Clinical and Practical Psychopharmacology

Intellectual Disability After Gestational Exposure to Antidepressant Drugs:
The Confidence Interval as a Compatibility Interval

Chittaranjan Andrade, MD

Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

ABSTRACT

There is growing concern among statisticians that the use of *P* value dichotomization into significant and nonsignificant outcomes does not appropriately describe research findings; in fact, misconceptions about the very meaning of the *P* value abound. Many alternate approaches to interpreting data have been suggested. The use of the 95% confidence interval (CI) as a compatibility interval is one such approach. By this approach, the study estimate for an outcome is considered to be the most compatible with the population value for that outcome, and all the values in the 95% CI around that estimate are also considered to be compatible with the population value with decreasing likelihood of compatibility the greater the distance of the value from the study estimate. This concept is explained with the help of a study that examined the risk of intellectual disability (ID) in children who had been gestationally exposed to antidepressant drugs (ADs). A conventional interpretation of the study findings is that, after adjustment for confounders, AD exposure was not associated with an increased risk of ID; this, in fact, is also how the authors expressed their conclusions. However, when the 95% CI in the main analyses, subgroup analyses, and sensitivity analyses in this study are examined as compatibility intervals, it becomes apparent that, even after adjustment for confounders, a very sizeable range of values indicating increased risk is compatible with the population value. In contrast, the range of values indicating decreased or no risk is very small. The implication, therefore, is that adjustment for confounding attenuates the risk, but the risk probably remains elevated. There is uncertainty in this subjective conclusion, but this conclusion is more truthful than a conclusion that a significant risk was rendered nonsignificant by adjustment for confounding.

J Clin Psychiatry 2019;80(3):19f12912

Statistical procedures are applied to research data to test hypotheses. These procedures usually end with a *P* value. If the *P* value is less than .05, the research finding is considered statistically significant; this is usually a cause for celebration because it suggests that something important has been discovered or confirmed and that the finding is worthy of a scientific publication.

However, there are several elephants in the room. Examining the *P* value in relation to a .05 threshold conveys little information about the study finding beyond whether or not it meets a definition for statistical significance; in contrast, 95% confidence intervals (CIs) and measures of effect size allow a far better understanding of the data.[1] The value .05 for statistical significance is not carved in stone; it may need to be corrected for false-positive errors.[2] There is disagreement about what the primary *P* value must be to be considered statistically significant; values as low as .005 have been suggested in lieu of .05.[3] The *P* value is a continuous value that lies between 0 and 1, and all intervening values are meaningful; in contrast, setting a .05 threshold for statistical significance creates an artificial dichotomy that results in a considerable loss of information.[4] The very concept of statistical significance implies that findings that are nonsignificant have no value, actionable or otherwise; this is completely false.[5] Finally, for what it is worth, very few students, clinicians, and even researchers know what the *P* value actually means; many misconceptions about the *P* value have been described.[6]

There is an increasing recognition that statistical inferences based on the *P* value and the associated concept of statistical significance need to be replaced by better ways of examining data and drawing conclusions therefrom.[7,8] One of many suggestions has been to reconceptualize the 95% CI as a compatibility interval.[5] This suggestion is explained in the context of a recent article[9] that examined the risk of intellectual disability (ID) in children whose mothers had used antidepressant medication during pregnancy.

Adverse Outcomes After Gestational Exposure to Antidepressant Drugs

There are probably more data available on pregnancy, neonatal, childhood, and even adult outcomes after gestational exposure to antidepressant drugs than to any other class of drugs in the pharmacopeia. Adverse outcomes that have been associated with gestational exposure to antidepressants include spontaneous abortion, premature delivery, small-for-dates and low

Chittaranjan Andrade

birth weight, morphological teratogenicity, neonatal distress or complications requiring neonatal intensive care admission, persistent pulmonary hypertension of the newborn, autism spectrum disorder in childhood, attention-deficit/hyperactivity disorder in childhood, and psychiatric disorders in adult life. Most of the research data suggest that these adverse outcomes are associated with gestational antidepressant exposure through confounding by indication. That is, the socioeconomic environment of depression, the dysfunctional behaviors associated with the depressed state, and the neurobiology of depression (that is severe enough to necessitate antidepressant use during pregnancy) may be responsible for the risk of adverse outcomes rather than the antidepressant exposure, itself. Thus, antidepressant exposure is a (proxy) marker of risk rather than the cause of the risk.

Intellectual Disability After Gestational Exposure to Antidepressant Drugs

ID is a little-studied outcome following gestational exposure to antidepressant drugs. In this context, Viktorin et al[9] conducted a population-based cohort study of children, identified through different Swedish registers, to determine whether antidepressant use during pregnancy is associated with an increased risk of ID in offspring.

The sample comprised 179,007 children studied at a mean age of 7.9 years. The sample was 51.5% male. ID had been diagnosed in 0.9% (n = 37) of 3,982 antidepressant-exposed children and in 0.5% (n = 819) of 172,646 unexposed children.

