revised Trauma Score [16].

### Model outcome

The outcome used to develop the clinical prediction models was all-cause mortality within 30 days of trauma.

### Study outcome

ISS > 15 was used as the gold standard to define trauma severity as major trauma, and hence, patients with ISS ≤ 15 were considered to have minor trauma [17]. Overtriage was defined as the event when a clinical prediction model classified a patient with ISS ≤ 15 as having major trauma, and undertriage was defined as the event when a clinical prediction model classified a patient with ISS > 15 as having minor trauma. The overtriage rate was defined as the number of overtriaged patients divided by the total number of patients. The undertriage rate was defined as the number of undertriaged patients divided by the total number of patients. The mistriage rate was defined as the sum of the over- and undertriage rates.

## Data sources and measurements

Model predictors, outcome, participant characteristics and study outcome as outlined above were all obtained from SweTrau. The method of measurement of the model predictors is not specified in the registry entries. In Swedish emergency rooms, patient parameters are usually obtained by a registered nurse or assistant nurse and are assumed to be accurate. Whether the patient is dead or alive 30 days after trauma is manually entered into the registry by each respective hospital. The NISS and ISS are calculated by hospital personnel based on the patient’s injuries.

## Bias

Data analysis was conducted according to a prearranged analysis plan. The analysis plan and statistical analysis code were finalised using simulated data. These efforts were made to avoid confirmation bias. The analysis plan was reviewed by an experienced statistician and programmer prior to implementation. Neither the outcome nor variables were blinded during analysis, which made a structured approach important to ensure objectivity.

## Study size

All patients matching the eligibility criteria were included. Four data sets were used to study the transfer of clinical prediction models, with each data set representing a different care context. The data sets and sample size considerations are outlined below.

## Quantitative variables

GCS, SBP, and RR were modelled as continuous linear terms. In the original study plan it was specified that SBP and RR would be modelled using restricted cubic splines with four knots. This approach however resulted in an unstable imputation model and we therefore decided to simplify the analysis model. ISS was presented as dichotomous, using ISS > 15 as the cutoff.

## Statistical methods

### Data sets

The complete SweTrau cohort was split into three overlapping sets of data, each representing a plausible care context for model transfer. These data sets were further divided into samples as follows:

### Data set 1: High- and low-volume centres

Two samples were derived from the SweTrau cohort based on the number of patients. High-volume centres were those within the top quartile in terms of the number of patients received. The rest were low-volume centres.

### Data set 2: Metropolitan and non-metropolitan centres

This data set was also split into two samples: a metropolitan sample consisting of patients from greater Stockholm, greater Gothenburg and greater Malmö, as defined by Swedish statistics, and a second sample consisting of patients from non-metropolitan areas.

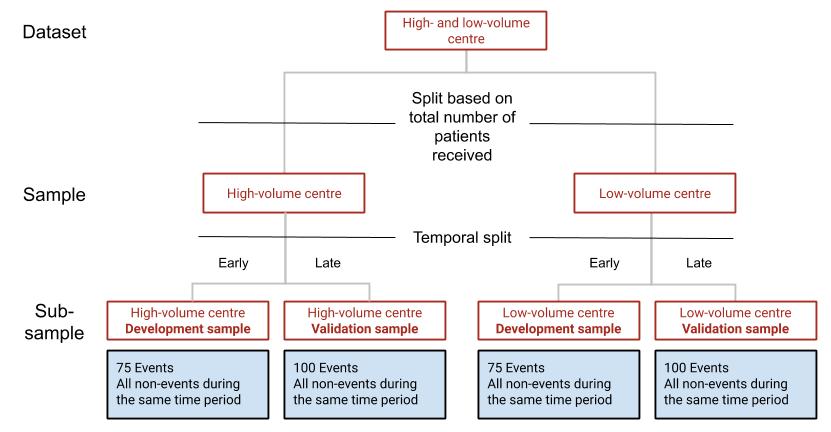
### Data set 3: Multiple and single centre data

In this data set, two samples were created. One centre with a large enough sample sizes to develop and validate a model was its own sample. The multiple centres sample consisted of the combined data from all centres.

A fourth dataset was initially created, consisting of individual centres that had large enough sample size to develop and validate a model. However, this data set was removed before analysis due to a lack of centres with a sufficiently large sample size.

