

# UNCERTAINTY, HEURISTICS AND INJURY PREDICTION

– Written by Mladen Jovanovic, Serbia

Predicting injuries in high-performance sports is of great importance for both players and clubs, but also for fans. Having high-calibre players and athletes healthy and available for both training practices and, most importantly, games or meets is most likely the number one priority of the supporting staff (i.e. performance managers, sport scientists, strength and conditioning coaches and medical personnel). With this in mind, predicting the likelihood of injury and intervening with appropriate actions to reduce that likelihood is a task that is pursued both by practitioners and researchers.

Injury prediction can be represented with a simplistic causal model (Figure 1).

The inputs go by the different names, such as predictors, independent variables, features or sometimes just variables<sup>1</sup>. The output variable, in this case the injury, is often called the response or dependent variable<sup>1</sup>.

What we, practitioners and researchers, are trying to do is to understand this complex relationship between input and output by representing it with simpler and usable models.

According to McElreath<sup>2</sup> and Savage<sup>3</sup> it is useful to make a distinction between large world and small world. McElreath<sup>2</sup> explains:

*“All statistical modelling has these same two frames: the small world of the model itself and the large world we hope to deploy the model in. Navigating between these two worlds remains a central challenge of statistical modelling. The challenge is aggravated by forgetting the distinction. The small world is the self-contained logical world of the model. Within the small world, all possibilities are nominated.*

*(...) The large world is the broader context in which one deploys a model. In the large world, there may be events that were not imagined in the small world. Moreover, the model is always an incomplete representation*

*of the large world and so will make mistakes, even if all kinds of events have been properly nominated. The logical consistency of a model in the small world is no guarantee that it will be optimal in the large world. But it is certainly a warm comfort.”*

In simpler words, we are creating a map ('small world') of the reality ('large world'), by using statistical modelling. We are doing this for three main reasons: prediction and inference<sup>4</sup> and most importantly intervention<sup>5</sup>.

Before getting into each of these reasons, it is important to expand further about statistical models and the nature of the 'large world'.

The analogy that McElreath<sup>2</sup> presented is to think of statistical models as golems – creatures made of clay, created with specific purpose, very powerful, but also very dangerous, because they always do what they are told. It is important to keep this in mind since models involve a

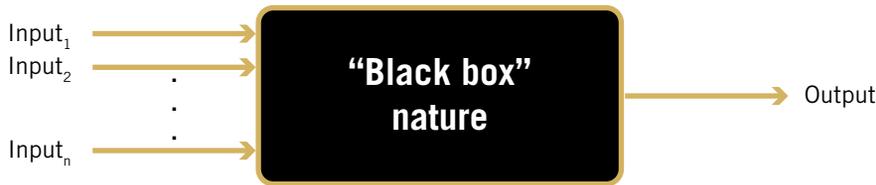


Figure 1: Simplistic causal model of injury prediction.

