Reproductive Toxicology 67 (2017) 125–128

Contents lists available at ScienceDirect

Reproductive Toxicology

journal homepage: www.elsevier.com/locate/reprotox

Review

Publishing negative findings and the challenge of avoiding type II errors in studies of suspect teratogens: Example of a recent ondansetron publication

Stanley A. Edlavitch

UMKC School of Medicine, Department of Psychiatry, 1000 E24th Street, Kansas City, MO 64108, United States

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form
22 November 2016
Accepted 5 December 2016
Available online 15 December 2016

Keywords:
Birth defects
Ondansetron
Hyperemesis gravidarum
Type II error

ABSTRACT

It is important that negative, as well as positive, studies be published to complete the available picture in areas of scientific inquiry. At the same time, it is critical that the implications of a negative study not be overstated and generalized when major issues of study design and data accuracy may be the reason that no relationship was discovered. The challenge of avoiding type II errors in interpreting negative findings has major public health implications, especially when the relationship of an exposure to birth defects is the concern. This is particularly important when interpreting the report by Fazio et al. (June issue of Reproductive Toxicology) on the relationship of ondansetron exposure to pregnancy outcome and birth defects. This review addresses the study design and conclusions and suggests that an alternative concluding statement would be more apropos, given the limitations of the data.

Contents

1. Statement of hypothesis 126
2. Sample 126
3. Validation of data 127
4. Analyses 127
5. Generalizability/Conclusions 127
References 128

There is widespread agreement that it is important negative, as well as positive, studies be published to complete the available picture in an area of scientific inquiry. At the same time, it is critical that the implications of a negative study not be overstated and generalized when major issues of study design and data accuracy may be the reason that no relationship was discovered. Numerous authors have suggested increasing the power of studies to minimize the probability of making a type II error, i.e. the error that occurs when study findings are generalized incorrectly to conclude no difference exists when a real difference exists [1,2]. Increasing sample size does not necessarily address study design and implementation concerns [3].

The challenge of avoiding type II errors in interpreting negative findings has major public health implications, especially when the relationship of an exposure to birth defects is the concern [4]. This is particularly important when interpreting the report by Fazio et al. in the June issue of *Reproductive Toxicology* [5] on the relationship of ondansetron exposure to pregnancy outcome and birth defects. Fazio et al. reported an analysis of fetal outcome in pregnancies exposed to ondansetron to treat Hyperemesis Gravidarum (HG) from a retrospective cohort study of women recruited through advertising on the Hyperemesis Education and Research Foundation Web between 2007 and 2014. The inclusion criteria was a self-reported diagnosis of HG in a singleton pregnancy and treatment with IV fluids and/or total parenteral nutrition/nasogastric feeding tube. Each eligible woman was asked to recruit one acquaintance with at least 2 pregnancies lasting beyond 27 weeks to participate as a control.

Outcomes were collected on 1070 pregnancies exposed to ondansetron, compared to outcomes in two control groups: 771 pregnancies in women with a history of HG with no ondansetron

E-mail address: edlavitchs@umkc.edu

http://dx.doi.org/10.1016/j.reprotox.2016.12.001
0890-6238/

exposure and 1555 pregnancies with neither a history of HG nor ondansetron exposure. Ventricular septal defects were reported in 2/952 of infants in the HG/Ondansetron-exposure group and 4/1286 in the No HG/No Ondansetron-exposure group. Cleft palate was reported in 1/952 live births in the HG/Ondansetron and 2/1286 in the No HG/No Ondansetron-exposure groups. Women with a history of HG who took ondansetron reported less miscarriages and terminations, and higher live birth rates. The authors concluded that the overall results do not support evidence of teratogenicity of ondansetron.

The publication by Fazio et al., addresses an unanswered question about the teratogenesis of ondansetron exposure during the first trimester of pregnancy. As published, this was a poorly designed and implemented study and only confuses the important public health risk questions surrounding ondansetron exposure in pregnancy.

Though ondansetron is approved primarily for nausea associated with chemotherapy and surgery, it is widely used off-label to treat morning sickness in normal pregnancies and for women experiencing Hyperemesis Gravidarum, (HG), particularly after trials of other drugs have failed. The U.S. Food and Drug Administration (FDA) reports that in a population of over 2.1 million pregnancies (Sentinel database), the prevalence of ondansetron use anytime in pregnancy was 14.5%. Ondansetron use increased from <1% of pregnancies in 2001–22.1% of pregnancies in 2014. Much of the increase is attributable to oral ondansetron (beginning in 2006). Use was highest in the first trimester and decreased in the second and third trimesters. [6]. In light of this degree of use, questions of the effects of ondansetron on pregnancy outcome and fetal defects clearly have important public health implications.