### Development and validation sample

The samples in each data set were split into two subsamples using a temporal split based on the date of trauma. The earlier subsample was the development sample, and the later subsample was the validation sample. The development sample contained 75 events (events being patients who died within 30 days of the trauma) and all non-events (non-events being patients who survived 30 days past the trauma) during the same time. The rationale for including 75 events was that 25 events per variable have been shown to result in less biased coefficient estimates compared to the more standard 10 events per variable [18, 19]. The validation sample contained 100 events and at least 100 non-events, which was suggested as the minimum number by Vergouwe in 2005 for external validation samples [20]. Figure 1 shows the development of the samples using the high- and low-volume centre data set as an example. The minimum sample sizes of the development and validation samples were 150 and 200, respectively. We performed analyses only on data sets for which all samples included at least the minimum sample size.

  
**Fig. 1: Sample development in the high- and low-volume centre data set.** Initial split based on the number of patients. Temporal split made using the date of the trauma.

### Sequence of analysis

The programming language R was used for all analyses [21], and all code has been made publicly available [10]. Analyses were performed in the sequence of model development, model validation and finally model comparison. These steps were repeated in each data set. Below, we use the transfer of a model from the high-volume centre sample to the low-volume centre sample as an example to describe the complete procedure.

### Model development

In the model development step, a clinical prediction model was developed in the high-volume centre development sample. The model was developed using logistic regression. The dependent variable was all-cause mortality within 30 days of trauma, and the independent variables SBP, RR, and GCS were modelled as previously described. To avoid overfitting the model, a bootstrap procedure was used to estimate a linear shrinkage factor that was then applied to the model coefficients [22]. The shrunk model was used to estimate the probability of all-cause 30-day mortality in the development sample. A grid search was performed across estimated probabilities in the development sample to identify the cutoff value that optimised overtriage while keeping undertriage at less than 5% [23]. This cutoff value was then used to classify patients as having major or minor trauma.

### Model validation

In the model validation step, the model performance was assessed in the high-volume centre validation sample and in the low-volume centre validation sample. First, the model was used to estimate the probability of all-cause 30-day mortality in each sample. Then, the probability cutoff value identified in the development sample was applied to the validation samples, patients were classified as having major or minor trauma, and the model performance was estimated.

### Model comparison

Finally, the difference in model performance between the high- and low-volume centre validation samples was calculated. Empirical bootstrapping was used to estimate the 95% confidence intervals (CI) around the performance and the differences in the performance estimates. Both bootstrap procedures used 1 000 bootstrap samples drawn with replacement of the same size as the original samples. The three steps of model development, model validation, and model comparison were repeated in all four sets of data.

### Performance measures

Model performance was assessed in terms of over-, under-, and mistriage rates as defined above.

### Missing data

We used multiple imputation using chained equations (mice) to handle missing data [24]. The number of imputations created for each data set was equal to the highest percentage of missing data in that data set. Quantitative variables were imputed using predictive mean matching, and qualitative variables were imputed using logistic regression. All analyses outlined above were then conducted separately in each imputed dataset. We present the combined results as the median point estimate across imputations along with an empirical bootstrap of the 25th and 75th percentiles across imputations, i.e. the lower bound of the presented interval is the lower bound of a 95% CI of the 25th percentile, and the upper bound is the upper bound of a 95% CI of the 75th percentile. This combined CI was referred to as an uncertainty interval (UI) and was used to express the added uncertainty associated with the imputation procedure to handle missing data; as such, it is more conservative than a standard 95% CI.

## Ethical considerations

The study was approved by the regional ethics review board in Stockholm, Sweden. The ethics review numbers are 2015/426-31 and 2016/461-32. All patients in SweTrau received letters detailing their inclusion in the database and the possibility of their data being used for scientific purposes. It was assumed that the patients who consented to the use of their SweTrau data for research wanted to contribute to the improvement of care, and thus, by performing this study, the authors hope to honour their wishes.

# Results

We analysed data from 26965 trauma patients to investigate the effects of transfer on clinical prediction models (Table 1), after excluding 78 patients with missing date and time of trauma. The total number of missing observations across all variables was 9984 in the entire study cohort. The sample with the highest percentage of missing observations was the non-metropolitan sample, with 48% incomplete observations. The variable with the highest number of missing values was the ed\_rr\_value, with 8296 missing values, or 31% of the total values for this variable. The percentages of missing values for the other model predictors were 8% for GCS and 9% for SBP.