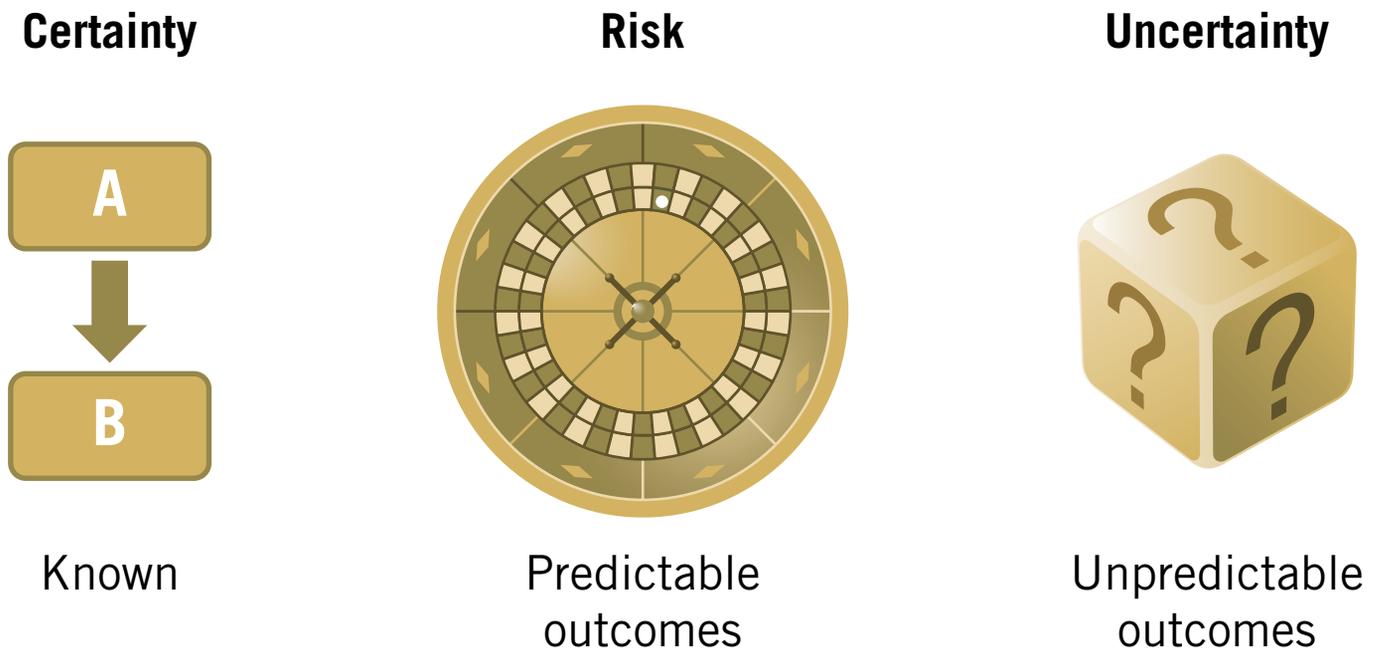


Figure 2: Difference between certainty, risk and uncertainty. Based on the work of Knight<sup>6</sup>.

lot of assumptions; from the way input and output variables are measured and represented, to the assumptions of probability distributions<sup>2</sup>.

Based on the work of Knight<sup>6</sup>, Gert Gigerenzer<sup>7-9</sup> presented differences between certainty, risk and uncertainty in the large world (Figure 2).

Certainty is related to known knowns; risk is related to known unknowns and uncertainty is related to unknown unknowns. This is very similar to Cynefin Framework<sup>10</sup> by Dave Snowden. The problem that Gert Gigerenzer and also Nassim Taleb<sup>11</sup> warn us about is that we are using very complex golems to estimate risks in an uncertain world. So, we are pretty much confusing risk for uncertainty, using models that assume calculable probabilities. Nassim Taleb calls this ludic fallacy<sup>11</sup>.

Similar to Cynefin Framework, Mousavi and Gigerenzer<sup>9</sup> differentiate between different decision-making strategies in certainty, risk and uncertainty. As opposed to Daniel Kahneman<sup>12</sup>, Gert Gigerenzer<sup>7-9</sup>

believes that 'rational' decisions that work very well in the risk world of calculable probabilities, might underperform in the uncertain world compared to heuristics or fast and frugal rules.

The question to be asked is whether issues with injuries belong to calculable world of risk or we need to approach it from the aspects of uncertainty and complexity (for example see Bittencourt et al<sup>13</sup>)? More will be covered regarding this topic in the interventions section of this article, including the discussion regarding heuristics.

#### PREDICTION

As mentioned in the introduction, we are doing statistical modelling for two main reasons: prediction and inference<sup>14</sup>, all with the goal of making better interventions<sup>5</sup>.

With the prediction question, we are approaching problems as a black box (in the sense that one is not typically concerned with the exact mechanisms inside the box, provided it yields accurate predictions)

– we are mostly interested in predictive accuracy<sup>14,14</sup> or how well the model will predict the future data.

It is important here to make a distinction between retrodictive and predictive performance of the model<sup>2</sup>. If we feed data into the model that is used to estimate the parameters, we can ask how well the model reproduces the data used to educate it. But because the model can overfit (learning too much from the data or confusing noise for the signal), this can yield overly optimistic performance<sup>1,2,14</sup>. It is beyond the scope of this paper to go deeper in the topic of overfitting, model bias and variance, trade-off between model accuracy and interpretability and how to deal with those, so the interested reader is directed to explore this further<sup>1,2,14</sup>.

Something readers should be warned about is the fact that not many research papers dealing with injury prediction estimate the predictive performance of the model they use, mostly because they are using and depending on the inferential analysis and questions (see next section).

