Injury, Int. J. Care Injured 50 (2019) 2–3

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Editorial

Injury scoring: Then, now, and into the 21st century

Accurately predicting the outcomes of injury is an important aspect of the clinical and administrative management of injured patients. But are we close to defining the perfect predictor? Is a perfect predictor even possible?

The urge to prognosticate the outcomes from injury at least as old as the history of medicine. Four thousand years ago the Egyptian physician Imhotep categorized injuries as survivable ("I can heal"), possibly survivable ("I will fight with"), and unsurvivable ("cannot be healed") [1]. In the 20th century continuous measures of the probability of death were developed, but the approach remained unchanged: using past experience to predict outcomes for future patients. In this essay we review the recent history of injury outcome prediction and suggest how the future may play out. We will find that the measure we are currently using (ISS) has outlived its usefulness because that newer, better measures are available. Indeed, it is likely that modern outcome models are about as good as they can be. Further improvements will have to come from improved approaches to injury description and capture.

Modern mortality models based on anatomically defined injuries have three components. First, a lexicon that partitions the continuous landscape of injury into individual "injuries", care being taken to ensure that all events aggregated into a given "injury" present similar risk of death. Second, individual injuries must each be assigned a severity. Lastly, a mechanism must be conceived that combines individual injury severities into a single prediction for each patient. Baker and colleagues solved these three challenges at a stroke when they conceived the Injury Severity Score (ISS) [2] over 40 years ago. Using the Association for the Advancement of Automotive Medicine's (AAAM) lexicon to define individual injuries along with their expert assigned severities [3] Dr. Baker simply asserted that the ISS was the sum of the squares of the severities of the worst injuries from each of the three most severely injured body regions. The resulting score ranged from 0 (uninjured) to 75 (maximal injury), although only 44 scores were valid due to the vagaries of the "sum of squares" model. Over time other anomalies emerged: the ISS was non-linear, the ISS was non-monotonic, and because the ISS allowed only three injuries to participate in a score it sometimes incorporated less severe injuries in place of more severe injuries [4]. Nevertheless, the ISS performed adequately, especially when other types of information such as physiologic measures (Systolic Blood Pressure, Glasgow Coma Scale, etc.) and measures of patient resilience (age, comorbidities) were added to the injuries only model. Indeed, the ISS was recently incorporated as a predictor in the Trauma Quality Improvement Program model [5] currently used to evaluate trauma centers in the USA.

The ISS was conceived in the last century however, when datasets were small, statistical software was primitive, and expectations for accuracy were modest. In the 21st century the advent of the National Trauma Databank with millions of patients has made it possible to compute individual AIS code injury severities from data using simple regression techniques rather than rely upon experts to estimate these values by considering every possible injury as either present (1) or absent (0) and then fitting a model with all 1300 possible injuries as binary predictors. Once armed with these better estimates of individual injury's severity, regression-based prediction models can be created that allow more nuanced approaches to combining individual injuries into accurate survival predictions. One such AIS based model (Trauma Mortality Prediction Model (TMPM-AIS) [6]) substantially outperformed the venerable ISS, leading Dr. Baker herself to join a call for TMPM-AIS to replace ISS [7]. A similar model based on ICD-9-CM codes [8] also significantly outperforms ISS; this latter result is surprising because while AIS coding specifically captures injury severity, ICD-9-CM coding is much less scrupulous in this regard. These observations confirm that ISS is failing to optimally incorporate the information available in AIS coded data.

The TMPM-AIS model relies on a simple mathematical technique described almost 100 years ago (probit regression, a method very similar to logistic regression), so it is perhaps surprising that a recent attempt [9] to improve on TMPM-AIS' prediction using up-to-the-minute machine learning techniques failed. Although such newly available "big data" techniques (neural networks, random forests, deep learning, etc.) have had considerable success in many arenas, because many important injuries occur rarely or patients seldom survive long enough to be included in injury datasets, predicting outcomes following injury may be a problem uniquely resistant to newer statistical techniques that rely upon massive data. Fortunately, classical statistical techniques are more than adequate for the task of injury outcome prediction. While the examples given so far refer to the AIS and ICD lexicons, these techniques can be used with any lexicon that provides sufficiently granular injury description. Indeed, these techniques are completely general, and can be applied to any binary outcome when the available predictors are also binary.

Although the mathematics underlying mortality models hasn't changed in the last hundred years, the lexicons used to record injuries have changed several times. The AIS lexicon was introduced in 1971 and has been updated numerous times, and

https://doi.org/10.1016/j.injury.2018.11.002

the International Classification of Disease injury descriptors have changed twice: in 1979 ICD-8-CM was replaced by ICD-9-CM, and in 2015 ICD-9-CM was replaced by ICD-10-CM. Because mortality models are tightly coupled to their lexicon, each change in lexicon has required a new mortality model, but because the principles of such models are well understood, creating new models is not difficult. The recent advent of ICD-10-CM left the trauma community without a model to predict outcomes based on this lexicon, but with the arrival of the NTDB's 2016 release with ICD-10-CM coded injuries a prediction model based on ICD-10-CM was quickly created [10], a model which seems adequate for anatomical injury risk adjustment and far better than ISS. In passing we note that outcome models based on ICD-10 coding which lacks augmentation using clinical modifiers [11] is unlikely to be successful, because without clinical modifiers ICD-10 often conflates injuries that have very different clinical implications into a single "injury" (e.g. ICD-10 provides only a single code for "splenic injury" (S360), while in reality splenic injuries range from minor contusion major laceration, distinctions that ICD-10-CM provides for with 5 different grades of injury.)

