SPORTSCIENCE sportsci.org
Perspectives / Research Resources

# Commentary on *Making Meaningful Inferences About Magnitudes*

Stephen W Marshall

Sportscience 9, 43-44 (sportsci.org/jour/05/swm.htm)

Departments of Epidemiology and Orthopedics, University of North Carolina at Chapel Hill, Chapel Hill NC 27599-7435, USA. Email.

Reprint pdf · Reprint doc

Batterham and Hopkins have proposed a new approach for reporting the statistical findings from research studies. Their technique combines information on the magnitude of the estimate of the effect (e.g., mean difference), the degree of imprecision about that effect (e.g., the confidence interval), and the smallest difference that has real-world (or clinical) meaning. This information is combined into an overall set of likelihood statistics and a set of short descriptors (*likely beneficial*, etc.) are proposed. In this commentary, I address three issues: Is this approach better than using p-values? Is this approach useful with observational data? What are the drawbacks of this approach? I particularly comment on the usefulness of their approach for epidemiologists and other researchers who work with observational data.

## Is This Approach Better Than Using P-Values?

Batterham and Hopkins correctly object to the painful reductionism associated with formal tests of statistical significance (the null-hypothesis test). As they point out, the statistical theory underlying this approach is counter-intuitive and mysterious to almost all scientists who use it. Most researchers fail to comprehend that failure to reject the null is not the same as accepting the alternative. More importantly, researchers frequently ignore important information in their data purely because the magical p-value of 0.05 has not been obtained. An approach based on the magnitude of the estimate is vastly preferable to the unfortunate binary mindset of accept/reject that null-hypothesis testing engenders (Rothman, 1978; Poole, 1987; Poole, 2001; Wolf and Cumming, 2004). The Batterham and Hopkins approach has the advantage of being a sophisticated quantitative alternative. It incorporates a Bayesian approach but circumvents the issues involved in defining priors. As such, it is highly attractive and researchers should be encouraged to adopt it.

## Is This Approach Useful With Observational Data?

Batterham and Hopkins largely develop their approach from within the paradigm of experimental statistics. Observational studies, however, have additional complexities associated with the use of non-random samples and independent variables (risk factors, or exposures) that are not randomized, and perhaps can never be randomized (Greenland, 1990). Randomization may be impossible for ethical reasons (e.g. if the effect of interest is cigarette smoking) or logistical reasons (e.g. if the effect of interest is air pollution), or both. In observational data, issues associated with confounding, selection factors, and misclassification of effects pose at least as a great source of uncertainty as the imprecision of estimates (Greenland, 1990; Greenland, 1998). Thus, the Batterham and Hopkins approach, to have maximum utility for epidemiologists, should incorporate quantitative information on the likely effect of these non-random sources of error (bias). Recent work has investigated the use of simulation techniques for the quantitative assessment of bias (Lash and Fink, 2003; Fox 2005; Greenland 2004; Greenland 2005; Fox 2005), and these methods can be used to produce an *uncertainty interval*–an interval that is based not just on imprecision (random error) but also includes information on the effect of bias (systematic error).

The beauty of the Batterham and Hopkins approach in this regard is its flexibility. Although Batterham and Hopkins demonstrate the method using the confidence interval (random error only), their approach will work equally well using a simulated uncertainty interval (random and systematic error). Use of an uncertainty interval, rather than a confidence interval, will allow epidemiologists to include non-random sources of error into the Batterham and Hopkins approach. This makes the use of the method a very attractive tool for epidemiologists. As a minor point, epidemiologists applying the Batterham and Hopkins approach should note that, for measures of effect that are based on ratios (such as odds ratios, rate ratios, and hazard ratios), the X-axis in Figure 3 should be plotted on the logarithmic scale.

## What are the Drawbacks of This Approach?

Beyond pointing to Cohen's scales of magnitudes, Batterham and Hopkins do not provide any guidance about how to determine the smallest worthwhile effect. Does this need to be defined before the data analysis is conducted? Can one change one's mind about the smallest worthwhile effect after reviewing the data analysis? If so, what is the effect on the validity of the conclusions? The answers to these and other questions about determining the smallest worthwhile effect await further research and guidance. One thing seems clear: two groups of researchers who use different criteria for selecting the smallest worthwhile effect will, even given the same data, arrive at different conclusions. Although this sounds like a weakness, it could be seen as a strength of the method, since requires that researchers make explicit what number they consider to be the smallest worthwhile effect.

The great strength of the Batterham and Hopkins method is that it does incorporate information on the smallest worthwhile effect into the formal presentation of data. The great drawback is, in many cases, there may be little data and limited consensus on what the smallest worthwhile effect should be. The “ballpark” estimates often used to motivate power calculations in a research grant proposal are unlikely to be sufficiently refined to fulfill a useful function in the analysis phase of a study.

## Conclusion

In summary, Batterham and Hopkins have proposed a simple yet powerful method for presenting the findings of research studies. Their presentation combines information on the magnitude of the estimate, the degree of imprecision, and the smallest difference that has “real-world” (or clinical) meaning. For epidemiologist, their method can readily be extended to include sources of uncertainty other than random error using multiple bias models. However, I suspect that some clarification, guidance, and resolution of issues around selecting the numbers to be used as smallest worthwhile effects will be required if the Batterham and Hopkins method is to achieve its full potential. Despite this possible limitation, the technique provides a useful tool for discouraging the mindless dependence on null-hypothesis tests that pervades science. Use of the Batterham and Hopkins method will encourage a move towards less null-hypothesis testing and more estimation of effects. It is also expected to promote a thoughtful analysis of, and reflection upon, study data and findings.

## References

Greenland S (1990). Randomization, statistics, and causal inference. Epidemiology 1, 421-429

Greenland S (1998). Basic methods for sensitivity analysis and external adjustment. In: Modern Epidemiology (2nd Edition), Rothman K, Greenland S (Eds). Lippincott-Raven: New York, NY, pp.343-358

Greenland S (2005). Multiple-bias modelling for analysis of observational data. Journal of the Royal Statistical Society A 168, 267-306

Greenland S (2004). Interval estimation by simulation as an alternative to and extension of confidence intervals. International Journal of Epidemiology 33, 1389-1397

Fox MP, Lash TL, Greenland S (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. International Journal of Epidemiology Advance Access published September 19

Lash TL, Fink AK (2003). Semi-automated sensitivity analysis to assess systematic errors in observational data. Epidemiology 14, 451-458

Phillips CV (2003). Quantifying and reporting uncertainty from systematic errors. Epidemiology 14, 459-466

Poole C (1987). Beyond the confidence interval. American Jounral of Public Health 77, 195-199

Poole C (2001). Low p-values or narrow confidence intervals: which are more durable? Epidemiology 12, 291-294

Rothman KJ (1978). A show of confidence. New England Journal of Medicine 299, 1362-1363

Published Dec 2005