The preimplantation genetic testing debate continues: first the hype, then the tension, now the hypertension?

The study by Zhang and colleagues (1) in this month’s Fertility and Sterility asks an important question: Is preimplantation genetic testing (PGT) associated with an increased risk of obstetric complications? Given the physiologic role of the trophectoderm in implantation, invasion, and placentation (2), it seems biologically plausible that manipulation of these cells through biopsy sampling could result in an increased risk of aberrant placentation. The rapid increase in the clinical use of PGT further justifies more scientific inquiry into this question. Framing the clinical and research implications of this study requires several considerations: [1] the pros and cons of interpreting cohort studies, [2] the interpretation of small sample sizes that result in only one end point being statistically significant, and [3] the clinical context of numbers needed to harm coupled with the indication for PGT in the first place.

We all would prefer randomized controlled trials (RCTs) for every clinical question, but they are not financially feasible for every clinical scenario patients and providers encounter. Many important questions still need to be initially explored by cohort studies to assess whether an association may exist. However, cohort studies are limited in that they can only assess for associations and not causality. The authors have highlighted this fact in their conclusion and appropriately commented that their data warrant further investigation.

It is difficult to demonstrate that PGT causes preeclampsia with these data. A notable reason for this is the challenge of controlling for confounding variables in cohort studies. Patients who are clinically treated with PGT are different than the patients who undergo in vitro fertilization (IVF) without PGT. Researchers do their best to identify and control for these confounders, but the potential for unrecognized confounding bias always exists in cohort studies.

As just one example, RCT data and a large meta-analysis have demonstrated that patients who have frozen embryo transfers (FET) have an increased risk of hypertensive disorders, preeclampsia, and large-for-gestational-age offspring (3, 4). In the cohort study of Zhang et al. (1), 76% of PGT cycles were FET versus 69% of the controls ($P = .18$, chi-square test). In this case, a $P$ value is not designed to inform us as to whether the prevalence of FET increased the risk of preeclampsia in PGT cycles. The use of FET was not accounted for in the modeling of the primary outcome. A separate analysis was performed of only FET cycles, but it failed to show a statistically significant difference in preeclampsia between the IVF cycles with and without PGT. Furthermore, the 95% confidence interval (CI) of preeclampsia (including fresh transfers and FET) approached the null (95% CI lower limit, 1.10; $P = .02$), so even one or two additional cases of preeclampsia resulting from FET itself but not PGT could have changed the conclusions of this study. The authors employed extensive accounting for other variables, but both recognized and unrecognized confounding can have a substantial impact on cohort studies.

Small sample sizes and a single statistically significant finding must also be interpreted through a critical lens. Preeclampsia was the only statistically significant finding in this study ($P = .02$); gestational hypertension was not statistically significant ($P = 1.0$). If these two disorders are on the same clinical spectrum, one might anticipate a similar result in both findings. Can PGT result in increased preeclampsia but have no effect on gestational hypertension? Or is this the result of chance? Indeed, none of the other 22 outcomes were statistically significantly associated with PGT, and 18 of 22 of those outcomes had $P = 1.0$ as their comparison. If trophectoderm biopsy resulted in aberrant placentation, which resulted in an increased risk of adverse obstetric outcomes, one might expect to see this manifest across a spectrum of adverse outcomes. This was not the case in the present study.

Small studies are also at increased risk of type I error due to small perturbations in event counts. As an example, if a fair coin is flipped 10 times, observing that heads appear by chance 7 times or more happens about 1 in 6 times. If the same coin is flipped 1,000 times, the same relative finding (700 heads) will almost never occur because it happens at an extremely low frequency (less than 1 in a trillion). A difference of one or two cases of preeclampsia in this study would have changed the conclusion due to the high sensitivity of the statistics with such a small sample size.

Finally, the authors did not account for multiple comparisons when evaluating preeclampsia. This is a reasonable decision when there is a prespecified primary outcome in a trial. The reader is left to consider whether the finding of increased preeclampsia in PGT patients is due to chance, confounding factors, or both. The authors were honest about this potential, pointing to the very wide estimate of the effect size which suggests imprecision.

The clinical relevance of an effect cannot be determined by $P$ values or odds ratios. These are statistical constructs that inform us about the probability of the finding and the relative estimation of the effect. Clinical relevance requires consideration of the absolute risk and the numbers needed to treat or harm. If the findings of this study are true after further scientific interrogation, then the absolute increased risk of preeclampsia is 6.4% in women who have PGT, and the number of patients needed to harm is 16 with PGT to have one additional case of preeclampsia.

How would patients and providers interpret this risk? On the one hand, only 1 of 16 patients who get PGT will experience preeclampsia as a result of that decision. Further, even if this increased risk exists, there was no difference in neonatal intensive care admissions or overall neonatal outcomes. These risks might be interpreted very differently based on the indication for PGT. For a patient requiring PGT-M (monogenic/single gene defects) or PGT-SR (chromosomal structural rearrangements), these risks might be irrelevant in the larger picture of why PGT is chosen. For patients electing for PGT to reduce their miscarriage risk and time to pregnancy, these
risks might be interpreted with great variability from patient to patient. Given the provocative discussion over PGT among clinicians, these risks would likely be variably interpreted and argued to support positions already well entrenched in debate.

The study by Zhang and colleagues (1) is an important first step into asking the biologically plausible question of whether trophectoderm biopsy might lead to aberrant placentation, resulting in adverse obstetric outcomes. We have outlined the limitations of the present data, but there is value in what the authors address. We agree with the author’s conclusion that these data warrant further investigation, and we encourage past and future investigators of PGT to report subsequent obstetric outcomes from RCTs. We caution readers to not definitively conclude that PGT causes increased risk of preeclampsia based on this study alone.

George Patounakis, M.D., Ph.D. \textsuperscript{a,b}
Micah J. Hill, D.O. \textsuperscript{c}
\textsuperscript{a} Reproductive Medicine Associates of Florida, IVI-RMA Global, Lake Mary, Florida; \textsuperscript{b} Department of Obstetrics and Gynecology, College of Medicine, University of Central Florida, Orlando, Florida; \textsuperscript{c} Program in Reproductive Endocrinology and Infertility, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland

https://doi.org/10.1016/j.fertnstert.2019.04.028

You can discuss this article with its authors and other readers at https://www.fertstertdialog.com/users/16110-fertility-and-sterility/posts/48248-28143

REFERENCES