Interpreting Negative Results

Postpartum Interview Position Not Associated With Improved Outcomes

The randomized, blinded study by Valdes et al in this issue of the ARCHIVES examined mothers' satisfaction and their retention of information on common postnatal topics when the postpartum interview was conducted while the pediatrician was sitting in a chair near the bed, standing, or sitting on the mother's bed. The authors' study hypothesis was that postpartum mothers' satisfaction with pediatricians and their knowledge would improve when a seated pediatrician conducted postpartum visits. Seventy-five primiparous and multiparous women were enrolled as participants in the study on their first postpartum day in a university hospital newborn nursery. Mothers were randomly assigned to be interviewed with the pediatrician in 1 of the 3 positions. The primary study outcomes were the mother's perception of the duration of the pediatrician's interview, her satisfaction with the interview, the number of questions asked by the mother, and her retention of facts discussed during the interview. Differences among the 3 position groups were not statistically significant. Thus, the authors did not confirm their hypothesis that sitting near a postpartum mother improves her satisfaction with pediatricians or that different physician positions increase the retention of newly learned information.

See also page 327

We critiqued the study using the Users' Guides to the Medical Literature tool for therapeutic trials as a standard for addressing methodological issues. In this discussion, we evaluate the components of the clinical trial and discuss the unique issues surrounding a study that does not have statistically significant findings, often termed a negative-results study.

Randomization of Study Participants

Participants were randomized using computer-generated codes to 1 of 3 groups: chair, standing, or bed. Randomization was properly concealed by keeping position assignments in opaque envelopes, which were opened by the pediatrician immediately prior to each interview. Participants were appropriately analyzed in the groups to which they were randomized.

Randomization resulted in generally comparable groups that were similar with regard to type of delivery, duration of interview, opioid use, and amount of pain reported. None of the participants reported having a prepartment consultation with any other health care professionals. Age was the only variable that significantly differed among the 3 groups; participants in the bed group were older. However, other possible confounders that could affect results, such as socioeconomic status, level of social support, and timing of the interview relative to the time of birth, were not discussed. Although the educational level of the participants was briefly mentioned (one participant had a graduate degree), no further reporting of possible differences between participants with middle-school level vs high-school level education was included. Other factors that could influence the outcomes are important to consider. A mother who only completed the eighth grade may potentially be more influenced by social desirability than a mother who received a high school diploma. As a result, that mother might answer satisfaction questions in a more positive manner in order to please the outcomes assessor, resulting in biased data.

Blinded Assessment

Blinding is just as essential as randomization when designing clinical trials. Randomization removes the influence of confounding variables at baseline, whereas blinding removes the influence of confounding variables during the study and outcome assessment periods. An unblinded investigator may behave in a manner that influences study participants and results in biased outcomes or may introduce bias during data assessment. Participants who are aware of their group allocation may also introduce bias into the study. For example, a participant who is aware of her group allocation may overreport or underreport on a particular outcome being assessed.

The authors reported that this was a double-blinded study. Outcome assessors who collected the data were not aware of group allocation. The data collectors, who were blinded to the randomization of participants, asked a series of questions 15 to 60 minutes after each completed postpartum interview. However, we were not able to ascertain if patients were aware of the nature of the study or their group allocation. It is not clear whether the study hypothesis was stated to mothers, but they did have to provide informed consent.
to participate and were aware of the position of the interviewer. Most important, the 2 pediatricians who conducted the interviews were not blinded to the group allocation, and they were also the study investigators who knew the hypotheses. They attempted to account for interviewer bias by trying to keep the time of the interview and number of questions consistent. However, it is possible that the pediatricians might have unknowingly behaved differently in the various positions during the interviews. For example, a pediatrician who believed that sitting on the bed would make a significant difference might ask questions with a different intonation or change in body language, thus biasing the data to demonstrate positive results. Although the investigators were not blinded, the negative results of this study suggest that any investigator bias present did not significantly affect the results.