As mentioned above, adding other types of predictive information to an anatomic injury mortality model can substantially improve performance. It is almost always the case that adding more information, and especially more *types* of information, improve model performance. As an example, traumatic shock is a powerful predictor of mortality precisely because the best predictor of whether a patient will live or die is *whether that patient is living or dying*, exactly what traumatic shock ("the rude unhinging of the machinery of life" [12,13]) represents. However, because one never knows what predictors beyond anatomic injuries will be reliably available in a dataset it may be unwise to prespecify all the predictors in a model. Once the contribution of anatomic injuries to mortality is computed it can be added to a logistic or probit model that also includes other predictors available in the dataset of interest. It's the anatomic injury contribution that's challenging to compute; any other predictor can easily be added to a logistic or probit model on the fly.

Given the importance of factors other than anatomic injuries in determining outcome is it really so important that the anatomic portion of an outcome model be optimized? We feel the answer is yes, for at least two reasons. First, if only anatomic injury data is available for risk adjustment (e.g. administrative data) it is clear that using the best possible model is crucial. And secondly, as payment reform and public reporting of trauma care are introduced some hospitals may be tempted to upcode clinical measures of physiologic derangement. Anatomic injuries are less susceptible to such manipulation.

We also believe that all outcome models, and especially those used to evaluate performance of providers and institutions, should be fully transparent, with the details of their development and characteristics of their performance publicly available. It is only through the process of peer review better models can be developed and blind alleys avoided. The model used by the American College of Surgeons Trauma Quality Improvement Program (TQIP) [5] fails this test.

It is unlikely that mortality prediction will improve dramatically in the future, but improved lexicons that allow more accurate injury description as well as improved data capture will certainly lead to incremental improvements in prediction. It is also unlikely that newer mathematical approaches will have much of an impact, because available techniques are well adapted to the injury prediction problem. Newer graphical methods such as directed acyclic graphing and newer statistical techniques such as causal analysis [14] may help us to better understand why patients die, but these techniques are unlikely to substantially improve overall mortality prediction. We should continue to work toward better mortality models, but we must acknowledge that at some point we will be discovering how inaccurate such models must be rather than how accurate such models can be.

All prediction is an exercise in probability. Even with a perfect prediction model half of patients with a 50% predicted mortality will go on to die; unfortunately, we can never know which half. Our charge is to reduce the size of the 50/50 group in so far as possible, but this group of patients who are afflicted with conditions that physicians " . . . must fight with" will always be with us.

References

[1] van Middendorp J.J., Sanchez GM, Burridge AL. The Edwin Smith papyrus: a clinical reappraisal of the oldest known document on spinal injuries. Eur Spine J 2010;19(Novemer (11)):1815–23.
[2] Baker SP, o'Neill B, Haddon [34_TD$DIFF]Jr W, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma Acute Care Surg 1974;14(March (3)):187–96.
[3] CoMAoA S. Rating the severity of tissue damage. I. The abbreviated scale. Jama 1971;215:277–80.
[4] Kilgo PD, Meredith JW, Hensberry R, Osler TM. A note on the disjointed nature of the injury severity score. J Trauma Acute Care Surg 2004;57(September (3)):479–87.
[5] Hemmila MR, Nathens AB, Shafi S, Calland JF, Clark DE, Cryer HG, et al. The Trauma Quality Improvement Program: pilot study and initial demonstration of feasibility. J Trauma Acute Care Surg 2010;68(2):253–62.
[6] Osler T, Glance L, Buzas JS, Mukamel D, Wagner J, Dick A. A trauma mortality prediction model based on the anatomic injury scale. Ann Surg 2008;247(June (6)):1041–8.
[7] Cook A, Weddle J, Baker S, Hosmer D, Glance L, Friedman L, et al. A comparison of the injury severity score and the trauma mortality prediction model. J Trauma Acute Care Surg 2014;76(January (1)):47–53.
[8] Glance LG, Osler TM, Mukamel DB, Meredith W, Wagner J, Dick AW. TMPM–ICD9: a trauma mortality prediction model based on ICD-9-CM codes. Ann Surg 2009;249(June (6)):1032–9.
[9] Gorczyca MT, Glance LG, Cook AD, Hosmer DW, Cheng JD, Toscano NC, et al. Outcome prediction following injury: a comparison of machine learning and conventional statistical models. PLoS One 2019 In Press.
[10] Osler T, Glance LG, Cook A, Buzas JS, Hosmer DW. A trauma mortality prediction model based on the ICD-10-CM lexicon: TMPM-ICD10. J Trauma Acute Care Surg 2019 Under review.
[11] Gedeborg R, Warner M, Chen LH, Gulliver P, Cryer C, Robitaille Y, et al. Internationally comparable diagnosis-specific survival probabilities for calculation of the ICD-10–based Injury Severity Score. J Trauma Acute Care Surg 2014;76(February (2)):358–65.
[12] Gross SD. System of surgery: pathological, diagnostic, therapeutic, and operative. Blanchard and Lea; 1862. p. 414.
[13] Cairns CB. Rude unhinging of the machinery of life: metabolic approaches to hemorrhagic shock. Curr Opin Crit Care 2001;7(December (6)):437–43.
[14] Pearl J, Glymour M, Jewell NP. Causal inference in statistics: a primer. John Wiley & Sons; 2016 Mar 7.

Turner Osler*
Department of Surgery, University of Vermont, 789 Orchard Shore Road, Colchester, VT 05446, USA

Laurent G. Glance
Department of Anesthesiology, University of Rochester, USA

Jeffrey S. Buzas[a]
David W. Hosmer[a]
[a]*Department of Mathematics and Statistics, University of Vermont, USA*

* Corresponding author.
E-mail address: tosler@uvm.edu (T. Osler).