**Table 1. Sample characteristics of multiple imputed data.** Data are presented as medians with interquartile range (IQR) or counts with %, as applicable.

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Level | High volume | Low volume |
| n (%) |  | 20021 (74.2) | 6944 (25.8) |
| Age, years (median [IQR]) |  | 40.0 [25.0, 57.0] | 46.0 [27.0, 64.0] |
| Sex (%) | Male | 13134.5 (65.6) | 4680.5 (67.4) |
|  | Female | 6886.5 (34.4) | 2263.5 (32.6) |
| GCS (median [IQR]) |  | 15.0 [15.0, 15.0] | 15.0 [14.0, 15.0] |
| SBP (median [IQR]) |  | 139.0 [125.0, 153.0] | 136.0 [120.0, 152.0] |
| RR (median [IQR]) |  | 18.0 [15.0, 20.0] | 19.0 [16.0, 22.0] |
| 30-day survival (%) | Alive | 19270.9 (96.3) | 6546.5 (94.3) |
|  | Dead | 750.1 (3.7) | 397.5 (5.7) |
| ISS (median [IQR]) |  | 3.0 [1.0, 9.0] | 5.0 [1.0, 13.0] |
| NISS (median [IQR]) |  | 3.0 [1.0, 12.0] | 6.0 [1.0, 17.0] |
| ISS>15 (%) | No | 17343.7 (86.6) | 5530.0 (79.6) |
|  | Yes | 2677.3 (13.4) | 1414.0 (20.4) |

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Level | Metropolitan | Non-metropolitan |
| n (%) |  | 13042 (48.4) | 13923 (51.6) |
| Age, years (median [IQR]) |  | 41.0 [26.0, 58.0] | 42.0 [25.0, 60.0] |
| Sex (%) | Male | 8804.0 (67.5) | 9010.4 (64.7) |
|  | Female | 4238.0 (32.5) | 4912.6 (35.3) |
| GCS (median [IQR]) |  | 15.0 [15.0, 15.0] | 15.0 [15.0, 15.0] |
| SBP (median [IQR]) |  | 140.0 [125.0, 156.0] | 135.0 [121.0, 150.0] |
| RR (median [IQR]) |  | 18.0 [15.0, 20.0] | 18.0 [16.0, 21.0] |
| 30-day survival (%) | Alive | 12428.1 (95.3) | 13397.4 (96.2) |
|  | Dead | 613.9 (4.7) | 525.6 (3.8) |
| ISS (median [IQR]) |  | 4.0 [1.0, 10.0] | 3.0 [1.0, 9.0] |
| NISS (median [IQR]) |  | 5.0 [1.0, 16.0] | 3.0 [1.0, 11.0] |
| ISS>15 (%) | No | 10666.8 (81.8) | 12206.9 (87.7) |
|  | Yes | 2375.2 (18.2) | 1716.1 (12.3) |

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Level | Multiple centres | Single centre |
| n (%) |  | 26965 (81.9) | 5956 (18.1) |
| Age, years (median [IQR]) |  | 41.0 [25.0, 59.0] | 42.0 [27.0, 59.0] |
| Sex (%) | Male | 17813.4 (66.1) | 4153.2 (69.7) |
|  | Female | 9151.6 (33.9) | 1802.8 (30.3) |
| GCS (median [IQR]) |  | 15.0 [15.0, 15.0] | 15.0 [14.0, 15.0] |
| SBP (median [IQR]) |  | 138.0 [124.0, 153.0] | 140.0 [124.0, 157.0] |
| RR (median [IQR]) |  | 18.0 [16.0, 21.0] | 18.0 [15.0, 20.0] |
| 30-day survival (%) | Alive | 25827.7 (95.8) | 5628.2 (94.5) |
|  | Dead | 1137.3 (4.2) | 327.8 (5.5) |
| ISS (median [IQR]) |  | 4.0 [1.0, 9.0] | 5.0 [1.0, 14.0] |
| NISS (median [IQR]) |  | 4.0 [1.0, 12.0] | 9.0 [3.0, 22.0] |
| ISS>15 (%) | No | 22873.7 (84.8) | 4488.8 (75.4) |
|  | Yes | 4091.3 (15.2) | 1467.2 (24.6) |

GCS, Glasgow coma scale; SBP, systolic blood pressure; RR, respiratory rate; ISS, injury severity score; NISS, new injury severity score

## Development and validation

During model development, the number of imputations used for each data set was 43, 48 and 37 for the high- and low-volume data set, the metropolitan and non-metropolitan data set, and the multiple and single centre data set, respectively.