Breiman<sup>4</sup> critiques this dependency in his famous paper on the differences between two cultures in statistical modelling.

The author would like to propose two simple strategies to estimate predictive performance for future research. In short, predictive performance deals with how the model will perform with unseen data (data that is not used to build the model). The simplest model is to create a hold-out data set and estimate the model performance on that data (Figure 3). The split is usually around 20 to 50% of the original data.

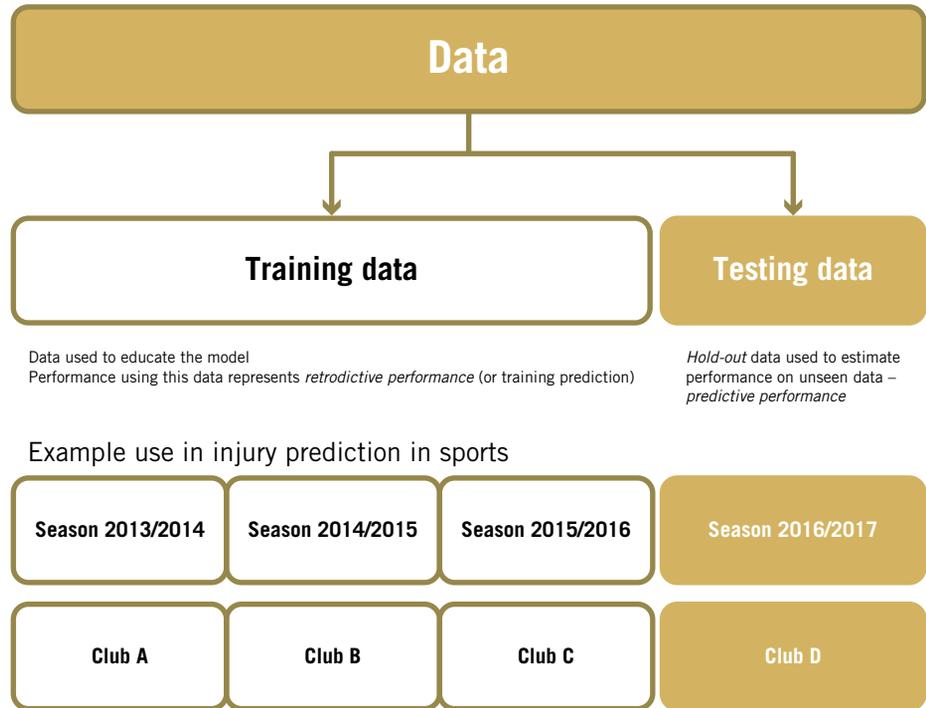
The easiest application of using the training/testing split is for example to build the model using two or three seasons of data and estimate the prediction accuracy on the last season (unseen by the model). To do this, researchers need to collect more data, which is not always feasible.

One way to work around this problem is to use resampling techniques<sup>14</sup>. Resampling techniques for estimating model performance operate similarly to the hold-out technique: a subset of samples are used to fit a model and the remaining samples are used to estimate the efficacy of the model. This process is repeated multiple times and the results are aggregated and summarised<sup>14</sup> (Figure 4).

