Injury Risk: Is it Predictable?

Greg Olsen, PhD
Bryce Patterson, MS

OVERVIEW

Without supernatural foresight, we can't predict specific injury events—who will be injured and when. However, for a given population of people and period of time, some people are predictably more likely to be injured than others. Reasons can include:

Some people are more exposed to injury risk through their environment and activities than others.

Some people are in a physiological state that is more prone to an injury than other people.

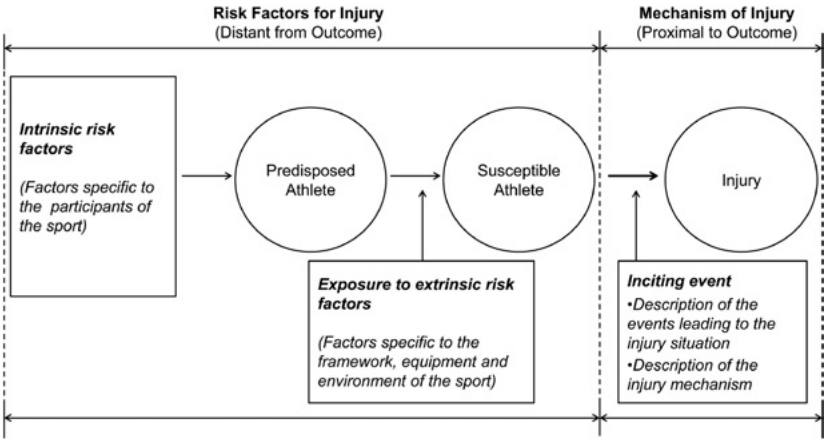

Figure 1: Visualization of model for injury causation

Knowledge of which individuals in a group are most at risk can enable targeted injury rate reduction efforts. Because of the high costs of injuries, even small reductions in injury rates are extremely valuable and that value grows with the size of an affected organization.

Despite this potential value, efforts to assess injury risk have been very limited. The nature of injury events and varying circumstances across populations makes it very difficult to address at a public health level. More focused efforts by organizations like sports teams, medical providers, military units, and employers have not been efficient or scalable. Without objective and reliable metrics, these organizations need to rely on subjective assessments and the scarce capacity of trained professionals. Rapidly expanding capabilities and availability of data collection and machine learning technologies, however, are making injury risk assessment at an organizational level a reality.

A DATA-DRIVEN APPROACH

The foundation for a data-driven approach to relative injury risk assessment is straight forward:

DATA

- Regular capture of “feature” data about each person that can be reliably correlated with injury risk
- Consistent tracking and characterization of injury events over time

RISK PREDICTION MODELS

- Training of ML models that quantify relative injury risk from feature data
- Regular update of models as feature and outcome data accumulate

DATA TO SUPPORT RELATIVE INJURY RISK ASSESSMENT

The availability of data that can support movement health related analyses has substantially improved in recent years. These sources include:

Figure 2

Wearable sensors and associated data collection systems (Figure 2):

Watches, rings, and other body attached sensors have become practical for widespread use.

Figure 3

Video-based motion capture and analysis systems (Figure 3):

Kinematic information capture from consumer phones or other low-cost technology has become a reality.

Figure 4

Force plate based testing and analysis systems (Figure 4):

Force plate systems can fill the role of weight scales and capture rich time series data related to movement health in short duration, low cost tests.

Each of these sources of data have been extensively studied in biomechanical research laboratories for years, and have shown the ability to measure physiological characteristics that are related to injury risk.

To make injury risk assessment operational, this type of data and injury occurrence data needs to be systematically collected at scale. Improved availability and reduced costs of the required technology make it feasible for individual organizations to deploy this type of capability themselves.

RISK PREDICTION MODELS

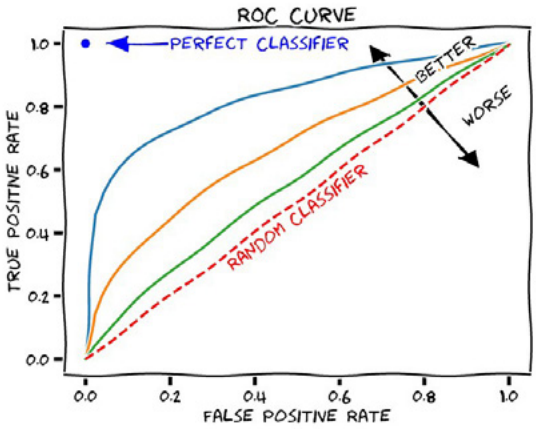

Figure 5: A generalized visualization of a receiver operating characteristic curve

Reliably predicting that an individual is even just 2X more likely to be injured in any given month than a base group could provide the guidance an organization needs to make meaningful reductions in injury rates. This type of relative risk prediction model is well suited to supervised machine learning techniques that naively look for patterns in feature data to classify outcomes. Even what is typically considered low accuracy models (e.g. an only slightly better than random models where AUC > 0.50) can support 2X relative risk ratios.

The accuracy of prediction depends on the amount of available training data, and how well the training data matches the population of a given organization. One would not necessarily expect, for example, a model trained on data collected from field hockey players to perform well for a special forces unit's intensive training regimen. The solution then is for organizations to have models continuously trained on data from their own population (or populations closely matching theirs).

In many cases, existing models may not provide sufficient accuracy for a given population and the organization will need to wait for a critical mass of data to produce a useful model. Further, these models are not something you build, put on the shelf, and bring out for use every now and then, the models must evolve along with the population in focus. Frequent retraining of models as new data arrives ensures that model accuracy will improve over time. Organizations that aggressively collect data and collaborate with their data analysis partners will enjoy the fastest path to accurate models.

SUMMARY

Data-driven relative injury risk prediction is a useful tool to any organization that is attempting to reduce injury rates and minimize the impact of injuries for the people they manage. It requires data aggregation and technology integration partnership from both ML platform providers and human health and performance stakeholders. Risk prediction models do not remove the need for human performance professionals (e.g. physical therapists, trainers, physicians, sport scientists), but they do help those professionals to scale and to focus their efforts more efficiently.

IMAGE CREDITS

Figure 1: Spörri, J., Kröll, J., Amesberger, G., Blake, O. M., & Müller, E. (2012). Perceived key injury risk factors in World Cup alpine ski racing—an explorative qualitative study with expert stakeholders. *British Journal of Sports Medicine*, 46(15), 1059-1064.

Figure 5: [Visualization of a receiver operating characteristic curve] (2018). Create Better Data Science Projects With Business Impact: Churn Prediction with R. *Towards Data Science*, https://towardsdatascience.com/create-better-data-science-projects-with-business-impact-churn-prediction-with-r-f609c23a6287