Exposed and unexposed children differed in several regards. For example, mothers and fathers were more likely to have a psychiatric diagnosis in the exposed group than in the unexposed group (72.0% and 21.0% vs 13.6% and 10.5%, respectively).

In the unadjusted analysis of data from the full sample, antidepressant use during pregnancy was associated with an almost doubled risk of ID in the offspring (relative risk [RR], 1.97; 95% CI, 1.42–2.74). In the fully adjusted model, which adjusted for all potential confounders for which data had been acquired, the risk was considerably attenuated (RR, 1.33; 95% CI, 0.90–1.98). The confounders that were adjusted for included birth date, maternal and paternal age, paternal psychotropic medication use that overlapped the pregnancy, maternal and paternal education, and maternal and paternal psychiatric diagnoses.

In a subgroup analysis of children of women who had a prenatal diagnosis of anxiety or depression, 2,372 exposed children were compared with 4,976 unexposed children. The RR for ID in the exposed children was 1.57 (95% CI, 0.96–2.59) in the unadjusted model and 1.64 (95% CI, 0.95–2.83) in the fully adjusted model.

A similar pattern of findings was identified in almost all the other subgroup and sensitivity analyses; that is, there was a progressive attenuation of risk from the unadjusted analysis to the fully adjusted analysis. These subgroup and sensitivity analyses included analyses conducted separately for selective serotonin reuptake inhibitors (SSRIs), non-SSRI antidepressants, and nonantidepressant psychotropics; analyses conducted in women who had received only 1 prescription for an antidepressant during pregnancy; analyses conducted separately for different categories of antidepressant in children with mild to moderate ID, children with severe ID, and children with ID without autism spectrum disorder; analyses conducted in males and females separately; and others.

Conventional Interpretation of the Study[9] Findings

The RR is calculated as the risk in the group of interest divided by the risk in the comparison group. If the risk in these two groups is identical, the RR is 1.00. If the 95% CI for the RR includes the value 1.00, it means that the RR is not statistically significant when $P < .05$ is set as the threshold for significance.[1,10]

In the Viktorin et al[9] study, the unadjusted RR for the full sample was 1.97 (95% CI, 1.42–2.74). This is a statistically significant increase in risk. However, the same RR attenuated to 1.33 (95% CI, 0.90–1.98) in the fully adjusted analysis, that is, the analyses that took into account the effects of all the potential confounders. The 95% CI of the RR in the fully adjusted analysis includes the value 1.00 (because 1.00 lies between 0.90 and 1.98). Therefore, even though the RR value of 1.33 represents a 33% increase in risk, the RR is interpreted as being not statistically significant.

So the conventional interpretation of the data is that gestational exposure to antidepressants is associated with an increased risk of ID; however, the risk is no longer statistically significant after adjusting for confounders. In the subgroup of women who had a prenatal diagnosis of anxiety or depression, the 95% CI for the RR included 1.00 in unadjusted as well as adjusted analyses. Therefore, these RRs are also not statistically significant. Similar conclusions can be drawn from the subgroup and sensitivity analyses.

Therefore, it would seem that the risk of ID is explained by the confounders and not by the antidepressant exposure. The conclusion would seem to be that this study does not support the hypothesis that gestational antidepressant exposure is associated with an increased ID risk.

This study is an excellent example of why the use of statistical significance should give way to better ways of interpreting research findings. Why this is so is explained in later sections of this paper.

How the Authors[9] Interpreted Their Findings

In the Statistical Analysis section of their paper, Viktorin et al[9] stated that "all tests of statistical hypotheses were performed on the 2-sided 5% level of significance." That is, they set .05 as the threshold for declaring results as statistically significant. However, they did not state P values for any of the findings in either the published paper or the supplementary materials. In a way, this is to their credit because they broke free from the shackles of drawing conclusions from P values. In a way, this is also a discredit because P values are not useless numbers; they do convey meaningful information when P is considered as a continuous variable.[4]

In their discussion, the authors observed that with increased adjustment for confounding, the association between antidepressant exposure and ID "gradually attenuated to a statistically nonsignificant RR." They did not state "significant" or "nonsignificant" to describe their results anywhere else in their published paper and the supplementary materials. This avoidance of dichotomous interpretations is also creditable.

In their discussion, the authors correctly stated that "the association between offspring ID and maternal antidepressant medication may, to a large extent, be explained by confounding by the covariates included and adjusted for in the current analyses." However, they did make one substantially erroneous statement, and in a prominent place, to boot. In the Conclusions subsection of their abstract, they stated "After adjustment for confounding factors, however, the current study did not find evidence of an association between ID and maternal antidepressant medication use during pregnancy." This statement is completely wrong, as will be explained in later sections.

