Transferring clinical prediction models for early trauma care can lead to increased mistriage

M. Henriksson, BScN1; D. D Saulnier, MSc1; J. Berg, MD1,2; M. Gerdin Wärnberg, MD, PhD1

1 Department of Global Public Health, Karolinska Institutet,

Stockholm, Sweden

2 Department of Emergency Medicine, Skåne University Hospital,

Malmö, Sweden

Corresponding author: Martin Henriksson

And reprints Department of Public Health Sciences

Karolinska Institutet

Tomtebodavägen 18

171 65 Solna, Sweden

martin.henriksson@stud.ki.se

[https://orcid.org/0000-0002-0993-1501](https://orcid.org/0000-0001-6069-4794)

Phone +46 73 070 73 25

*Word count*: 2,990 (Without tables, figures and references)

*Tables*: 4

*Figures*: 1

*Trial registration:* The protocol for this study was uploaded to github.com/613Martin/transfer-effect-mistriage before the research was undertaken.

*Declaration of interest:* The authors declare no conflict of interest.

*Source of funding:* This work was supported by the Swedish National Board of Health and Welfare. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

*Acknowledgements:* The analyses were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at LUNARC.

*Ethics committee approval:* This study has been approved by the Regional Ethical Review Board in Stockholm. Ethical review numbers 2015/426-31 and 2016/461-32.

Transferring clinical prediction models for early trauma care might lead to increased mistriage

April 16, 2020

**Abstract**

*Objective:* This study aimed to assess how transfers of clinical prediction models for early trauma care between different care contexts within a single health system affected mistriage rates.

*Study design and setting:* Patients aged 15 years or older, registered between 2011 and 2016 in the Swedish national trauma registry, SweTrau, were included. Three dataset groups 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 dataset group and transferred within groups. Model performance was evaluated in all dataset groups using 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 dataset groups.

*Results:* A total of 26 965 patients were included. Changes in mistriage rates after transfer ranged from -0.25 (95% UI -0.21 – 0.04) to 0.29 (95% UI 0.13 - 0.39). Both overtriage and undertriage rates were affected.

*Conclusions:* Transferring clinical prediction models can lead to uncertain outcomes. Depending on the care context, model transfer led to either increased or decreased mistriage. Overtriage was more affected by model transfer than undertriage.

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

**What’s new?**

* Transferring clinical prediction models should warrant caution, as the outcomes might vary depending on the care context.
* Overtriage is more affected by model transfer than undertriage. Some transfers leading to improved overtriage also lead to worse performance in terms of undertriage.
* Clinical prediction models for early trauma care could be developed in settings with a high volume of trauma patients, as these models generally perform better after transfer, but only in terms of mistriage.

# Introduction

Trauma, defined as physical injuries to a host by outside objects [1], accounts for approximately 458.5 million hospital visits 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 trauma due to 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. 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, and are then implemented in different contexts [9].

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, that study assessed model transfers between substantially different settings (India and US) 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

The study is a registry-based cohort study that used…

## Source of data

A registry-based cohort study was conducted using Sweden’s national trauma registry, SweTrau. At the time of the study, the register consisted of 55 000 trauma cases, recorded from 52 of Sweden’s 55 hospitals [11]. SweTrau data was 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].

## 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 [12]. The Swedish guidelines for trauma activation define children as individuals aged <15 years [13]. Patients with missing date and time of trauma were excluded.

**Outcome**

### Model outcome

The outcome used to develop the clinical prediction models was all-cause mortality within 30 days of trauma. This outcome was selected as many established prediction models were developed with mortality as the outcome [14-17].

### Study outcome

Injury Severity Score (ISS) is a common scoring system used to assess trauma severity. ISS >15 was used as the gold standard to define trauma severity as major trauma; patients with ISS <15 were considered to have minor trauma [18]. ISS >15 was used as the cutoff as this is how major trauma is conventionally defined [19]. Overtriage was defined as the event when a clinical prediction model classified a patient with ISS <15 as having major trauma. 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 major trauma patients. The mistriage rate was defined as the sum of the over- and undertriage rates.

