Discussion: Prenatal exposure to anti-D immune globulin and autism risk by Croen et al

In the roundtable that follows, clinicians discuss a study published in this issue of the Journal in light of its methodology, relevance to practice, and implications for future research. Article discussed:

Croen LA, Matevia M, Yoshida CK, Grether JK. Maternal Rh D status, anti-D immune globulin exposure during pregnancy, and risk of autism spectrum disorders. Am J Obstet Gynecol 2008;199:234.e1-234.e6.

DISCUSSION QUESTIONS

- What are the advantages and disadvantages of the study design?
- What alternative study designs could have been used?
- How were the exposure and outcome variables defined?
- Were the methods for identifying exposures and outcomes adequate?
- How would you describe the patient population?
- In identifying cases, might bias have been introduced?
- Did the study have adequate power to address the study question?
- Were the analytic methods appropriate for the study design?

From the Washington University in St Louis, School of Medicine, St Louis, MO

Moderator
Emily DeFranco, DO
Third-Year Fellow, Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine

Discussants
Gilad Gross, MD
Associate Professor, Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine

Anthony Shanks, MD
First-Year Fellow, Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine

Traci Johnson, MD
Third-Year Resident, Department of Obstetrics and Gynecology

Tammy Shen
Fourth-Year Medical Student, Washington University in St Louis, School of Medicine

0002-9378/$34.00

doi: 10.1016/j.ajog.2008.07.026

INTRODUCTION

Over past decades, the prevalence of autism spectrum disorders (ASDs) has climbed from 4-5 cases per 10,000 children to more than 10 per 10,000; a current estimate puts the US prevalence at 1 in 166.[1,2] Because a temporal relationship exists between the rising rate of ASDs and an increase in the recommended number of childhood vaccines, researchers have wondered whether exposure to certain vaccines or vaccine constituents is to blame. Whereas no connection has been demonstrated in scientific studies, media attention has continued to escalate. This month, Journal Club members discussed a study that sought a link between thimerosal-containing anti-D immune globulin (RhIg) and ASDs.

Emily DeFranco, DO
and George A. Macones, MD, MSCE

BACKGROUND

DeFranco: The uncertain etiology of autism and related disorders has led to a great deal of speculation about possible associations between prenatal and early childhood exposures and later development of ASDs. Prior studies attempting to examine etiologic influences had been hampered by small sample sizes and study-design constraints. The authors of the paper we are discussing today looked for a relationship between exposure to RhIg and risk for autism. To accomplish this, they used a large health care database and a study design that could optimize the evaluation of rare outcomes and factors associated with those outcomes.

The researchers said they found no significant association between Rh-negative status, prenatal RhIg exposure, and ASDs. In our discussion today, we will consider a number of methodologic issues related to the study design used in this paper and comment on how these issues could have influenced the results; how the authors addressed the study's limitations; and whether we agree that the findings are valid and generalizable.

STUDY DESIGN

DeFranco: ***What are some of the advantages and disadvantages of the study design?***

Johnson: A case-control design was used in this study. One of the advantages of this design is efficiency in obtaining cases and controls to power the study. When investigating the relationship between prenatal treatment with RhIg—in this study, RhoGAM (Ortho-Clinical Diagnostics, Inc, Raritan, NJ)—and autism risk, one would assume that the frequency of exposure would be low, given the prevalence of Rh negativity and incidence of autism in the general population. If, for example, a prospective cohort design were to be used, researchers would encounter a problem during data collection; the investigator would have to follow up many women who received RhIg to detect each case of autism in a child. In a case-control study, researchers assemble cases by first finding the outcome of interest in medical records; they then go back through the records to determine whether any relevant exposures occurred. In this study, they identified children with at least 1 diagnosis of an ASD and then established whether their mothers had been treated with RhIg by examining prenatal medical

records. Using a case-control design also allows one to choose an appropriate control group that limits selection bias, thereby increasing validity of the comparison. Disadvantages of the design include its retrospective assessment of risk. The reliance on medical records to adequately determine risk is not ideal.