Confidence Intervals as Compatibility Intervals

The findings of a study describe only the sample that was studied. However, researchers and particularly readers are interested not just in the sample but in the population at large. If the sample is representative of the population, the findings from the study can reasonably be generalized to the population. However, samples are seldom truly representative of the population. This is most commonly so because samples are often convenience samples that are further restricted by study-specific inclusion and exclusion criteria. Thus, generalization from the sample to the population can be problematic.

No matter how good the sampling, findings from a sample are only approximations of what characterizes the population, and it is the 95% CIs that help us understand where the population value of a finding may lie.[1] Viktorin et al[9] found that, in the fully adjusted model, gestational exposure to antidepressants was associated with a 33% increased risk of ID (RR, 1.33; 95% CI, 0.90–1.98). This means that there is a 95% chance that the population value for the adjusted RR lies between 0.90 and 1.98. Expressed otherwise, there is a 95% chance that the confounder-adjusted risk for antidepressant exposure–associated ID may be reduced by as much as 10% to increased by as much as 98%.

Interpreting study results using a threshold for statistical significance provides an artificial element of certainty to the conclusions drawn. This certainty is false because findings from a single sample cannot define the population values and because replicatory studies will almost certainly yield different estimates. In this context, Amrhein et al[5] suggested that researchers shun the dichotomy of statistical significance and, instead, embrace uncertainty by regarding the 95% CI as compatibility intervals.

How is this done? Using the same example examined earlier in this section, we would consider that all values

between 0.90 and 1.98 are compatible with the population value for the adjusted RR; the estimated RR, 1.33, is the most compatible value, and values within the CI are progressively less compatible (but nevertheless still compatible) the further away they lie from 1.33. The concept of statistical significance does not come into the picture at all.

Reinterpretation of the Study[9] Findings

As already stated, the conventional interpretation of the study[9] findings is that gestational exposure to antidepressants is associated with a significantly increased risk of ID and that the risk is no longer statistically significant after adjusting for confounders. The implication is that the risk of exposure-related ID arises from the confounders and not from the antidepressant exposure, itself. In short, the study suggests that gestational antidepressant exposure is not associated with an increased ID risk. The authors stated as much in their abstract: that their study "did not find evidence of an association between ID and maternal antidepressant medication use during pregnancy."

When the reader rejects the dichotomous interpretation and examines the 95% CI as compatibility intervals, it will immediately become apparent that, *for most if not all the analyses in the study, RR values that indicate an increased risk are most compatible with the population risk*. With regard to the full sample, fully adjusted analysis, for example, the 95% compatibility interval is 0.90 to 1.98. That is, the risk (in the population) may be decreased by up to 10% or increased by as much as 98%. Because most of the values in the compatibility interval suggest an increased risk, one cannot easily dismiss the possibility that, even after adjusting for all measured confounders, gestational antidepressant exposure remains associated with an increased risk of ID. This also applies to the CIs in almost all the other analyses in the study; that is, the findings are consistent.

This interpretation, based on a consideration of the compatibility intervals, indicates the likelihood of an increased risk in the population and is very different from the conventional interpretation of the findings and the interpretation stated by the authors in their abstract.

A Parting Note

The reinterpretation does not imply that, to reduce the risk of ID in offspring, women must avoid antidepressant drug use during pregnancy. This is because the analyses in the study[9] adjusted for only some confounders, and only to the extent that these confounders were accurately measured. The authors[9] listed many confounders that they could not adjust for. Furthermore, it would not have been possible for them to adjust for unknown confounders, such as genetic risks. So the best take-home message is that antidepressant drug use during pregnancy is associated with an increased risk of ID in the offspring; the risk remains elevated but is substantially attenuated after adjusting for measured confounders; how unmeasured and unknown confounders further affect the risk is unknown.

Chittaranjan Andrade

REFERENCES

1. Andrade C. A primer on confidence intervals in psychopharmacology. *J Clin Psychiatry*. 2015;76(2):e228–e231.
2. Andrade C. Multiple testing and protection against a type 1 (false positive) error using the Bonferroni and Hochberg corrections. *Indian J Psychol Med*. 2019;41(1):99–100.
3. Ioannidis JPA. The proposal to lower *P* value thresholds to .005. *JAMA*. 2018;319(14):1429–1430.
4. Andrade C. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian J Psychol Med*. In press.
5. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305–307.
6. Goodman S. A dirty dozen: twelve *P*-value misconceptions. *Semin Hematol*. 2008;45(3):135–140.
7. Wasserstein RL, Lazar NA. The ASA's statement on *P* values: context, process, and purpose. *Am Stat*. 2016;70(2):129–133.
8. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond *P*<0.05. *Am Stat*. 2019;73(suppl 1):1–19.
9. Viktorin A, Uher R, Kolevzon A, et al. Association of antidepressant medication use during pregnancy with intellectual disability in offspring. *JAMA Psychiatry*. 2017;74(10):1031–1038.
10. Andrade C. Understanding relative risk, odds ratio, and related terms: as simple as it can get. *J Clin Psychiatry*. 2015;76(7):e857–e861.