Table 2 shows the model validation performance. The model with the lowest validation mistriage rate was the non-metropolitan model, with a median mistriage rate of 0.18 (0.18 - 0.44). The performance in terms of undertriage and overtriage rate for the same model was 0.05 (0.04 - 0.07) and 0.12 (0.11 - 0.40), respectively. The highest validation mistriage rate (i.e. worst validation performance) was found in the low-volume sample, with a model median mistriage rate of 0.64 (0.47 - 0.73).

**Table 2. Model validation performance, i.e. performance when applied to the validation subsample of the sample in which the model was created.** Data are presented as point estimates with 95% uncertainty intervals (95% UI).

|  |  |  |
| --- | --- | --- |
|  | High-volume | Low-volume |
| Mistriage | 0.23 (0.15 - 0.39) | 0.64 (0.47 - 0.73) |
| Undertriage | 0.05 (0.04 - 0.06) | 0.03 (0.02 - 0.04) |
| Overtriage | 0.17 (0.09 - 0.35) | 0.60 (0.43 - 0.71) |

|  |  |  |
| --- | --- | --- |
|  | Metropolitan | Non-metropolitan |
| Mistriage | 0.35 (0.25 - 0.48) | 0.18 (0.18 - 0.44) |
| Undertriage | 0.05 (0.04 - 0.07) | 0.05 (0.04 - 0.07) |
| Overtriage | 0.30 (0.18 - 0.43) | 0.12 (0.11 - 0.40) |

|  |  |  |
| --- | --- | --- |
|  | Multiple centres | Single centre |
| Mistriage | 0.28 (0.14 - 0.44) | 0.56 (0.40 - 0.62) |
| Undertriage | 0.05 (0.03 - 0.06) | 0.04 (0.03 - 0.07) |
| Overtriage | 0.23 (0.08 - 0.41) | 0.52 (0.33 - 0.59) |

## Comparison

Model performance after transfer was determined for each model after being transferred to the validation subsample in the other sample, in the data set (Table 3). When transferred, the model with the lowest mistriage rate was the non-metropolitan model, with a median mistriage rate of 0.22 (0.24 - 0.42). The model with the highest mistriage rate after transfer was the low-volume model with a median mistriage rate of 0.67 (0.46 - 0.79).

**Table 3. Model transfer performance, i.e. performance when transferred to the other validation subsample in the same data set.** Data are presented as point estimates with 95% uncertainty intervals (95% UI).

|  |  |  |
| --- | --- | --- |
|  | High-volume | Low-volume |
| Mistriage | 0.27 (0.20 - 0.41) | 0.67 (0.46 - 0.79) |
| Undertriage | 0.06 (0.04 - 0.08) | 0.02 (0.01 - 0.03) |
| Overtriage | 0.21 (0.12 - 0.36) | 0.65 (0.43 - 0.77) |

|  |  |  |
| --- | --- | --- |
|  | Metropolitan | Non-metropolitan |
| Mistriage | 0.38 (0.22 - 0.52) | 0.22 (0.24 - 0.42) |
| Undertriage | 0.04 (0.03 - 0.05) | 0.07 (0.06 - 0.11) |
| Overtriage | 0.34 (0.17 - 0.50) | 0.15 (0.14 - 0.35) |

|  |  |  |
| --- | --- | --- |
|  | Multiple centres | Single centre |
| Mistriage | 0.30 (0.19 - 0.40) | 0.62 (0.40 - 0.69) |
| Undertriage | 0.08 (0.05 - 0.09) | 0.02 (0.02 - 0.04) |
| Overtriage | 0.22 (0.11 - 0.35) | 0.60 (0.37 - 0.67) |

Table 4 shows the transferred model performance when compared to the validation performance in the validation subsample receiving the transferred model. The highest mistriage rate when compared to the receiving sample validation model performance was found in the low-volume model, with a median mistriage rate difference of 0.44 (0.17 - 0.60), meaning this model performed worse when transferred to the other sample in the same data set. In clinical terms, this model transfer means that among 100 trauma patients, 44 more patients would be wrongly classified as having major or minor trauma.

