

Jottings

Statistical methods are important. We have all seen the impact of shifting from p values to confidence intervals and from relative (RRR) to absolute measures (ARR, NNT) of effect, and how these can influence the interpretation and implementation of study results. A continuing controversy is over the use of multiple outcomes. In this issue, we thought the results of the study by Crowther *et al* important enough to be one of the 20 articles we chose from the 100,000 published in the past 2 months. But we've accompanied this by an editorial on the interpretation issues. We would welcome your comments on this difficult issue.

If you enjoy discussion about EBM topics, then you might wish to sign on to the Evidence-Based Health email list at: www.jiscmail.ac.uk/lists/EVIDENCE-BASED-HEALTH.html

In addition to alerts about forthcoming courses related to EBM, the list has a lively discussion of topics in evidence-based health care. Some recent threads have included an explanation and history of Simpson's paradox, outcomes research, the impact of dishonest reporting of research on

EBM, and of course, that perennial discussion about the relative merits of RCTs and observational studies. In the course of that thread Badri Badrinath pointed out the fascinating article by John Ioannidis (Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005 Jul 13;294:218–28), in which the law of initial results ("first results are always spectacular, and subsequent ones are mediocre") was tested using the results of subsequent research. The author found that "most of the observational studies had been contraindicated, or subsequent research shown substantially smaller effects, but most randomised studies had results that had not been challenged." Happy reading,

THE EDITORS
Evidence-Based Medicine

EBM notebook

How should clinicians interpret results reflecting the effect of an intervention on composite end points: should I dump this lump?

In this issue, *Evidence-Based Medicine* highlights and comments on a landmark randomised trial, the ACHOIS trial.¹ In this trial, Crowther and her colleagues enrolled 1000 pregnant women with mild gestational diabetes who were either informed of their diagnosis and received treatment (including individual dietary advice, instruction on self monitoring of blood glucose concentrations, and insulin therapy as needed) or informed that they did not have gestational diabetes and assigned routine prenatal care unless subsequent findings suggested diabetes.

The ACHOIS trial was designed to determine the extent to which aggressive treatment of mild gestational diabetes affects outcomes in the mothers (induction of labor and caesarean section) and in the infants (a composite of death, shoulder dystocia, bone fracture, and nerve palsy dubbed "any serious perinatal complications").

For the maternal outcomes, 189 of 490 (39%) of mothers in the intervention group and 150 of 510 (29%) in the routine care group required induction of labour; 152 (31%) and 164 (32%) required caesarean delivery. We will examine the implications of these results using numbers needed to treat (NNTs) under the assumption that all women with mild gestational diabetes are at the same risk of adverse outcomes and that their risk approximates that of the average woman enrolled in the ACHOIS trial (women at identifiably higher risk of these outcomes would have smaller NNTs, and *vice versa*). While the trial evidence suggests no difference in the rate of caesarean sections, it tells us that for every 11 (95% CI 7 to 31) women screened and managed for gestational diabetes, 1 additional woman will require induction of labour. Furthermore, clinicians admitted 357 of 506 (71%) and 321

of 524 (61%) infants in the intervention and routine care groups to the neonatal nursery, respectively. Thus, for every 11 (CI 7 to 29) women intensively managed, 1 additional infant will require admission to the neonatal nursery.

One might justify the effort of screening and managing gestational diabetes, the higher rate of labour inductions, and the increase in neonatal nursery admissions if the intervention reduced serious perinatal complications in the infants. Seven (1%) and 23 (4%) of the infants suffered the composite end point "serious perinatal complications," a difference that was statistically significant ($p = 0.04$) and indicated that clinicians must screen and manage 34 (CI 20 to 103) women with mild gestational diabetes to prevent 1 serious perinatal complication.

Many would find the 3% absolute reduction in the risk of the combined end point "serious perinatal complications" (relative risk reduction [RRR] 67%, CI 25% to 86%) a favourable tradeoff. But should they interpret the study results this way? Or should readers rather focus on the effects of treatment on each of the individual components of the composite?

We have recently suggested a series of questions that can help clinicians interpret studies using composite end points.² We will apply these guides in examining the validity of the composite end point "any serious perinatal complications."

ARE THE COMPONENT OUTCOMES OF SIMILAR IMPORTANCE TO PATIENTS?

If parents considered the death of an infant, nerve palsies, shoulder dystocia, and bone fractures of similar importance, then it would not matter how the 67% RRR or the 3% absolute risk reduction in the composite end point was distributed across its components. It is certain, however, that

parents would consider perinatal death more important than the other components.

Greene and Solomon,³ in an editorial accompanying the *New England Journal of Medicine* publication of the ACHOIS trial,¹ cite the US Preventive Services Task Force summary of the evidence² to note that “only a fraction of deliveries complicated by shoulder dystocia result in birth trauma, and, in most cases, such trauma (clavicular and humeral fractures and brachial-plexus injuries) does not result in permanent injury”.

Thus, shoulder dystocia without birth trauma may be of less patient importance than the bone fractures and nerve palsies that result from birth trauma. In summary, we conclude there is a large gradient of patient importance across the components of the composite end point, with perinatal death as the most important and shoulder dystocia as the least important.

DID THE MORE AND LESS IMPORTANT OUTCOMES OCCUR WITH SIMILAR FREQUENCY?

