ELSEVIER

Journal of Clinical Epidemiology ■ (2016) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Traumatic transfers: calibration is adversely affected when prediction models are transferred between trauma care contexts in India and the United States

Martin Gerdin[a,*], Nobhojit Roy[a,b,c], Li Felländer-Tsai[d], Göran Tomson[a,e], Johan von Schreeb[a], Max Petzold[f,g], On behalf of the Towards Improved Trauma Care Outcomes in India (TITCO) Consortium

[a]*Health Systems and Policy, Department of Public Health Sciences, Karolinska Institutet, SE-171 77 Stockholm, Sweden*
[b]*Department of Surgery, Bhabha Atomic Research Centre Hospital, Mumbai, Maharashtra 400085, India*
[c]*School of Habitat, Tata Institute of Social Sciences, Chembur, Mumbai, Maharashtra 400088, India*
[d]*Division of Orthopedics and Biotechnology, Department of Clinical Science Intervention and Technology, Karolinska Institutet, Alfred Nobels allé 8, SE-141 52 Huddinge, Sweden*
[e]*Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, SE-171 77 Stockholm, Sweden*
[f]*Centre for Applied Biostatistics, Occupational and Environmental Medicine, Sahlgrenska Academy, University of Gothenburg, PO Box 414, SE-405 30 Gothenburg, Sweden*
[g]*School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, 7 York Rd, Johannesburg 2193, South Africa*

Accepted 4 January 2016; Published online xxxx

Abstract

Objective: We evaluated the transferability of prediction models between trauma care contexts in India and the United States and explored updating methods to adjust such models for new contexts.

Study Design and Settings: Using a combination of prospective cohort and registry data from 3,728 patients of Towards Improved Trauma Care Outcomes in India (TITCO) and from 18,756 patients of the US National Trauma Data Bank (NTDB), we derived models in one context and validated them in the other, assessing them for discrimination and calibration using systolic blood pressure, heart rate, and Glasgow coma scale as candidate predictors.

Results: Early mortality was 8% in the TITCO and 1−2% in the NTDB samples. Both models discriminated well, but the TITCO model overestimated the risk of mortality in NTDB patients, and the NTDB model underestimated the risk in TITCO patients.

Conclusion: Transferability was good in terms of discrimination but poor in terms of calibration. It was possible to improve this miscalibration by updating the models' intercept. This updating method could be used in samples with as few as 25 events.

Keywords: Prediction modeling; Trauma; Transferability; Discrimination; Calibration; Global health

1. Introduction

Prediction models supplement clinical decision making in most areas of medicine [1−7]. A prediction model can be defined as an algorithm, often derived using statistical methods and based on two or more parameters, that may be used to estimate the risk of a specific outcome in an individual [8]. Methodological and reporting guidelines emphasize that prediction models must show external validity before they can be used in clinical practice [8−11]. To assess external validity, models are evaluated in contexts or populations different from the contexts or populations in which they were derived [12]. Such evaluations may reveal substantial problems with external validity and question the usefulness of published prediction models [13].

Conflict of interest: None.

Funding: This study was funded by grants from the Swedish National Board of Health and Welfare and the Laerdal Foundation for Acute Care Medicine, Norway. The funding agencies had no influence on the conduct of the study, the contents of the article, or the decision to send the article for publication.

* Corresponding author. Tomtebodavägen 18A, 171 65 Solna, Sweden. Tel.: +46 708 53 95 98.

E-mail address: martin.gerdin@ki.se (M. Gerdin).

http://dx.doi.org/10.1016/j.jclinepi.2016.01.004
0895-4356/

What is new?

Key findings

- We demonstrate that prediction models for trauma care transfer well between different trauma contexts in terms of discrimination but not calibration.

What this adds to what was known?

- We show that as few as 25 events may be an adequate effective sample size to perform simple updating of logistic prediction models.

What is the implication and what should change now?

- Clinicians and policy makers should ensure that prediction models have been validated in similar contexts to that in which they will be used.
- Furthermore, we propose that researchers explicitly incorporate an updating component in the design and conduct of validation studies.

Several prediction models have been developed to support clinical judgment in trauma care [14–16]. With few exceptions [17,18], such trauma prediction models have been developed in high-income countries [6], although over 90% of the five million trauma deaths that occur each year happen in low- and middle-income countries [19–21]. This discrepancy highlights a potential problem with regard to the external validity and transferability of trauma prediction models. There are several reasons why a prediction model developed in one trauma care context may not be transferable to another, for example, differences in prevention, care, and patient case mix [22,23].

However, almost no research has assessed the transferability of prediction models between trauma contexts in high-income versus low- and middle-income countries. Furthermore, only limited attention has been paid to if and how trauma prediction models that transfer poorly can be updated to better fit a new context. Such updating methods have been suggested and applied in other medical fields [24,25]. Finally, although updating already existing predicting models to new contexts offer the theoretical advantage of requiring smaller samples sizes compared to developing models from scratch [26], there is a dearth of applied research that formally tests this assumption.

To bridge these knowledge gaps, we assessed the transferability of trauma prediction models between two substantially different trauma care contexts. We also assessed how updating methods can be used to recalibrate such models and studied the sample size requirements of these updating methods.

2. Methods

2.1. Study design and contexts

We used data from the Towards Improved Trauma Care in India (TITCO) project, a prospective cohort project rolled out in public university hospitals across urban India and from the US National Trauma Data Bank (NTDB), a data set with data from trauma centers in the United States maintained by the American College of Surgeons [27].

India is a lower middle-income country and accounts for about 20% of global trauma mortality with more than one million annual trauma deaths. Systems for prehospital care are rare, as is the capacity for intensive and critical care. Hence, trauma patients generally arrive unannounced and without prior triage. Trauma care is often the responsibility of junior clinicians with little experience who work long shifts round the clock. For this study, we used TITCO data collected between October 2013 and August 2014 from three centers located in Delhi, Kolkata, and Mumbai. The data collection process has been described elsewhere [17].

The United States is a high-income country where trauma patients are generally dealt with within a well-developed trauma system with integrated prehospital care. Patients are triaged in the prehospital setting and transferred to designated trauma centers. These trauma centers are classified as level I, II, or III depending on level of trauma care provided. NTDB includes data from trauma centers all levels. The full NTDB contains data on more than five million incidents from over 900 trauma centers. For our study, data from 2012 were used, which at the time of writing was the most recent data available.

2.2. Variables

Our outcome was early mortality, defined as death in hospital within 24 hours of the time when the first set of vital signs was recorded in the centers participating in this study. If the time when the first set of vital signs was recorded was not available, we used the time when the patient arrived to hospital instead. We considered systolic blood pressure (mm Hg), heart rate (beats/min) and Glasgow coma scale as potential predictors.

2.3. Analyses, statistical methods, and sample size considerations

To simulate the model development process, we conducted our analyses in three steps (Fig. 1). In step 1, we derived one model using TITCO data and one model using NTDB data. In step 2, we updated the TITCO model in NTDB subsamples and updated the NTDB model in TITCO subsamples (step 2). In step 3, the original unadjusted TITCO model from step 1 as well as the updated TITCO models from step 2 were then validated in an independent NTDB sample. We used the same procedure for NTDB models in a TITCO validation sample (step 3). Model performance was then compared across models.