## Predictors

### 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 [20].

## *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.

## *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.

## Sample size

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

### **Missing data**

We used multiple imputation using chained equations (mice) to handle missing data [21]. Each sample was imputed separately and the number of imputations created for each sample was equal to the highest percentage of missing data in that sample. 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 sample. 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 [22, 23].

**Statistical analysis methods**

### Dataset groups

The complete SweTrau cohort was split into three overlapping groups of data, each drawn from the complete patient cohort, and each representing a plausible care context. These dataset groups were further divided into datasets as follows:

### Dataset group 1: High- and low-volume centres

Two datasets 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.

### Dataset group 2: Metropolitan and non-metropolitan centres

This dataset group was also split into two datasets: a metropolitan sample consisting of patients from greater Stockholm, greater Gothenburg and greater Malmo, as defined by Swedish statistics, and a second dataset consisting of patients from non-metropolitan areas.

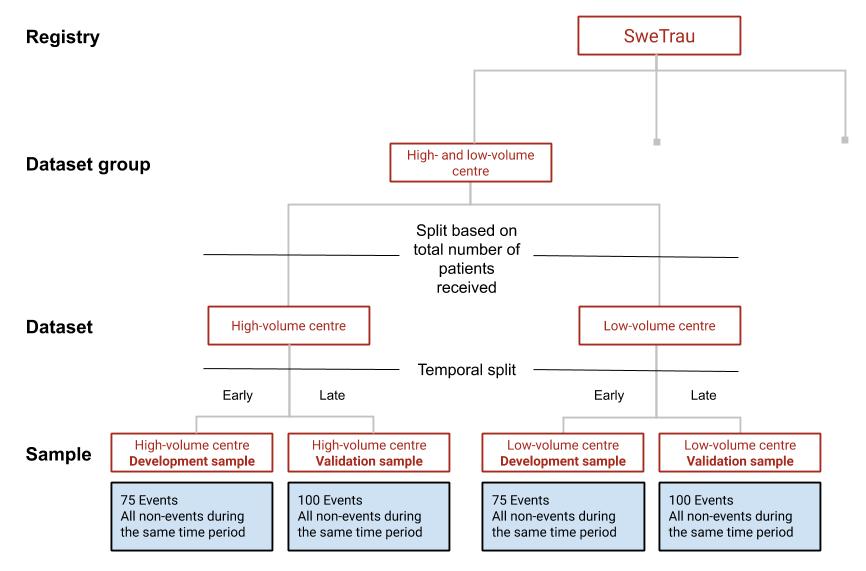
### Dataset group 3: Multiple and single centre data

In this dataset group, two datasets were created. The single centre dataset with large enough sample sizes (as defined below) to develop and validate a model was its own dataset. The multiple centres dataset consisted of the combined data from all centres.

A fourth dataset group was initially created, consisting of individual centres. However, this dataset group was removed due to a lack of centres with a sufficiently large sample size.

### Development and validation sample

The datasets in each dataset group were split into two samples using a temporal split based on the date of trauma. The earlier sample was the development sample, and the later sample 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 [24, 25]. The validation samples all contained 100 events and at least 100 non-events, which was suggested as the minimum number by Vergouwe in 2005 for external validation samples [26]. Figure 1 shows the development of the samples using the high- and low-volume centre dataset group as an example. The minimum sample sizes of the development and validation samples were 150 and 200, respectively. We performed analyses only on datasets for which all samples included at least the minimum sample size.

  
**Fig. 1: Sample development using the high- and low-volume centre dataset group as an example.** 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 [27], 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 dataset group. Below, we use the transfer of a model from the high-volume centre dataset to the low-volume centre dataset 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 based on the calibration slope, it was then applied to the model coefficients [28]. 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% [19]. 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 using the Boot MI Percentile method as presented by Barlett and Hughes was used to estimate the 95% confidence intervals (CI) around the performance and the differences in the performance estimates [29]. 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 dataset groups.