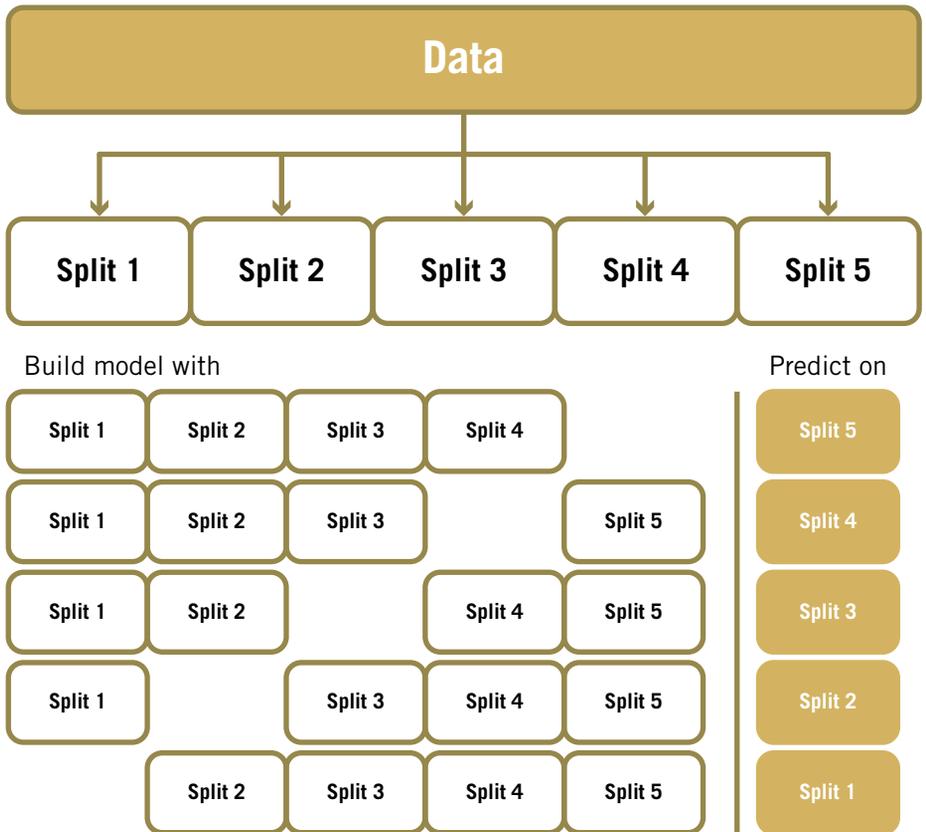
There are numerous resampling techniques (i.e. cross-validation, bootstrap and leave-one-out) and interested reader is directed to books by Kuhn<sup>14</sup> and James et al<sup>1</sup>.

The implementation of a resampling technique might involve building a model for all players in the data set, but leaving one injured player out and then estimating prediction accuracy for him, then repeating this technique for all injured athletes in the data set. This way we get an estimate how the model will perform on the new athlete unseen by the model (for example new athlete joining the club). The split can also be performed on the club level (if the data set involves multiple clubs), where we might want estimate of the model performance on the unseen club data. Deeper discussion on the ways this can be applied is beyond the scope of this paper.

The key message is that injury prediction researchers should provide estimates of the predictive performance of their model, rather than relying solely on inferential analysis and statistics. Using a hold-out data



**Figure 3:** Estimating predictive performance of a model on unseen data using the hold-out technique.



**Figure 4:** Estimating predictive performance using the resampling technique (cross-validation).

set or at least providing estimates using resampling techniques is much-needed in sport injury prediction research.

#### INFERENCE

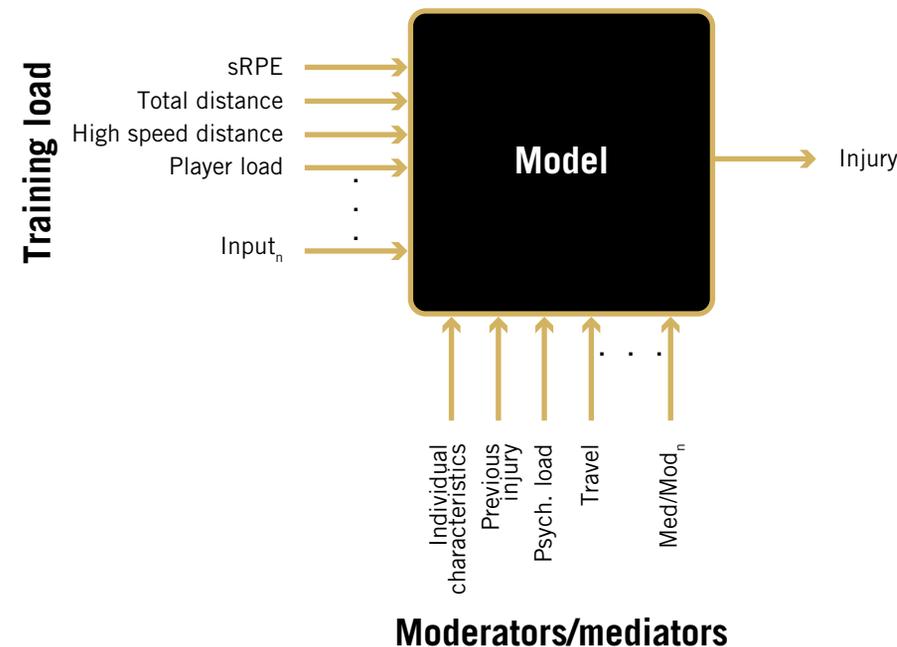
Having a model with good predictive accuracy is not enough if we are interested in answering why injury happened and what predictors are associated with it<sup>4,5</sup>. We cannot answer these questions if we use the same 'black box' approach as we did for prediction. For example, we might be interested in how much injury likelihood will be increased if we increase training load (while adjusting for confounders). This is an inferential problem. If we are only interested in estimating injury likelihood from known predictors, we have a prediction problem. Depending on our goal and questions asked, we might lean more on one approach over the other. Breiman<sup>4</sup> critiqued too much emphasis on an inferential approach (termed Data Modelling Culture, while the black box predictive approach is termed Algorithmic Modelling Culture):

