

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

In this issue, the *ACP Journal Club* highlights and comments on a landmark randomized study, the ACHOIS trial (1). In this trial, Crowther and 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 levels, 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 cesarean 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 labor, and 152 (31%) and 164 (32%), respectively, required cesarean 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 for adverse outcomes and that their risk approximates that of the average woman enrolled in the ACHOIS trial (women at identifiably higher risk for these outcomes would have smaller NNTs, and vice versa). While the trial evidence suggests no difference in the rate of cesarean 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 labor. Further, clinicians admitted 357 of 506 (71%) and 321 of 524 (61%) infants in the intervention and routine-care groups, respectively, to the neonatal nursery. 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 labor inductions, and the increase in neonatal nursery admissions, if the intervention reduced serious perinatal complications in the infants. 7 (1%) and 23 (4%) of the infants in the intervention and routine-care groups, respectively, had the composite endpoint, “serious perinatal complications,” a difference that was statistically significant ($P = 0.04$) and indicated that clinicians must screen and manage 34 (CI 19 to 98) women with mild gestational diabetes to prevent 1 serious perinatal complication.

Many would find the 3% absolute risk reduction (ARR) for the combined endpoint, “serious perinatal complications” (relative risk reduction [RRR] 68%, CI 29 to 86), a favorable tradeoff. But should they interpret the study results this way? Or should readers focus instead on the effects of treatment on each of the individual components of the composite endpoint?

We have recently suggested a series of questions that can help clinicians interpret studies by using composite endpoints (2). We will apply these questions in examining the validity of the composite endpoint “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 68% RRR or the 3% ARR in the composite endpoint 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 accompanying editorial to the ACHOIS trial (1), cite the U.S. Preventive Services Task Force summary of the evidence (4) 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 less important to patients than the bone fractures and nerve palsies that result from birth trauma. We conclude that there is a large gradient of patient importance across the components of the composite endpoint, 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 endpoint has alerted us to a potential problem. If the more important components occur less frequently than relatively unimportant components, our concern will rise further.

5 (0.95% of patients) perinatal deaths occurred in the routine-care 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 endpoint, accounted for 77% of all events. The difference in events between intensive-management and routine-care groups (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 routine-care group warrants notice. Yet again, however, half of the overall absolute reduction in risk for the composite endpoint comes from the reduction in the risk for 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 that one should focus on the effect of the intervention on the individual component outcomes and dismiss the effect on the composite endpoint. There is, however, one more question to consider.

ARE THE COMPONENT OUTCOMES LIKELY TO HAVE SIMILAR RRRs?

Similar RRRs 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 endpoint. In this case, the intervention (including tight glycemic 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 hyperglycemia and shoulder dystocia and birth trauma is macrosomia (the intervention reduced the risk for macrosomia by 53%, CI 36 to 66), how hyperglycemia 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 for 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 similarities.

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 that the composite endpoint 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 focus on the effects of the intervention on the components of the composite endpoint, 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 (5), and perhaps only a meta-analysis of these and other randomized controlled trials will offer sufficient data on individual out-

comes 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 labor and admission to a neonatal nursery.

When event rates are low, the use of composite endpoints 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 endpoint can be profoundly misleading. We hope *ACP Journal Club* readers will find our questions helpful in deciding when to accept the effect of treatment on the composite endpoint as a valid measure of the effect of treatment and when to ignore the composite endpoint and focus on individual components.

ACP Journal Club has recognized the importance of this issue. Henceforth, the journal will report the event rate for each component outcome when reporting on trials using composite endpoints. The welcome beginning to this policy is this issue's report of the ACHOIS trial.

*Victor M. Montori, MD, MSc
Rochester, Minnesota, USA*

*Jason W. Busse, DC, MSc
Hamilton, Ontario, Canada*

*Gaieta Permanyer-Miralda, MD
Barcelona, Spain*

*Ignacio Ferreira, MD
Barcelona, Spain*

*Gordon H. Guyatt, MD, MSc
Hamilton, Ontario, Canada*

References

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.