PARTICIPANT FOLLOW-UP

The authors fully account for all the subjects in the study. One participant assigned to the chair group fell asleep and was excluded from the study. Another participant, assigned to the standing group, was excluded after being interviewed by a physician not involved in the study. Follow-up was completed for all study participants. Outcomes were assessed within 1 hour of the intervention. Given the short duration of follow-up, it is possible that a clinically significant effect might have become evident after a longer follow-up period. For instance, if knowledge of postnatal topics had been assessed 2 weeks later it is possible that improved retention of information learned might have been demonstrated vs a mere repetition of facts.

VALIDITY OF RESULTS

In this study, the authors do not report statistically significant treatment effects for their identified outcomes. Thus, they reject the hypothesis that a physician’s sitting position improves the mother’s satisfaction with the postpartum interview or her ability to learn new information.

Negative results must be considered using criteria similar to those used to judge positive results. Are these results valid? Two key concepts to consider before accepting that an intervention is ineffective are power and sample size. Power is the probability of detecting an effect or association when there is an effect or association and is used to calculate the sample size needed to detect clinically significant differences between groups. Many reviewers and readers are familiar with type I, or \( \alpha \), error, in which a significant difference is reported when there is none. Without a sufficiently large sample, studies are also at risk for a type II, or \( \beta \), error—failing to reject the null hypothesis when there is a difference not attributable to chance alone. Type II error is important because a researcher does not want to prematurely dismiss a hypothesis as invalid if it has not been fully tested in a study with a sufficient sample size.

In evaluating a study, it is important to determine if it is powered adequately. The significance level, \( \alpha \), was set at .05, and a power of 80% was chosen. One-way analysis of variance was used to detect differences among the 3 groups.

The study was powered to detect a minimum difference of 2 minutes in the estimated encounter duration. In their analysis, the authors determined that they needed 20 patients per group to detect this difference. To account for possible dropouts, they recruited 25 patients per group.

The authors neglected to provide the sample means and standard deviations of their results, which would allow one to determine if their study was powered sufficiently. Instead, they provided median values and interquartile ranges. We know that with increasing variance a larger sample size is needed. If, for example, the sample standard deviation were 6 instead of the predicted 2, the sample size needed to maintain a power of 80% would more than double. In this instance, the current sample size would be insufficiently powered to detect the difference among the 3 groups. The authors, in this circumstance, would fail to reject the null hypothesis when it was false.

In a negative-results study that tested multiple outcomes, one must also consider validity when only a single outcome was reported in the determination of sample size. The authors based their sample size on estimated encounter duration. This is despite later saying that there is considerable evidence that satisfaction with an encounter, another of their main outcomes, poorly correlates with perceived encounter duration. It is important to determine the sample size based on the most important outcome of the study. Otherwise, one may run the risk of being unable to confidently answer the research question, despite using rigorous methods.

CONCLUSIONS

This study is interesting for several reasons. One, it provides a unique example of a negative-results trial. Such studies are less commonly accepted for presentation and publication due to a bias toward studies that report significant associations. Publication bias toward positive results is important because therapies and interventions may be considered clinically significant rather than being recognized as chance results.

Second, this study allows readers to think about how important it is to critically assess studies with negative results. Negative-results trials have been demonstrated to often lack the statistical power needed to determine that a clinically relevant difference is present. Thus, it is vital for all types of studies to have adequate sample sizes to provide enough power to detect differences and to measure appropriate outcome variables.

Lastly, this study examines how a simple clinical intervention that is accessible to all might affect patient satisfaction and learning. With no real potential harms or costs, the treatment benefits of such an intervention would be worthwhile. Although the authors did not find significant differences among the 3 groups, until an additional study with a larger sample size is performed, we
must continue to study methods of enhancing physician-patient communication.

Iris R. Mabry, MD, MPH
Tracy Richmond, MD
Adriana Bialostozky, MD
Jerry Rushton, MD, MPH
Ann Arbor, Mich

Corresponding author: Iris R. Mabry, MD, MPH, Division of General Pediatrics, University of Michigan Health System, 6D04 NIB, 300 N Ingalls St, Ann Arbor, MI 48109-0456 (e-mail: imabry@med.umich.edu).

REFERENCES

2. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature, II: how to use an article about therapy and prevention, A: are the results of the study valid? JAMA. 1993;270:2598-2601.
3. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature, II: how to use an article about therapy and prevention, B: what were the results, and will they help me in caring for my patients? JAMA. 1994;271:59-63.