Perspectives / Research Resources

Commentary on *Linear Models and Effect Magnitudes*

Alan M Batterham

School of Health and Social Care, Teesside University, Middlesbrough TS1 3BA, UK. Email.
Reprint pdf · Reprint doc

This latest contribution to the pool of resources at Sportscience is an excellent learning, teaching, and research resource valuable for researchers and research consumers at all levels. In the form of a PowerPoint presentation, it may be used for upper level teaching (in whole or in parts) and also serves as a reference source for experienced researchers. The presentation builds on and complements the Magnitude Matters slideshow and the Progressive Statistics article published in January 2009 in MSSE.

For me, the crux of the presentation is on Slide 6, emphasizing the fact that the right question is not *whether* there is an effect but *how big* is the effect. As Will highlights, answering this question requires an a priori definition of the smallest worthwhile effect. This is by no means a trivial task, but it is one that must not be shirked in hiding behind a null-hypothesis testing framework. (As Will once remarked famously from the podium in an ACSM symposium, "if you don't know what matters for your patients or clients, quit the field!") An illustration of the importance of this problem is the recent call by the UK Medical Research Council/ National Institute for Health Research Methodology Research Programme for proposals concerned with "how to specify the targeted difference for a randomised controlled trial." Dr Jonathan Cook (University of Aberdeen) is now leading this project, which will result in draft guidance for researchers and funding bodies, including separate sections for different types of trials and on different ways in which the outcomes of a treatment might be measured.

There are three main methods for arriving at a minimum important difference; anchor-based methods, distribution-based methods, and opinion seeking. Will notes in the presentation that clinicians can't agree on a value for the smallest worthwhile effect and that in the absence of clinical consensus we need a statistical default. The approach Will takes is therefore an example of a distribution-based method, in which changes in scores on an outcome are evaluated in relation to the variability in scores for that outcome (e.g., thresholds for the standardised mean difference). In anchor-based methods the aim is to establish the change in the outcome being measured required to result in a meaningful change on another measure which has already proven to be clinically or practically important to the individual. For example, a single-anchor method might involve assessing the change in maximum oxygen uptake required for people to rate their health-related quality of life (the anchor) as much improved. In my experience, robust anchor-based approaches are rare in our field, and a statistical distribution-based default is sensible. Moreover, some work has suggested a reconciliation of anchor-based and distribution-based approaches, with a near-linear relationship between effect size and the proportion of patients benefiting from a treatment (Norman et al., 2001).

My remaining comments relate to specific sections of the presentation.

- Slide 7 gives an example of two predictors (Strength = a + b*Age + c*Size) with the statement that such models allow us to work out the "pure" effect of each predictor: "That is, yeah, kids get stronger as they get older, but is it just because they're bigger, or does something else happen with Age? The *something else* is given by the 'b'. It's that simple!" I would add a caveat here to check for potentially degrading collinearity. This is pertinent to the example given, as age and body size may be highly related in growing and maturing children. Collinearity does not violate any of the assumptions of ordinary least-squares regression and thus gives unbiased predictions from the linear combination of predictors. However, if your goal is explanation relating to the *relative importance of individual predictors*, then collinearity could be a problem, as it may be difficult to determine the separate influence of each. Collinearity results in large standard errors for the af-

fected coefficients (variance inflation) and is essentially a data problem: insufficient data information (signal) relative to the noise. Sophisticated collinearity diagnostics are available in many statistical software packages, including SAS and SPSS.

- On Slides 15 or 16 it would have been helpful to the reader if there were a note or link to the source or derivation of the Hopkins scale of effect magnitudes (for example, the progressive statistics paper), given that it differs from Cohen's scale and that this presentation may be the first stop for some researchers.
- People often get very confused about the difference between *partial* and *semi-partial* correlations, and which is better, so I found the plain-language explanations on Slide 21 very useful.
- It crossed my mind when you were dealing

with distributional issues, non-uniformity, and transformations that bootstrapping should get a mention somewhere. Bootstrapping provides trustworthy confidence limits when some of the assumptions underlying the linear model are violated, including one you didn't mention directly, independence of the observations.

Norman GR, Sridhar FG, Guyatt GH, Walter SD (2001). Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. Medical Care 39, 1037-1038

Back to article/homepage

Published July 2010
©2010