Measurement issues in sports performance analysis: the impact of unrepresentative data and limited reliability

Peter O'Donoghue.

Cardiff School of Sport, Cardiff Metropolitan University, Cyncoed Campus, Cardiff, UK.

Introduction

This keynote address contains two parts. The purpose of the first part is to consider validity and objectivity using a computer algorithm to automatically identify challenges in soccer (Bangsbo and Peitersen, 2002). The second part considers reliability and representativeness of sports performance data using the percentage of successful passes made by professional soccer players.

Methods

Study 1: A set of 10 criteria were used by a computerised algorithm to identify challenges in soccer using a combination of player tracking and match event data. Four of these were essential to identify 174 potential challenges events within the match with different combinations of the other factors applied in an exploratory approach. The challenges identified by the algorithm were compared with those identified by the author's subjective observation of the match video.

Study 2: A synthetic population of 250,000 players with 20 matches per player were was created using within- and between-player distributions for the percentage of successful passes variable. There were positional differences but no regional differences. A set of 1000 samples were drawn from the data to compare positional and regional effects using different sample sizes, matches per player and simulating different percentage error values.

Results

Study 1: 36 of the 37 challenge events identified by video observation were successfully identified by the algorithm when no criteria applying to the receiving player were applied. However, there were also 111 "false positive" challenges identified by the algorithm. When criteria relating to the receiving player were applied, the algorithm only identified 3 of the 37 challenges that were identified during video observation. However, the number of false negative challenges reduced to 27.

Study 2: Sample size had the largest impact on the conclusions drawn about positional effect using a one-way ANOVA test. When sample size was 30, fewer than half of the samples confirmed the positional effect in the population. More than 90% of the samples confirmed this effect when sample size was greater 150 and player data was based on 3 or more performances per player. There was little additional benefit to representing typical player performance by more than 3 performances. Where reliability was limited with 5% or 10% error in the percentage of successful passes, there was a slight reduction in the number of samples confirming the positional effect. When examining regional effect with an independent samples t-test, fewer than 10 of the 1000 samples led to a significant result (a Type I Error) for any combination of 5 sample sizes, 10 values for matches per player and 3 percentage error values. Interestingly, using unrepresentative data from only one performance per player protects against Type I Errors as does limited reliability (10% error).

Discussion & Conclusion

The validity of a tactical analysis algorithm needs to consider the number of misidentified cases as well as the relative seriousness of "false positive" and "false negative" cases. False negative cases are more serious in the example used because the analyst will not be alerted to these by the

algorithm. In studying sports performance, it appears that it is better to have larger sample sizes than smaller sample sizes with players being represented by multiple match performances. Measurement error increases the probability of a Type II Error but may protect against making Type I Errors.

References

Bangsbo, J. and Peitersen, B. (2004), **Offensive soccer tactics: how to control possession and score more goals**, Champaign, Il: Human Kinetics.