The study published by Fazio et al. and used as an example for this paper, was restricted to ondansetron exposure in women who reported they had suffered from HG requiring hospitalization and IV fluids. In other studies, HG has been associated with an increased risk of pre-term birth and small for gestational age neonates [7,8]. Though 50% of pregnant women report nausea and 25% report vomiting during pregnancy, severe symptoms diagnosed as HG affects only 0.3-1.5% of pregnant women.

Two previous epidemiologic studies, which included all pregnant women, suggested a possible increased risk of cleft palate and ventricular septal defects (VSD) associated with ondansetron exposure in pregnancy [9,10]. These relationships have not been found in other studies [11].

The limitations in study design, implementation and interpretation are substantial and subsequently cannot support the authors' contention that "The overall results do not support evidence of teratogenicity of ondansetron" and increase the likelihood of making a type II error in interpreting the findings. The publication provides little to no evidence about the teratogenicity of ondansetron in pregnant women without a reported history of HG and little evidence about the lack of teratogenicity of ondansetron in women who experienced HG.

A more accurate statement, less prone to misinterpretation is "The overall results do not support evidence of teratogenicity of ondansetron, but at the same time does not rule out this possibility" or alternatively "The overall results do not support evidence of teratogenicity of ondansetron, but at the same time due to the study's limitations do not rule out this possibility". Without such a statement, the article published by US News and World Report ("Study: Zofran Not Tied to Birth Defects") or in professional magazines, such as Nursing Standard ("Morning sickness drug is not linked to increased risk of birth defects") are typical of how this study can be generalized without an understanding of its limitations [12,13].

The issues which raise concern about making a type II error, are major questions about:

1 Statement of hypothesis
2 Sample
3 Validation of data
4 Analyses
5 Conclusions

1. Statement of hypothesis

The study is intended to be a descriptive study, as opposed to a hypothesis testing study. No hypothesis is explicitly stated. This distinction becomes increasingly important when the authors interpret their findings. For example, they state that "This well-controlled study shows no statistically significant increase in the overall reporting of major and minor birth defects in women with a history of HG exposed to ondansetron compared to women with a history of HG who did not take ondansetron". The analyses are based on pregnancies, not women. The HG group of 772 women contributed information on 1070 pregnancies with ondansetron exposure and 771 pregnancies in the same women without ondansetron exposure. Each woman who reported a history of HG contributed data on the outcome of 2.4 pregnancies (Average of 1.4 pregnancies with ondansetron exposure and 1.0 pregnancies without).

In the control group, there were 563 women who contributed data on the outcome of 1555 pregnancies or an average of 2.8 pregnancies per control. The statements about the findings being applicable to pregnant women are not worded precisely and can lead to a misleading interpretation of the results. One question which stands out is why were some of the HG group women exposed to ondansetron during one pregnancy and not necessarily others?

The first step in conducting any research is to determine a clear testable statement of the study hypotheses. A part of stating the hypotheses clearly is to develop a protocol that includes a clear definition of the outcomes that will be measured and how exposures of interest and possible confounders will be measured. This study includes women who reported a history of HG requiring IV fluids during one of multiple pregnancies to controls who did not experience severe nausea and vomiting in apparently a larger number of pregnancies per woman, without identifying which pregnancy (1st, 2nd, 3rd,) is contributing to the outcome data. They do not report validation of HG diagnoses, pregnancy outcomes or medication exposure. Taken together, this study does not meet these mandatory minimum requirements of stating a clear testable hypothesis, confirmation of outcomes or evidence of a tightly designed observational study.

2. Sample

This retrospective cohort study is part of a larger survey investigating the genetics and epidemiology of Hyperemesis Gravidarum (HG). Eligible patients were primarily recruited through advertising on the Hyperemesis Education and Research Foundation Web site. The inclusion criteria for women with a history of HG were a self-reported diagnosis of HG in a singleton pregnancy and a self-report of treatment with IV fluids and/or total parenteral nutrition/nasogastric feeding tube. Controls were acquaintances of HG participants with at least 2 pregnancies lasting more than 27 weeks, with no reported IV fluid use.

The authors state that "Albeit rare, some women may have normal nausea/vomiting in one pregnancy and HG in another, and therefore, selecting controls with a minimum of 2 pregnancies with normal or no nausea and vomiting in pregnancy (NVP) helps minimize enrollment of those types of controls. The criteria excludes controls who deliver singleton babies with birth defects,