*“Data modelling has given the statistics field many successes in analysing data and getting information about the mechanisms producing the data. But there is also misuse leading to questionable conclusions about the underlying mechanism.*

*(...) an algorithmic model can produce more and more reliable information about the structure of the relationship between inputs and outputs than data models.*

*(...) The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions and has kept statisticians from working on a large range of interesting current problems. Algorithmic modelling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modelling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.*

*(...) With data gathered from uncontrolled observations on complex systems involving unknown physical, chemical or biological mechanisms, the a priori assumption that*



**Figure 5:** Simplistic causal model with moderators/mediators of the effect of training load on injury likelihood.

*nature would generate the data through a parametric model selected by the statistician can result in questionable conclusions that cannot be substantiated by appeal to goodness-of-fit tests and residual analysis. Usually, simple parametric models imposed on data generated by complex systems, for example, medical data, financial data, result in a loss of accuracy and information as compared to algorithmic models.”*

It is beyond the scope of this paper to dwell on these issues in the statistical community, but it is important for the readers to become aware of this tug-of-war between predictive and inferential approaches.

In short, with inferential analysis we are interested in causal knowledge. Potentially the most important use of causal knowledge is for intervention. We don't just want to learn why things happen; we want to use this information to prevent or produce the outcomes<sup>5</sup>.

To do this we must use golems, statistical models with a priori knowledge or structure (for example knowing that training load is a key predictor and age and gender are confounders that need to be adjusted for) and assumptions. Imposing this structure to get causal knowledge is something that Breiman<sup>4</sup> warns about.

To give an example of the issues in creating the small world (model) from the large world (reality) and the level of

assumptions involved, we will use very simple model where training load is the key predictor whose relationship with injury is moderated and mediated by confounders (Figure 5).

- First, what is the unit of analysis – season, club, individual, week, day? How are we representing the data?
- Does the injury refer to all time-missed injuries or only to non-contact injuries? Do we differ between specific location and type of the injury (e.g. hamstring strain vs groin strain)? How do we account for the injuries that are not in the direct interest of the model (i.e. contact quad injury from the game to groin/hamstring/quad pull we want to predict)?
- When it comes to training load, what do we include as predictors? For example, we might use only GPS and sRPE (session rate of perceived exertion) variables, but disregard load done in the gym, because it is not easy to quantify or collect. This is a big assumption – that gym loads do not represent important training load, whereas in the large world we know that doing same running workload after a heavy squat session is not the same as doing it without it.
- How is the training load aggregated? Why use 7- and 28-day rolling ave-rages and not some other combination or

using rolling standard deviation and other descriptive statistics?

- How do we model the latent period – it is known from the research that injury might lag behind highest peak in the training load<sup>15</sup>?
- How do we represent the confounders such as age, gender, previous injury (how is that represented?), training location, temperature, travel, jet lag and so forth?
- How is readiness to perform represented (for example subjective wellness scores or objective performance measures such as groin squeeze, hamstring strength or depth jump reactive strength index)?
- What about psychological load<sup>15</sup>? For example, imagine a team having a winning streak for the last five games or having a losing streak for the last five games. Will doing same running load in training have the same effect on the body under these different psychological conditions?

Taking all of the above assumptions into consideration, we can only conclude that we are indeed creating a golem – a small world we want to use in the large world (making interventions in the complex reality). So we must be very careful in making inferences from the small world to the large world. We must restrain from making sweeping statements that ‘training load predicts injuries’ because of the lack of predictive performance of the model and the fact that we are dealing with a complex golem that has a lot of assumptions. Unfortunately, the general readership often doesn’t understand the difference between the small world and the large world and all the assumptions and data representations involved in the golem.

