Letter

How should clinicians interpret results reflecting the effect of an intervention on composite end points?

This article contains extra text on the EBM website

Evidence-Based Medicine recently published a thoughtful commentary by Montori *et al*,[1] addressing the above question, prompted by the Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS),[2] which was abstracted in the same issue.[3] The ACHOIS investigators compared a screen-and-treat programme for gestational diabetes with routine pregnancy care and used as their main end point a composite of 4 outcomes: shoulder dystocia, Duchenne-Erb palsy, fracture, and death—a natural choice given that a main purpose of gestational glycaemic control is to prevent the baby from being born too big. The topic is important and several points are worth noting.

Montori *et al* listed 3 questions to help decide the appropriateness of a composite outcome:

(a) Are the component outcomes of similar importance to the patients?

(b) Do the more and less important outcomes occur with similar frequency?

(c) Are the component outcomes likely to have similar relative reductions of the risk (RRRs)?

It seems to me that a clinically meaningful composite endpoint analysis would require either condition (a) or (c) to be satisfied, whereas condition (b) is a secondary matter. In the ACHOIS study, shoulder problems were much more common, and obviously much less serious, than death. With respect to RRRs, Montori *et al* were uncertain how to answer question (c). If (b) had been satisfied, the decision problem would be a little easier to explain to expectant mothers, but it is (a) and (c) that decide whether a logically defensible lumped analysis can be devised.

What happens if a patient experiences 2 of the outcomes? The composite endpoint approach in ACHOIS implies that, if the infant dies, a broken clavicle does not matter. This statement would also be true in the case of co-occurrence of, say, palsy and fracture. The endpoint definition capitalises the form ''one or more of the following…''. In other words, co-occurrence of component undesirable outcomes does not really matter. Otherwise, the questions above have to be modified to deal intelligently with co-occurrence.

When deciding what should define a good or poor final result, one should respect the relative seriousness of the imaginable overall results (by ''overall results'' I mean elementary outcome features and combinations thereof). Shoulder dystocia, a poor result, must be defined as ''shoulder dystocia or any neonatal event, or combination of events, more serious than that,'' and that is precisely how the ACHOIS team defined it. The frequency of shoulder dystocia alone is a bad summary statistic because it lumps deaths and uneventful deliveries into the category of ''no shoulder dystocia.'' Or, to be more precise, it provides no guidance as to which regimen is better and may actually mislead the reader; it does add a little to our body of knowledge about shoulder problems during delivery, but that is all.

In summing up, I agree with Montori *et al* that composite statistics are easy to interpret when—and perhaps only when—the components either have similar RRRs (which will then also be the composite RRR) or are of similar importance to the patient (so that their composite RRR decides the clinical issue). In principle, one should treat a bundle of co-occurring outcome features as a single outcome when listing candidate components for a composite end point (''consider relevant intersections in selecting relevant conjunctions''). Serious difficulties exist in handling composites of non-simultaneous events as used, for example, by many research teams in cardiology, including my own.[3] Finally, composite end points should be so designed as not to lump together good and poor courses of illness.

Note from editors: This letter was shortened because of space considerations. Please see www.evidence-basedmedicine.com for the full version of this letter.

JØRGEN HILDEN, MD
University of Copenhagen, Department of Biostatistics
Copenhagen, Denmark

1 Montori VM, Busses JW, Miralda GP, *et al*. How should clinicians interpret the effect of an intervention on composite end points: should I dump this lump? *Evid Based Med* 2005;**10**:162–3.
2 Crowther CA, Hiller JE, Moss JR, *et al*. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med* 2005;**352**:2477–86.
3 Jespersen CM, Als-Nielsen B, Damgaard M, *et al*. Randomised placebo controlled multicentre trial to assess short term clarithromycin for patients with stable coronary heart disease: CLARICOR trial. *BMJ* 2006;**332**:22–27.

Correction

In the October 2006 issue of *Evidence-Based Medicine*, calcuation errors were detected in the table of the abstract for the article by Prince *et al*.[1] The numbers needed to treat (NNTs) for compliant patients were incorrect.

The correct NNTs for calcium *v* placebo in compliant patients are these:

Any fracture NNT 21 (95% CI 13 to 235)
Appendicular fracture NNT 22 (CI 13 to 225)
Upper limb fracture NNT 34 (CI 24 to 238)

1 Calcium did not prevent fractures in elderly women [abstract]. *Evid Based Med*. 2006;11: 149, Abstract of:Prince RL, Devine A, Dhaliwal SS, *et al*. Effects of calcium supplementation on clinical fracture and bone structure:results of a 5-year, double-blind, placebo-controlled trial in elderly women.*Arch Intern Med*, 2006;**166**:869–75.