**Table 4. Comparison mistriage, i.e. sample validation performance subtracted from transferred performance.** Data are presented as point estimates with 95% uncertainty intervals (95% UI).

|  |  |  |
| --- | --- | --- |
|  | High-volume | Low-volume |
| Mistriage | -0.36 (-0.50 - -0.14) | 0.44 (0.17 - 0.60) |
| Undertriage | 0.03 (0.01 - 0.05) | -0.03 (-0.05 - -0.01) |
| Overtriage | -0.38 (-0.54 - -0.16) | 0.47 (0.18 - 0.64) |

|  |  |  |
| --- | --- | --- |
|  | Metropolitan | Non-metropolitan |
| Mistriage | 0.21 (-0.11 - 0.35) | -0.13 (-0.34 - -0.07) |
| Undertriage | -0.01 (-0.03 - 0.01) | 0.02 (0.01 - 0.06) |
| Overtriage | 0.22 (-0.12 - 0.37) | -0.15 (-0.40 - -0.09) |

|  |  |  |
| --- | --- | --- |
|  | Multiple centres | Single centre |
| Mistriage | -0.24 (-0.24 - 0.08) | 0.32 (0.10 - 0.45) |
| Undertriage | 0.04 (-0.01 - 0.04) | -0.03 (-0.04 - -0.01) |
| Overtriage | -0.28 (-0.27 - 0.09) | 0.35 (0.12 - 0.49) |

# Discussion

This study aimed to assess how transfers of clinical prediction models for early trauma care between different contexts within a single health system affect mistriage rates. Varying degrees of uncertainty were observed when comparing the transferred performance to the receiving sample validation performance.

The most notable effect on model performance following model transfer was observed after transferring the low-volume model. This transfer resulted in an increased mistriage rate of 0.44. Mainly contributing to this was an increase in overtriage. For this transfer, the 95% UI ranged from 0.17 to 0.60, indicating a high level of uncertainty. In contrast, the transfer of the high-volume model to the low-volume sample reduced mistriage, primarily due to reduced overtriage.

Low-volume centres do by definition receive fewer trauma patients, and therefore less data is available to establish accurate clinical prediction models. It is reasonable to assume this is the reason for the marked increase in mistriage when transferring the low-volume model to the high-volume sample. This also explains why the low-volume model and the single-centre model showed the highest levels of mistriage during model validation (i.e. when applied to the sample in which the model was created). Further confirming this, the transfer of the high-volume model to the low-volume sample reduced mistriage.

In a clinical context, this suggests that clinical prediction models for early trauma care should be developed in settings with a high number of trauma patients. These models could then be transferred to settings with less trauma patients, which should then result in a decreased mistriage rate in those settings. In 2018, Granström et al. found that using a criteria-based triage system (i.e. a clinical prediction model) for trauma triage was effective in reducing overtriage without increasing mortality [7]. Therefore, improving clinical prediction models to reduce mistriage is clinically beneficial, potentially saving lives by minimizing undertriage, and likely saving resources by minimizing overtriage. Possibly, with registries as SweTrau, the analysis of model transfer could be performed continuously to optimize the models currently in use and advise which model transfers might be detrimental (or potentially beneficial).

The transfer of clinical prediction models in trauma care has not previously been studied extensively. In 2016, Gerdin et al. found that model transfer adversely affected model performance in terms of calibration but that this could be improved by updating the model [9]. This study did not examine model performance in terms of calibration, nor were the effects of updating the model explored.

External validation - the evaluation of a clinical prediction model in a setting in which it was not originally developed - has been studied more extensively. Studies using simulated clinical prediction models based on different predictors showed a decline in model performance when externally validated [22, 25]. A study from 2018 examining the external validity of prediction models for coronary artery disease (CAD) [26] showed that the models significantly underestimated the probability of CAD. The combined results of this study, the studies on model transfer and the studies on external validation would suggest indiscriminate model transfers could lead to detrimental clinical consequences, but also that clinically beneficial model transfers could be identified.

## Strengths and limitations

### Strengths

The study design realistically reflects potential model transfers: clinical prediction models for trauma are usually developed in large metropolitan centres with a high volume of patients and then transferred as clinical recommendations to minor low-volume centres. The results should therefore be of practical importance.