DeFranco: ***What other study designs could have been used to address this same question?***

Johnson: As noted, a prospective cohort study might have been used, in which patients who received RhIg are followed up for 8-10 years. At the same time, a cohort of patients who did not receive RhIg is tracked. Data are collected over time to ascertain the incidence of ASDs among the children of these populations, so a comparison of ASD rates can be made between the 2 groups. This design, however, would be time consuming, labor intensive, and expensive. Many subjects might be lost to follow-up as well.

DeFranco: ***How were the exposure and outcome variables defined?***

Johnson: Infants diagnosed with an ASD (cases), the outcome of interest, and infants without an ASD diagnosis (controls) were identified from a cohort of infants who were born at a Kaiser Permanente of Northern California (KPNC) facility between January 1995 and December 1999. The outpatient databases were electronically scanned for children with at least 1 diagnosis of an ASD, including autism, Asperger's disorder, or pervasive developmental disorder not otherwise specified (PDD-NOS). One randomly sampled control per case was selected from this database for comparison. Information on maternal Rh status and thimerosal-containing RhIg exposure was abstracted from prenatal medical records. The influenza vaccine also contains thimerosal; thus, information on its receipt was also gathered. Data on maternal characteristics, infant characteristics, and number of exposures were recorded and compared.

DeFranco: ***Were the methods for identifying exposures and outcomes adequate?***

Shen: The methods for identifying exposures and outcomes were adequate overall. The exposures (maternal Rh status, RhIg administration, and receipt of influenza vaccine during pregnancy) were identified by abstracting prenatal medical records using a standardized form. This method avoided differential-recall causes of information bias, which might otherwise have been substantial, considering that the media has sensationalized the putative association between mercury exposure and autism. However, the study made no mention of whether those abstracting the medical records were blinded to the case or control status of the participants; if not, that could lead to potential information bias. Missing information in medical records might also provide another source of information bias. For example, among the 9 Rh-negative women in the study who did not receive RhIg, 7 had an Rh-negative partner. It is possible that the remaining 2 women did receive RhIg in the prenatal period but their immunization status was undocumented.

Nevertheless, several exposures were somewhat externally validated. The frequency of Rh-negative status was similar to published rates by race/ethnicity. The mean gestational age at the first prenatal injection was 27 weeks (SD 6 weeks), in accordance with the American College of Obstetricians and Gynecologists' recommendations.

As mentioned, the outcomes were identified as children with at least 1 diagnosis of ASD, including autism and Asperger's disorder, and diagnoses were identified by electronically scanning the KPNC outpatient clinical databases, which contained all diagnoses made in outpatient visits occurring at plan facilities and outside approved facilities. This method was comprehensive in identification of cases within the study population and reduced human error.

DeFranco: I agree. When considering the internal validity of a study, it is important to identify which study design was chosen; whether it was an adequate method to address the study's hypothesis; and how the design limitations were minimized by the methods in which the study was conducted.

DeFranco: ***Who was included in the patient population studied?***

Shen: The patient population studied included members of KPNC, a group model, integrated health plan that provides care for more than 3.2 million Northern California residents. Cases and control patients were born at a KPNC facility between January 1995 and December 1999; remained KPNC members for at least 2 years following birth; and were aged 4-7 years at the time the databases were scanned.

DeFranco: ***Was this a population-based sample?***

Shen: The sample used in the study should not be considered population based. The study population was identified from KPNC membership, which represents approximately 30% of the insured population in the Northern California region. Therefore, uninsured patients were not included in the study, which may be a potential source of selection bias. Even within the counties served by KPNC, the very poor and very wealthy were underrepresented.

DeFranco: It is not uncommon for study populations to be reported as population based. When critically evaluating the internal validity of a study, it is important to consider who was included in the study population and how those who were not included could have differed either in exposure or outcome status. The authors of this study used a large cohort of mothers and their offspring, with a breadth of information regarding demographic, obstetric, medical, and follow-up variables, but this was limited to the population who received medical care through KPNC hospitals. We should keep in mind that women who might have received prenatal care through other centers in the same area or women who did not receive prenatal care at all may have differed from the population studied in this analysis.