### Performance measures

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

## 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.

# Results

We analysed data from 26 965 trauma patients (Table 1), after excluding 78 patients with missing date and time of trauma. The total number of missing observations across all variables was 9 984 in the entire study cohort. The dataset with the highest percentage of missing observations was the non-metropolitan dataset, with 48% incomplete observations. The variable with the highest number of missing values was RR, with 8 296 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]. Count data are presented as a count followed by (%) in the table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Level** | **High-volume** | **Low-volume** |
| n (%) |  | 20021 (74) | 6944 (26) |
| Age, years |  | 40 [25, 57] | 46 [27, 64] |
| Sex (%) | Male | 13134 (66) | 4680 (67) |
|  | Female | 6886 (34) | 2263 (33) |
| GCS |  | 15 [15, 15] | 15 [14, 15] |
| SBP |  | 139 [125, 153] | 136 [120, 152] |
| RR |  | 18 [15, 20] | 19 [16, 22] |
| 30-day survival (%) | Alive | 19271 (96) | 6546 (94) |
|  | Dead | 750 (4) | 397 (6) |
| ISS |  | 3 [1, 9] | 5 [1, 13] |
| NISS |  | 3 [1, 12] | 6 [1, 17] |
| ISS>15 (%) | No | 17344 (87) | 5530 (80) |
|  | Yes | 2677 (13) | 1414 (20) |
| **Characteristic** | **Level** | **Metropolitan** | **Non-metropolitan** |
| n (%) |  | 13042 (48) | 13923 (52) |
| Age, years |  | 41 [26, 58] | 42 [25, 60] |
| Sex (%) | Male | 8804 (67) | 9010 (65) |
|  | Female | 4238 (33) | 4913 (35) |
| GCS |  | 15 [15, 15] | 15 [15, 15] |
| SBP |  | 140 [125, 156] | 135 [121, 150] |
| RR |  | 18 [15, 20] | 18 [16, 21] |
| 30-day survival (%) | Alive | 12428 (95) | 13397 (96) |
|  | Dead | 614 (5) | 526 (4) |
| ISS |  | 4 [1, 10] | 3 [1, 9] |
| NISS |  | 5 [1, 16] | 3 [1, 11] |
| ISS>15 (%) | No | 10667 (82) | 12207 (88) |
|  | Yes | 2375 (18) | 1716 (12) |
| **Characteristic** | **Level** | **Multiple centres** | **Single centre** |
| n (%) |  | 26965 (82) | 5956 (18) |
| Age, years |  | 41 [25, 59] | 42 [27, 59] |
| Sex (%) | Male | 17813 (66) | 4153 (70) |
|  | Female | 9152 (34) | 1803 (30) |
| GCS |  | 15 [15, 15] | 15 [14, 15] |
| SBP |  | 138 [124, 153] | 140 [124, 157] |
| RR |  | 18 [16, 21] | 18 [15, 20] |
| 30-day survival (%) | Alive | 25828 (96) | 5628 (94) |
|  | Dead | 1137 (4) | 328 (6) |
| ISS |  | 4 [1, 9] | 5 [1, 14] |
| NISS |  | 4 [1, 12] | 9 [3, 22] |
| ISS>15 (%) | No | 22874 (85) | 4489 (75) |
|  | Yes | 4091 (15) | 1467 (25) |

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 sample was 47, 57 and 38 for the high- and low-volume dataset group, the metropolitan and non-metropolitan dataset group, and the multiple and single centre dataset group, respectively.

Table 2 shows the model validation performance. The model with the lowest mistriage rate was the non-metropolitan model, with an undertriage and overtriage rate of 0.05 (0.05 - 0.08) and 0.20 (0.13 - 0.38), respectively.