Also, including all these predictors makes a model complex and decreases its interpretability, besides, it might overfit as explained previously. It is then important to prune the model and decide on the most important predictors that gives us a simpler

model and acceptable precision without overfit. This is important to give us fast and frugal rules or heuristics<sup>2,7-9</sup> that can be used quickly by the practitioners in the large (uncertain) world (Figure 6).

**INTERVENTION**

The point of having all the above causal knowledge is to intervene in the large world. Similar to screening tests to predict injury<sup>17</sup>, where we might be interested in differences between screening-based intervention and intervention alone, we still don’t have the proof in the form of randomised controlled trials to evaluate the effects of load monitoring interventions (e.g. making sure that acute:chronic ratio stays between 0.8 and 1.3) compared to a control group (e.g. usual loading strategies or no monitoring) on injury rates<sup>15</sup>.

It is then important not to jump to conclusions, such as that because training load is associated with injury, we can expect that making interventions and policies

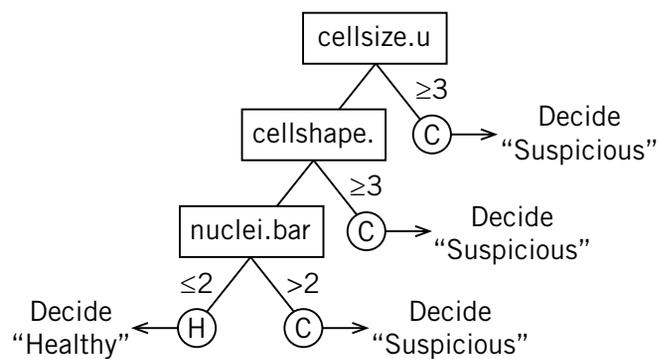
**Logistic regression**

**VS**

**Fast and frugal trees (FFTrees)**

$$p(\text{Suspicious}) = \frac{1}{1 + e^{-(-10.1 + 0.54 \times x_1 - 0.01 \times x_2 + \dots)}}$$

- $x_1 = \text{thickness}$
- $x_2 = \text{thickness}$
- $x_3 = \text{thickness}$
- $x_4 = \text{thickness}$
- $x_5 = \text{thickness}$
- $x_6 = \text{thickness}$
- $x_7 = \text{thickness}$
- ...



**Figure 6:** Simplicity of using the Fast and Frugal Trees compared to the logistic regression in the practice. Adapted with permission from Nathaniel Philips<sup>16</sup>.

		Internal 'readiness'		
		Worse than 'normal'	'Normal'	Better than 'normal'
External workload	More than 'normal'	<b>Inspect</b>	<b>Coping well, but keep an open eye</b>	<b>Inspect</b>
	'Normal'	<b>Inspect</b>	<b>Continue as normal</b>	<b>Coping really well, increase in workload possible</b>
	Less than 'normal'	<b>Inspect</b>	<b>Inspect</b>	<b>Possible detraining</b>

**Figure 7:** Heuristic matrix that can be used to make decisions in the training process using training load and readiness metrics.

based on the training load decreases injury rates (see the Lucas Critique<sup>18</sup>, as well as Taleb<sup>11</sup> and Kleinberh<sup>5</sup>).

The Lucas Critique criticises using estimated statistical relationships from past data to forecast the effects of adopting a new policy, because the estimated model parameters are not fixed, but will change along with a new policy application.

The Lucas Critique might also be involved during data collection, where practitioners might intervene based on the predictions during the study (if we know that an injury is likely to happen, we will not just let it occur, as this is unethical, but will intervene) and hence cancel the effects (see Taleb<sup>11</sup>).

Unfortunately, so far we do not have causal proof that making interventions based on historical data will reduce the likelihood of injury.

Another interesting question to be asked is whether we should let the athletes know the estimated injury risk once we deploy a valid model. How is this going to affect them and will this induce a self-fulfilling prophecy (believing that one is at risk actually increases the risk) or allow players to have a reason to underperform. These are all valid questions to be asked when deploying the model.

Besides the lack of the above-mentioned causal knowledge, we must remember that we are dealing with small world models and trying to use it in the large world, with its associated uncertainties. As mentioned in the previous section, there are a lot of assumptions involved in making a golem and practitioners must realise that they are dealing with much more uncertainty than is explained in the model. That way the estimated injury likelihoods need to be taken together with subjective decisions and beliefs of the practitioner to create an educated guess and the best course of action.