DeFranco: ***In identifying cases, might bias have been introduced into the study?***

Shanks: Case-control studies can be an efficient and powerful study design—especially when the incidence of the outcome being studied is rare. However, the reliability of the results lies in the ability to adequately identify appropriate cases and controls.

In this particular study, cases were defined as children with at least 1 diagnosis of an ASD, including autism and Asperger's disorder or PDD-NOS. Cases were identified by scanning a database of outpatient facilities affiliated with KPNC when the patients were 4-7 years of age. Although the actual criteria for obtaining the diagnosis of ASD were not mentioned in the study, the authors noted that more than 90% of children with an ASD diagnosis on their KPNC medical records met the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* criteria for autism. Use of *International Classification of Diseases, Ninth Revision* codes should allow for standardized measurement of cases.

It is worthwhile to mention that the population being assessed may not be a true reflection of patients at risk. Whereas KPNC may be an accurate reflection of the insured population, it is notable that the very poor and very wealthy are underrepresented. Excluding certain populations might miss potentially important associations. The methods state that the diagnoses were made during outpatient visits that occurred at plan and outside approved facilities. It would be important for all potential cases to have access to these resources.

DeFranco: ***Were the methods used to identify controls adequately described?***

Shanks: Controls were matched for sex, birth year, and hospital of birth. Matching for age is important because the diagnosis of ASDs was made in children between 4 and 7 years of age. This is notable because symptoms of autism are commonly recognized by age 2 years, whereas symptoms of Asperger's disorder might not be apparent until later (4-6 years). Selecting controls at too early an age might miss potential cases. Also, matching controls by the hospital of birth implies that both controls and cases had the same access to health care.

The authors acknowledged that there were 13 case mothers and 5 control mothers who each had 2 children in the original study cohort. The authors chose to sample only 1 child from these mothers to ensure independence of observations pertaining to maternal characteristics. Mothers of children with an ASD tended to be older, and they had significantly more education; a higher percentage of them had college and postgraduate education.

By not carefully matching controls to cases, confounding could enter the study. Control selection is vital in case-control studies. The methods state that controls were selected randomly from a cohort of births without an ASD disorder. Initially, 5 controls were selected per case, but this was altered to 1 randomly sampled control per case. In studies with small numbers of cases, increasing the number of controls can improve the study's power.

It is worthwhile to mention that controls were selected at random, although details regarding this process were not provided. The authors may have done an appropriate job ensuring a random selection process, but ultimately it is up to readers to decide whether that is so and to determine the study's applicability to clinical practice. Given the importance of control selection in a study of this design, more detail about the selection process would have been beneficial.

STATISTICAL ANALYSES

DeFranco: ***Did this study have adequate power to address the study question?***

Gross: The study's research question is essentially whether an association exists between the exposure—RhIg—and the outcome, in this case autism. Although not clearly stated by the authors, the null hypothesis is taken to be that RhIg exposure is not associated with an increased risk for autism. The same could be said for the association of other covariates, such as Rh-negative status itself.

The power of a study represents the probability that a statistical test will reject a false null hypothesis, meaning, a type II error will not be made. Type II errors essentially are false-negative findings, also known as beta errors. To determine power, several factors are taken into account. Among these is the level of significance (the typical arbitrary choice is 0.05); the power to detect an effect (typical arbitrary choice, 0.80); the effect size; and variation in the response variables. Sample size is also taken into consideration.

In this study, sample size is considered to be fixed because all outcomes (autism) have already occurred. Hence, the outcome sample size is fixed, which means that the power analysis has to be done after data collection. This is called a post hoc analysis. The authors were looking to detect a minimum odds ratio of 1.9, given a background exposure prevalence of 10%. This means that, given that 10% of the population is known to be Rh negative, and it is assumed that close to 100% of patients that are Rh negative will receive the intervention (RhIg), there would be nearly double the amount of autistic children in the exposure group vs the nonexposure group. Epidemiologically, if you can double the rate of an outcome, it typically becomes a statistically significant association. Thus, it would appear that given the post hoc nature of the design, the study was indeed powered appropriately.