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

|  |  |  |
| --- | --- | --- |
|  | **High-volume** | **Low-volume** |
| Mistriage | 0.27 (0.16 - 0.38) | 0.51 (0.42 - 0.65) |
| Undertriage | 0.06 (0.05 - 0.08) | 0.04 (0.02 - 0.06) |
| Overtriage | 0.21 (0.08 - 0.32) | 0.47 (0.36 - 0.62) |
|  | **Metropolitan** | **Non-metropolitan** |
| Mistriage | 0.34 (0.26 - 0.45) | 0.24 (0.21 - 0.44) |
| Undertriage | 0.07 (0.05 - 0.09) | 0.05 (0.05 - 0.08) |
| Overtriage | 0.27 (0.17 - 0.39) | 0.20 (0.13 - 0.38) |
|  | **Multiple centres** | **Single centre** |
| Mistriage | 0.31 (0.15 - 0.41) | 0.56 (0.44 - 0.63) |
| Undertriage | 0.07 (0.03 - 0.07) | 0.04 (0.03 - 0.07) |
| Overtriage | 0.24 (0.08 - 0.37) | 0.52 (0.37 - 0.60) |

## Comparison

Model performance after transfer was determined for each model after being transferred to the validation sample in the other dataset, in the dataset group (Table 3).

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

|  |  |  |
| --- | --- | --- |
|  | **High-volume model in low-volume validation sample** | **Low-volume model in high-volume validation sample** |
| Mistriage | 0.31 (0.22 - 0.40) | 0.52 (0.39 - 0.72) |
| Undertriage | 0.07 (0.05 - 0.10) | 0.03 (0.02 - 0.05) |
| Overtriage | 0.24 (0.13 - 0.34) | 0.48 (0.34 - 0.70) |
|  | **Metropolitan model in non-metropolitan validation sample** | **Non-metropolitan model in metropolitan validation sample** |
| Mistriage | 0.37 (0.25 - 0.51) | 0.26 (0.24 - 0.40) |
| Undertriage | 0.03 (0.03 - 0.05) | 0.09 (0.06 - 0.11) |
| Overtriage | 0.33 (0.20 - 0.47) | 0.17 (0.14 - 0.34) |
|  | **Multiple centres model in single centre validation sample** | **Single centre model in multiple centre validation sample** |
| Mistriage | 0.31 (0.19 - 0.37) | 0.60 (0.46 - 0.67) |
| Undertriage | 0.08 (0.07 - 0.12) | 0.03 (0.02 - 0.04) |
| Overtriage | 0.23 (0.09 - 0.29) | 0.57 (0.43 - 0.65) |

Table 4 is derived from table 2 and 3, it shows the difference between the performance of the transferred model and the validation model in the validation sample to which the model was transferred. For example: The low-volume model transferred to the high-volume validation sample results in a mistriage rate of 0.52 (0.39 - 0.72). The high-volume model validation mistriage rate was 0.27 (0.16 - 0.38). The difference then being 0.52 – 0.27 = 0.25, meaning there was an increased mistriage rate of 0.25 (0.08 - 0.52) from this particular model transfer. In clinical terms, this means that among 100 trauma patients, 25 more patients would be wrongly classified as having major or minor trauma following this model transfer.

**Table 4. Model performance difference, i.e. transferred model performance minus model validation sample performance in the validation sample receiving the transferred model.** Data are presented as point estimates with 95% uncertainty intervals (95% UI).

|  |  |  |
| --- | --- | --- |
|  | **High-volume model in low-volume validation sample** | **Low-volume model in high-volume validation sample** |
| Mistriage | -0.20 (-0.40 - -0.08) | 0.25 (0.08 - 0.52) |
| Undertriage | 0.03 (0.01 - 0.06) | -0.03 (-0.05 - -0.01) |
| Overtriage | -0.23 (-0.45 - -0.09) | 0.28 (0.10 - 0.57) |
|  | **Metropolitan model in non-metropolitan validation sample** | **Non-metropolitan model in metropolitan validation sample** |
| Mistriage | 0.12 (-0.06 - 0.30) | -0.08 (-0.34 - -0.11) |
| Undertriage | -0.01 (-0.03 - -0.00) | 0.02 (0.02 - 0.06) |
| Overtriage | 0.13 (-0.06 - 0.33) | -0.10 (-0.39 - -0.13) |
|  | **Multiple centres model in single centre validation sample** | **Single centre model in multiple centre validation sample** |
| Mistriage | -0.25 (-0.21 - 0.04) | 0.29 (0.13 - 0.39) |
| Undertriage | 0.04 (-0.00 - 0.05) | -0.04 (-0.05 - -0.02) |
| Overtriage | -0.29 (-0.25 - 0.04) | 0.33 (0.15 - 0.44) |

# 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. The most notable effect on model performance following model transfer was observed after transferring the single centre model. This transfer resulted in an increased mistriage rate of 0.29. Mainly contributing to this was an increase in overtriage. In contrast, the transfer of the multiple centre model to the single centre validation sample reduced mistriage, primarily due to reduced overtriage. Interestingly, these transfers also seemed to have the opposite effect on undertriage, leading to decreased and increased undertriage respectively.

