BEYOND STATISTICAL SIGNIFICANCE: UNIFYING THE LANGUAGE BETWEEN SPORT SCIENTISTS AND COACHES

[1]Tara K. Whiton, [1]John P. Wagle, [1]Kimitake Sato

[1]East Tennessee State University

Introduction. It is the responsibility of a sport scientist to make research translatable to coaches and applicable to sport and athletes with the goal of improving training practices and ultimately performances. Batterham and Hopkins (2006) raised an important perspective regarding translatable research:

> "hypothesis testing is illogical, because the null hypothesis of no relationship or no difference is always false—there are no truly zero effects in nature. Indeed, in arriving at a problem statement and research question, researchers usually have good reasons to believe that effects will be different from zero".

Therefore, hypothesis testing alone may be illogical within the context of sports training and performance. In addition, two primary complications exist that may interrupt the valuable communication between scientists and coaches. First, the use of traditional statistics in the research and their potentially complex interpretation requirements. Second, collecting data on high-level athletes provides a unique challenge as traditional statistics do not always provide the most appropriate strategies or direct applications to the nature of sport performance. For example, the miniscule margins of victory in track and field or the influence that changing terrains has on tactics in trail or snow-based sports, make evaluating change difficult with traditional approaches. Additionally, small sample sizes are often the reality of athletic research populations due to limits set by team size, star-performers, accessibility, resources, and scheduling, to name a few. By identifying and acknowledging these potential shortcomings, awareness is brought to both sport scientists conducting the research and the coaches reading and applying the results. The purpose of the current manuscript is to briefly highlight alternative statistical analyses and to encourage sport scientists to consider incorporating these techniques, ultimately adding more context and applicability to the unique nature of sport.

Practical Data Analysis Methods To Overcome Lack Of Transferability From Traditional Statistics. Evidence-driven programming requires coaches to stay current with scientific literature. However, a primary barrier to the application of scientific research to sport lies within the interpretation of oftentimes complex statistical results. Traditional statistical analysis is rooted in null hypothesis testing, which is concerned with determining the likelihood that the null hypothesis is true - a system that is very binary. In sport, we are often concerned with the magnitude of change or the meaningfulness of that change. As a result, there can be instances when important changes in sport performances go undetected due to lack of statistical significance as athletic gains are often marginal at best. Additionally, not only is there a growing demand for collecting objective data to make informed decisions, but there have been major advances in the amount and usability of technology to meet such demands. Even with such robust, objective, near-lab-quality data at the practitioner's disposal, there is a disconnect in understanding what the data is measuring, and the statistical methods used in the field compared to research. Though this is an appropriate and necessary situation in many cases, it is not that simple in sport science as the sole purpose is to provide practical interpretations of treatment effects using a background of knowledge of a variety of scientific principles relating to physiology, biomechanics, nutrition, and psychology, to name a few, with the central goal of maximizing an athlete's performance. The purpose and foundation sport science is derived from is what drives a scientist's experimental and statistical design. If, however, the same data analysis and interpretive strategies were used by sport scientists and coaches, then coaches may have a better understanding of not only the potential for practical application of lab-based findings, but the quality of interpretation of their own field-collected data. This indirect quality control may help coaches be more confident in their decision making and avoid a reversion back to naturalistic decision making (Jeffreys & Moody, 2016). Therefore, it may be logical to transition towards two reporting techniques within the literature, where there is room for both more traditional statistical approaches and other strategies that can bridge the gap between sport science researchers and practitioners. Thanks to the efforts of pioneers like Will Hopkins, there is momentum towards such a reality. Two of the strategies that have been previously discussed as useful tools in sport science are percent change (Hopkins, 2000) and effect size (Hopkins, 2008). Both measures can provide either a stand-alone interpretation of the data or can augment traditional approaches within a broader context of interpretation.

Percent Change. Percent change is valuable in athlete monitoring because it considers the baseline value and the magnitude of changes observed that are relative to the starting point or the individual. Oftentimes, these pre-post changes are evaluated by performing a paired-samples t-test for repeated-measures. However, complementing the significance level with a percent change value can provide more relatable information, allowing the coach the opportunity to make a well-informed decision about future programming. For example, Nobel & Chapman (2017) used percent change alongside statistical analysis to communicate the rate of performance improvement to the readership as this more appropriately represented the relatively small changes in marathon time over the course of a year in elite African runners compared to non-African runners. Incorporating percent change into the results allowed the reader to see which group had sharper performance increases despite differences in absolute marathon time. Another common use of percent change in endurance sports is in identifying pacing strategy throughout a performance. A chosen pacing strategy is reflective of an athlete's ability to distribute work throughout an event and is inherently dictated by the athlete's ability to resist fatigue via peripheral sensory input (Abbiss & Laursen, 2008). Abbiss & Laursen (2008) profiled pacing strategies for a variety of endurance-based distance events by categorizing them as successful by adopting a negative (start slower, finish faster), positive (start faster, finisher slower), or evenly-paced splits throughout a race. Since tactics and terrain can often influence speed, calculating the percent change from one split or checkpoint to the next can provide a better indication of performance improvement and pacing strategy and allow a comparison from event to event.

