

## Data, More Data, Big Data

As scientists we live with data. As soon as we have a hypothesis to be tested or a problem that needs to be solved, we start gathering data. Usually we have much more data available than time to write manuscripts. For sure we all have files with data that never made it to publication, although we think that there is still a story in those data that needs to be told.

One of the projects during my PhD—30 years ago—was a longitudinal study on talented Dutch junior speed skaters. My objective was to find out why one talented athlete makes it to the elite level and the other promising junior athlete does not. I measured everything I could possibly measure and continued that for 5 consecutive years. I spent numerous long summer days manually digitizing kilometers of 16-mm high-speed-motion film taken during the winter. The skaters came multiple times to the laboratory for testing, and we even did some psychological testing. I had the feeling that I had captured everything from the athletes, so the “tons of data” could give me all the answers. However, although further in their career some of the skaters won Olympic medals and other talents faded away, my data were not giving me clear answers at all.

Nowadays the collection of data is easier than ever. It is even possible to collect data without knowing it. We are all part of one large data project. Our behavior on the Internet, what we buy online or in the store, where and how we travel with our smartphone, our interests on, for instance, ResearchGate or what we search on PubMed, it is all captured and stored in large databases. Artificial intelligence is used to analyze these data to predict or influence our behavior. This process is often called Big Data. It is easy to envisage that in the field of sport science Big Data can play an important role, as well. Beside this, it looks like “Big Data” is also the magic word in getting big research funding. With this in mind, together with the large piles of data on my shelf, I recently attended a meeting on Big Data in sports and came across the book of Timandra Harkness, *Big Data: Does Size Matter?* (Bloomsbury Sigma, 2016).

From the meeting on Big Data in sports I came home a little confused and disappointed. The data scientists proposed to put,

nationwide, all data collected on athletes into one large database so that they could process it into valuable information on performance-determining factors. I was thinking of my tons of data, collected with great precision and with all ethical amenability, mingled with data from less-secure origins. How will quality control be done? Who is taking care of the ethical issues? How can data based on the past be used to predict an athlete’s potential regardless of eventual training or behavioral changes? From the Harkness book I became aware that we only measure what we *can* measure and we make only *this* important in our analysis. There is much more left that we can’t quantify. I realized that the most important things in life could not be quantified!

Big Data is converting large amounts of data into algorithms to make quantitative decisions or predictions. That can be done with great precision, and I have no doubts about the craftsmanship of data scientists. If we make use of Big Data, these algorithms become our tools—tools like the tools we have in our laboratories, only a little bit different in nature. Here the importance for our journal comes into play. How should we report on these algorithms? Should they be fully available to the reader? I hope sincerely that the answer to the last question is yes. I’m not aware of manuscripts based on Big Data submitted to the *International Journal of Sports Physiology and Performance* yet, but I’m sure they will come. I hope that special attention in the review process will be given to the origin of the data, to the ethical procedures in gathering the data, and to the full disclosure of the algorithms on which the outcomes of the study are based. As Harkness says, the algorithms of Big Data can only quantify and not qualify. All information from Big Data needs human judgment, so extensive attention needs to be paid to the interpretation and qualification of the decisions and predictions of the analyses.

*Jos J. de Koning, IJSP Associate Editor*