In all samples, model transfer led to either increased or decreased mistriage. The model transfers which led to reduced mistriage also led to increased undertriage. In other words, all model transfers led to adverse outcomes, either by a large increase in overtriage, or by a smaller increase in undertriage, which could be considered more clinically severe. In general, transferring from centres with a high volume of patients seemed to reduce mistriage.

In a clinical context, this could suggest that clinical prediction models for early trauma care could be developed in settings with a high number of trauma patients. These models could then be transferred to settings with less trauma patients, which could result in a decreased mistriage rate. However, the results of this study also show this might lead to increased undertriage, which in a clinical context is unacceptable due to the potentially catastrophic patient consequences.

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.

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]. The aim of this study was to assess the effects of model transfer on mistriage, and therefore model calibration was not evaluated, neither were the effects of model updating.

External validation has been studied more extensively. Studies using simulated clinical prediction models based on different predictors showed a decline in model performance when externally validated [28, 30]. A study from 2019 examining the risk prediction models in the management of acute kidney injury [31] concluded that external validation often led to reduced model accuracy.

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 applied in contexts different to the ones in which they were originally developed. The results should therefore be of practical importance.

This study evaluated the performance of the clinical prediction models using over- and undertriage. This was done in order to more accurately map model performance after transfer using clinically relevant measurements. Traditional performance measures such as discrimination and calibration were omitted, as these are already well studied.

### Limitations

Originally, all single centres with a valid number of events were to constitute a dataset group, but only one centre provided a sufficient number of events, leading to the loss of the individual centres dataset group.

Another limitation was the method chosen to identify the cutoff: minimizing overtriage while keeping undertriage at <5% is dependent on the frequency of major trauma in the sample. This will lead to unstable cutoffs, as the incidence of major trauma differs between the samples.

## Conclusion

Depending on the care context, model transfer lead to either increased or decreased mistriage. Both over- and undertriage were affected by model transfer, and the effects on mistriage were in all cases primarily due to changes in overtriage. Datasets with a high number of patients had the lowest mistriage both during validation, and after transfer to the validation sample in the other dataset. However, the transfer of these models (while improving mistriage) also lead to increased undertriage. The opposite was also true, and the transfer of the low-volume model led to a significantly increased mistriage, 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. SweTrau. SweTrau: Minnesanteckning. 2018.

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

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

14. Champion HR, Sacco WJ, Copes W, Gann DS, et al. A revision of the trauma score. J Trauma. 1989;623–9.

15. Sartorius D, Le-Manach Y, David JS, Rancurel E, et al. Mechanism, glasgow coma scale, age, and arterial pressure (mgap): A new simple prehospital triage score to predict mortality in trauma patients. Crit Care Med. 2010;831–7.

16. Kondo Y, Abe T, Kohshi K, Tokuda Y, et al. Revised trauma scoring system to predict in-hospital mortality in the emergency department: Glasgow coma scale, age, and systolic blood pressure score. Crit Care. 2011;191.

17. Kobusingye OC, Lett RR. Hospital-based trauma registries in uganda. J Trauma. 2000;498–502.

18. 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.

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

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

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

22. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. BMC Med Res Methodol. 2009;57.

23. Brand J, Buuren S van, Cessie S le, Hout W van den. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. Stat Med. 2019;210–20.

24. 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.

25. 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.

26. 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.

27. The R team. R-project. 2018.

28. 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.

29. Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under uncongeniality and misspecification. 2019.

30. 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.

31. Hodgson L, Selby N, Huang T, Forni L. The role of risk prediction models in prevention and management of aki. Seminars in Nephrology. 2019;39:421–30.