The large gradient in importance between the components of the composite end point has alerted us to a potential problem. If the more important components occur less frequently than relatively unimportant components, our concern will rise further.

Five (0.95% of patients) perinatal deaths occurred in the control group (n = 524) and none in the intensive management group (n = 506) (RRR 100%, CI 20% to 100%); the corresponding figures for shoulder dystocia were 16 (3%) and 7 (1.4%) (RRR 55%, CI -6% to 81%), for bone fractures 1 (0.2%) and 0 (RRR 100%, CI -662% to 100%), and for nerve palsies 3 (0.6%) and 0 (RRR 100%, CI -30% to 100%). These data tell us that shoulder dystocia, the least important component of the composite end point, accounted for 77% of all events. The difference in events between treatment and control group (0.95% for perinatal death, 1.6% for shoulder dystocia, 0.2% for bone fractures, and 0.6% for nerve palsies) is somewhat similar, and the exclusive occurrence of death in the control group warrants notice. Yet again, however, half of the overall absolute reduction in risk of the composite comes from the reduction in the risk of shoulder dystocia.

The relative dominance of shoulder dystocia over the other components and the large gradient in patient importance across the component outcomes would suggest one should focus on the effect of the intervention on the individual component outcomes and dismiss the effect on the composite end point. There is, however, one more question to consider.

ARE THE COMPONENT OUTCOMES LIKELY TO HAVE SIMILAR RELATIVE RISK REDUCTIONS?

Similar risk reductions across the component outcomes would suggest that the investigators got the biology of the intervention right. In other words, similar effects across components support use of a composite. In this case, the intervention (including tight glycaemic control) should affect the causal pathways leading to perinatal death, shoulder dystocia, bone fractures, and nerve palsies in a similar way and to a similar extent. While the biological link between hyperglycaemia and shoulder dystocia and birth trauma is macrosomia (the intervention reduced the risk of macrosomia by 53%, CI 36% to 66%), how hyperglycaemia and macrosomia are linked to perinatal death remains unclear.

Strong inferences about uniformity of RRRs across individual outcomes come only from considering the point estimates and their confidence intervals. The reductions in risk of perinatal death, bone fractures, and nerve palsies were all 100%, while the RRR for shoulder dystocia was 55%. The paucity of events and resulting imprecision of the estimates of RRR for each of the component outcomes weakens any inferences about their similarity.

To review our answers to the 3 questions, parents will find perinatal death far more important than bone fractures and nerve palsy, and most will find shoulder dystocia less important than the other 3 components. All components occurred infrequently with shoulder dystocia occurring most frequently. The wide confidence intervals around the RRRs weaken any inference about their similarity, but there is no clear biological reason to expect these to be similar. These answers to our 3 questions suggest to us that the composite end point used in this trial is a suboptimal measure of the effect of this intervention (while appreciating the very considerable efforts that investigators went to document this effect!).

In considering how to apply this evidence in clinical practice, decision makers should therefore focus on the effects of the intervention on the components of the composite end point, particularly on the most important one: perinatal mortality. With only 5 perinatal deaths among 1030 births in the ACHOIS trial, is this enough evidence to justify a policy of screening for gestational diabetes and intensive treatment? Another clinical trial is ongoing in this area,³ and perhaps only a meta-analysis of these and other randomised trial evidence will offer sufficient data on individual component outcomes to draw confident inferences about the balance between a possible reduction in perinatal deaths and the currently much stronger evidence of an increase in induced labour and admission to a neonatal nursery.

CONCLUSIONS

When event rates are low, the use of composite end points in clinical trials allows investigators to reduce sample size and the duration of follow up. These advantages come at a price: the interpretation of the effect on the intervention is complicated, and the combined end point can be profoundly misleading. We hope *Evidence-Based Medicine* readers will find our questions helpful in deciding when to accept the effect of treatment on the composite as a valid measure of the effect of treatment and when to ignore the composite end point and focus on individual components.

Evidence-Based Medicine has recognised the importance of this issue. Henceforth, the journal will report the event rate for each component outcome when reporting on trials using composite end points. The welcome beginning to this policy is this issue's report of the ACHOIS trial.

VICTOR M MONTORI, MD, MSC
Mayo Clinic College of Medicine
Rochester, Minnesota, USA

JASON W BUSSES, DC, MSC
McMaster University
Hamilton, Ontario, Canada

GAIETÀ PERMANYER MIRALDA, MD
Hospital General Vall d'Hebron
Barcelona, Spain

IGNACIO FERREIRA, MD
Universitat Autònoma de Barcelona
Barcelona, Spain

GORDON H GUYATT, MD, MSC
McMaster University
Hamilton, Ontario, Canada

- 1 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.
- 2 Montori VM, Permanyer-Miralda G, Ferreira-Gonzalez I, *et al*. Validity of composite end points in clinical trials. *BMJ* 2005;**330**:594–6.
- 3 Greene MF, Solomon CG. Gestational diabetes mellitus—time to treat. *N Engl J Med* 2005;**352**:2544–6.
- 4 Brody SC, Harris R, Lohr K. Screening for gestational diabetes: a summary of the evidence for the U.S. Preventive Services Task Force. *Obstet Gynecol* 2003;**101**:380–92.
- 5 Landon MB, Thom E, Spong CY, *et al*. A planned randomized clinical trial of treatment for mild gestational diabetes mellitus. *J Matern Fetal Neonatal Med* 2002;**11**:226–31.