It is also important to realise that practitioners need fast and frugal heuristics (adaptive shortcuts, rules of thumb) that they might quickly use to make fast and informed decisions. It is hence important to simplify the model, which might also perform better in the uncertain world compared to the complex model that might overfit<sup>7-9</sup>. This, of course, needs to be the result of the thorough model building and simplification<sup>2</sup>.

Figure 7 shows a simple model of heuristics that could be used when analysing training load and readiness metrics and making training interventions for a single athlete.

#### SUMMARY

The goal of this paper is to bring to practitioners' awareness the issues with statistical modelling. It is important to realise that we are making statistical golems (models or small worlds) that we need to use in the large world that might be uncertain and complex. It is also important to be aware of the tug-of-war between two statistical cultures, as well as the lack of predictive performance of most injury prediction models. Understanding the complex assumptions that go into the golem makes one realise how difficult it is to deploy the model to the large world and make interventions based on the numbers. We need to be aware of the need for fast and frugal rules or heuristics that might help decision-making in uncertainty. Finally, practitioners must remember the words of famous statistician George Box: "all models are wrong, but some are useful", and use model estimates in combination with their own subjective expertise and beliefs.



## ***Injury prediction researchers should provide estimates of the predictive performance of their model, rather than relying solely on inferential analysis and statistics***



### References

1. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Berlin, Germany: Springer Science & Business Media 2013.
2. McElreath R. *Statistical Rethinking*. Boca Raton, Florida, USA: CRC Press 2016.
3. Savage LJ. *The Foundations of Statistical Inference*. London, United Kingdom: Methuen and Co. 1962.
4. Breiman L. *Statistical modeling: the two cultures*. *Statistical Science* 2001; 16:199-231.
5. Kleinberg S. *Why. Sebastopol, California, USA: O'Reilly Media Inc. 2015*.
6. Knight FH. *Risk, Uncertainty and Profit*. Frederic, Maryland, USA: Beard Books 2002.
7. Gigerenzer G. *Risk Savvy: How To Make Good Decisions*. London, United Kingdom: Penguin 2014.
8. Gigerenzer G, Gaissmaier W. *Heuristic decision making*. *Annu Rev Psychol* 2011; 62:451-482.
9. Mousavi S, Gigerenzer G. *Risk, uncertainty, and heuristics*. *Journal of Business Research* 2014; 67:1671-1678.
10. Snowden DJ, Boone ME. *A leader's framework for decision making*. *Harvard Business Review* 2007; 85:68-76.
11. Taleb NN. *The Black Swan: The Impact of the Highly Improbable*. New York, USA: Random House 2007.
12. Kahneman D. *Thinking, Fast and Slow*. London United Kingdom: Penguin 2012.
13. Bittencourt NF, Meeuwisse WH, Mendonca LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. *Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition narrative review and new concept*. *Br J Sports Med* 2016. [Epub ahead of print].
14. Kuhn M, Johnson K. *Applied Predictive Modeling*. Berlin, Germany: Springer Science & Business Media 2013.
15. Soligard T, Schweltnus M, Alonso JM, Bahr R, Clarsen B, Dijkstra HP et al. *How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury*. *Br J Sports Med* 2016; 50:1030-1041.
16. Philips N. *Making fast, good decisions with the FFTrees R package*. nathanieldphillips.com 2016. Available from: <http://nathanieldphillips.com/2016/08/making-fast-good-decisions-with-the-fftrees-r-package/>. [Accessed November 2016].
17. Bahr R. *Why screening tests to predict injury do not work-and probably never will...: a critical review*. *Br J Sports Med* 2016; 50:776-780.
18. Wikipedia. *Lucas critique*. Wikipedia - The Free Encyclopedia. Available from: [https://en.wikipedia.org/wiki/Lucas\\_critique](https://en.wikipedia.org/wiki/Lucas_critique) [Accessed November 2016].

Mladen Jovanovic M.Sc.  
Strength and Conditioning Coach and  
Sport Scientist  
Ph.D. Student  
Faculty of Sports and Physical Education  
University of Belgrade  
Belgrade, Serbia

Contact:  
coach.mladen.jovanovic@gmail.com