### Limitations

Originally, all single centres with a valid number of events were to constitute a data set, but only one centre provided a sufficient number of events, leading to the loss of the individual centres data set. This data set could have provided interesting results reflecting realistic situations.

This study used GCS, RR and SBP as predictors for 30-day mortality. Additional predictors such as age and mechanism of injury or newer predictors such as the shock index might increase model performance. Furthermore, we used 30-day mortality as an outcome, recognizing that late mortality, time in hospital, functional impairment, and morbidity are also important outcomes in the context of trauma care.

The predetermined analysis plan was to be followed in a step-by-step fashion. Programming challenges and late-time code inclusions required a flexible approach to this analysis plan, sometimes with revisions of previous steps. This is not optimal, and efforts were made to not deviate in any major way. Using GitHub, all code was made public for inspection and to ensure reproducibility [10].

## Conclusion

Model transfer resulted in varying levels of uncertainty in in terms of its effect on mistriage. In general, samples with a high number of patients had the lowest mistriage both during validation and after transfer to the other sample in the data set. The opposite was also true, and the transfer of the low-volume model led to an increased mistriage of 0.44 (0.17 - 0.60), noting that this observation was associated with considerable uncertainty. Recognizing the limitations of this study, the authors believe that further studies are warranted due to the potential economic costs and patient consequences of poor model transfers.

# References

1. Moore EE, Feliciano DV, Mattox KL. Trauma. 8th edition. Book. New York: McGraw-Hill Education; 2017.

2. James SL, Abate D, Abate KH, Abay SM, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the global burden of disease study 2017. The Lancet. 2018;1789–858.

3. Baker T, Gerdin M. The clinical usefulness of prognostic prediction models in critical illness. Eur J Intern Med. 2017;45:37–40.

4. Naghavi M, Abajobir AA, Abbafati C, Abbas K, et al. 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. The Lancet. 2017;1151–210.

5. World Health Organization. Injuries and violence: The facts. Report. World Health Organization; 2014.

6. American College of Surgeons. Advanced Trauma Life Support (ATLS). 10th edition. Book. Chicago: The Committee on Trauma; 2018.

7. Granstrom A, Strommer L, Schandl A, Ostlund A. A criteria-directed protocol for in-hospital triage of trauma patients. Eur J Emerg Med. 2018;25:25–31.

8. Rehn M, Perel P, Blackhall K, Lossius HM. Prognostic models for the early care of trauma patients: A systematic review. Scand J Trauma Resusc Emerg Med. 2011;19:17.

9. Gerdin M, Roy N, Fellander-Tsai L, Tomson G, Schreeb J von, 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.

10. Henriksson M. Github: Transfer effect mistriage. 2019.

11. Lennquist S. Traumatologi. 1st edition. Book. Stockholm: Liber AB; 2007.

12. Socialstyrelsen. Socialstyrelsen: Traumavard vid allvarlig handelse. 2015.

13. LOF. LOF: Nationella traumalarmskriterier 2017. 2017.

14. SweTrau. SweTrau: Minnesanteckning. 2018.

15. McFadyen JG, Ramaiah R, Bhananker SM. Initial assessment and management of pediatric trauma patients. Int J Crit Illn Inj Sci. 2012;2:121–7.

16. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the trauma score. J Trauma. 1989;29:623–9.

17. Palmer C. Major trauma and the injury severity score–where should we set the bar? Annu Proc Assoc Adv Automot Med. 2007;51:13–29.

18. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49:1373–9.

19. Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger TV. Performance of logistic regression modeling: Beyond the number of events per variable, the role of data structure. Journal of Clinical Epidemiology. 2011;64:993–1000.

20. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol. 2005;58:475–83.

21. The R team. R-project. 2018.

22. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: A clinical example. J Clin Epidemiol. 2003;56:826–32.

23. Rotondo MF, Cribari C, Smith RS. Resources for the optimal care of the injured patient. Report. American College of Surgeons; 2014.

24. Buuren S van, Groothuis-Oudshoorn C. MICE: Multivariate imputation by chained equations in r. Book. 2011.

25. Terrin N, Schmid CH, Griffith JL, D’Agostino RB, Selker HP. External validity of predictive models: A comparison of logistic regression, classification trees, and neural networks. J Clin Epidemiol. 2003;56:721–9.

26. 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;11:437–46.