Effect Size. As mentioned above, Batterham and Hopkins (2006) pointed out that in most cases, researchers have reason to believe that effects will be different from zero, therefore, hypothesis testing should be complemented with other inferential statistics. An effect statistic is a complementary measure that refers to the magnitude of an effect that one variable has on another. A non-significant result does not necessarily imply a lack of worthwhile effect, and, practical relevance in sport science is not necessarily seeking whether there is an effect, but how big the effect is (Batterham & Hopkins, 2006). Calculating effect sizes takes into consideration the variance of improvement in performance and can be accurately compared within or across groups since effect sizes have been standardized (Rhea, 1994). Hopkins revised the interpretation of Cohen's original effect magnitude scale (0.1, 0.3, and 0.5 for small, moderate, and large effects, respectively) to better suit sport science by adding 0.7 and 0.9 to the scale for very large and extremely large effect sizes (Cohen, 1988; Hopkins 2006, 2008). This not only demonstrates the need to adapt statistical interpretations within the context of sport, but also the need to have a more relatable language with which to communicate between researchers and practitioners.

Vikmoen et al. (2017) provided an example of this strategy by determining a large effect on endurance performance through the addition of resistance training. Though following identical endurance training prescriptions, those who added resistance training to their regimen experienced a meaningful reduction in oxygen consumption during the last hour of a prolonged cycling trial compared to those who performed endurance training alone (Vikmoen et al., 2017). By providing a magnitude threshold, practitioners can then have a clearer approximation of effect magnitude if they were to provide a similar intervention. Additionally, researchers can use the presence of an effect to engage in educated speculation as to what mechanisms may drive such a change. In any sport monitoring case, effect size can help determine efficacy of any training intervention by providing coaches with qualitative magnitudes and comparable standards.

A variety of metrics and indices can be analyzed using both percent change and effect sizes. A common example in sport monitoring is the reactive strength index modified (RSI-mod) which considers jump height and ground contact time (Barker et al., 2017). The components of RSI-mod are global performance indicators, and can therefore be used as a tool for monitoring performance improvement as well as acute fatigue management (Barker et al., 2017). However, simply collecting metrics such as indices does not provide enough context to the sport coach in terms of rate or magnitude of change. By implementing either percent change or effect size to monitor changes in RSI-mod throughout a training season, a coach can visualize changes relative to baseline measurements, as well as magnitude of change, providing tangible evidence with which to structure future training programs.

Summary. These few examples are provided to stimulate thought and debate about changing data analysis and reporting in both the research and coaching setting. There is a slight disconnect and lack of effective communication of research findings between coaches and scientists due to a scientist's reliance on traditional statistical models, and a coach's need for practical and transferable data to the field. A complex statistical model is not always necessary or appropriate for a coach to make an informed decision in training and competition, however, the inferential measures above can support both parties. While this list is nowhere near exhaustive, each example presents the use of practically meaningful methods that can easily co-exist with traditional statistical practices. If scientists use these strategies to communicate the findings of their research, and coaches use them in monitoring the progress of their athletes, then both groups are speaking the same language. In conclusion, we hope to unify both parties as now scientists can access greater readership from a broader audience, and the information provided is more contextual for coaches, allowing them to use a more evidence-based approach to programming.

REFERENCES:

Abbiss, C. R., & Laursen, P. B. (2008). Describing and Understanding Pacing Strategies during Athletic Competition. *Sports Medicine*, *38*(3), 239-252.

Barker, L. A., Harry, J. R., & Mercer, J. A. (2017). Relationships Between Countermovement Jump Ground Reaction Forces and Jump Height, Reactive Strength Index, and Jump Time. *Journal of Strength and Conditioning Research,* 1.

Batterham, A. M., & Hopkins, W. G. (2006). Making Meaningful Inferences about Magnitudes. *International Journal of Sports Physiology and Performance, 1*(1), 50-57.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*.

Hopkins, W. G. (2000). Measures of Reliability in Sports Medicine and Science. *Sports Medicine,30*(5), 375-381.

Hopkins, W. G. (2006). A new view of statistics. http://sportsci.org/resource/stats/effectmag.html

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2008). Progressive Statistics for Studies in Sports Medicine and Exercise Science. *Medicine & Science in Sports & Exercise,41*(1), 3-13.

Jeffreys, I., & Moody, J. (2016). *Strength and conditioning for sports performance*. New York, NY: Routledge

Noble, T. J., & Chapman, R. F. (2017). Elite African Marathoners Specialize Earlier than Elite Non-African Marathoners. *International Journal of Sports Physiology and Performance,* 1-19.

Rhea, M. R. (2004). Determining the Magnitude of Treatment Effects in Strength Training Research Through the Use of The Effect Size. *Journal of Strength and Conditioning Research, 18*(4), 918-920.

Vikmoen, O., Rønnestad, B. R., Ellefsen, S., & Raastad, T. (2017). Heavy strength training improves running and cycling performance following prolonged submaximal work in well- trained female athletes. *Physiological Reports, 5*(5).