DeFranco: ***Were the analytic methods appropriate for the study design?***

Gross: The analytical methods used by the authors, for the most part, were appropriate for this study design; however, they might have been able to strengthen their analysis by using a couple of additional methods. To understand this, it is important to readdress the study design, which was a case-control design, in which cases and controls are matched. Study groups were matched on 3 variables, including sex, birth year, and hospital of birth. In the end, the authors matched in a 1:1 ratio, despite originally selecting 5 controls per case.

Differences between cases and controls were compared using the χ^2 method, which on first glance, seems to be the most appropriate and accessible means to make a comparison. The same can be stated for the use of logistic regression analysis to estimate the odds ratios (both crude and adjusted) and the 95% confidence intervals.

In this study design, there are certain potential confounders that have been fixed in the process of matching. Because subjects were matched for these variables, no adjustments can be made for them. While investigators are

matching what they consider to be important variables, it cannot be stated with certainty that these actually are the most important variables. Hence, using analytical methods that take the matching process into consideration can further strengthen studies like this one. One example is McNemar's test, which uses 2×2 tables like the χ^2 test. This test is designed to detect differences in matched pairs of subject. Similarly, conditional logistic regression, unlike routine logistic regression, takes fixed variables into consideration.

However, the authors were seeking to find a possible association with a specific exposure, namely RhIg administration. They were not performing a broad analysis intended to find any variable achieving statistical significance. Thus, the researchers have not detracted from this particular study in any major way by not using McNemar's test and conditional logistic regression.

Conclusions

DeFranco: ***Did the discussion section adequately address the study's strengths and limitations?***

Gross: The large study sample is 1 of the strongest, if not the strongest, aspect of this study. The accessibility of the information and the manner in which it was collected (prospectively as part of prenatal care) serve to strengthen the overall validity of the study's findings. The study model is fairly immune to several biases that can often hinder retrospective studies, including recall, ascertainment, and/or selection bias. These are all recognized by the authors. Alternative study designs, which traditionally yield stronger results, such as cohorts or randomized prospective designs, are not addressed, and it may have been helpful for the authors to explain why these designs, although desirable, might not have been feasible to answer the study question at hand.

Additional limitations, such as outcome (autism diagnosis) validation, are recognized. A possible key weakness is the inability to control for other types of mercury exposure, such as environmental and occupational sources. Of interest is fish consumption, which has recently received a significant amount of attention.

The authors did not explain why they went from a 5:1 control-case ratio to a 1:1 design. Using more controls enhances the study's power, especially when the outcome is rare. Perhaps it was simpler to do a 1:1 analysis, and this might have been their intent.

In the end, the question of whether thimerosal is guilty by causation or by association will not be completely answered by this particular study. But it does a very good job of shedding some light on this issue. In addition, it might also assuage the fears of some practitioners and patients; this is very important if such concerns hamper potentially lifesaving treatment with RhIg. As more investigations of thimerosal-containing preparations are completed, more information will become available, a point the authors acknowledged.

DeFranco: After a thorough discussion of the study's methodology, we feel that the authors adequately considered the study's specific strengths and limitations and addressed them in the conclusion section. Therefore, we are generally satisfied with its internal validity.

Do the findings have external validity, that is, are they generalizable to readers' patients? Before physicians can make that determination, they must consider how their own patients might differ from those included in this analysis. ■

REFERENCES

1. E. Fombonne Epidemiological trends in rates of autism. Mol Psychiatry 2002;7(Suppl 2): S4-6.

2. Autism Information Center: frequently asked questions—prevalence. Centers for Disease Control and Prevention Web site. Available at: http://www.cdc.gov/ncbddd/autism/faq_prevalence.htm. Updated Jan. 30, 2008. Accessed July 7, 2008.