



contain serotonin; rather, they alter the availability in the nervous system of the serotonin produced by the body. The FDA categorizes Zoloft as a Pregnancy Category C drug.<sup>1</sup>

The parties agree that birth defects, including the cardiac birth defects alleged in this litigation, have occurred throughout history. Major congenital heart defects, which are among the most prevalent birth defects, occur in as many as 1% of live births. Expanding the scope to include all cardiac defects, one finds an incidence of approximately 7.5% of live births. Although some birth defects are caused by known genetic sources or environmental agents (such as certain viruses, radiation exposure, or teratogenic medications), most are due to currently unknown causes. Teratology is the scientific field which deals with the cause and prevention of birth defects.

Where plaintiffs allege that a medication, such as Zoloft, is a teratogen, it is common to put forth experts whose opinions are based upon epidemiological evidence. Although the “gold standard” for epidemiological studies is the double-blind, randomized control trial, such studies may not ethically be conducted on pregnant women. Therefore, the research on teratogens relies upon less rigorous observational studies. This does not mean that inferences about causation cannot be made; it simply means that the researchers (and the litigation expert) must more carefully examine the power of the study to detect associations, the role of chance, and any

---

<sup>1</sup> The FDA has established 5 categories to indicate the potential of a drug to cause birth defects if used during pregnancy. Category A means that there are adequate, well-controlled studies which have failed to demonstrate a risk to the fetus. Few drugs are in category A because controlled studies of medication use during pregnancy are ethically prohibited. Category B means animal studies show no risk, but there are no adequate and well-controlled studies of use by pregnant women. Category C means that animal reproduction studies have shown an adverse effect on the fetus, but there are no adequate and well-controlled studies in humans, and so pregnant women should weigh the potential benefits against the potential risks. Category D is used when there is positive evidence of human fetal risk based on adverse reaction data from investigational or marketing experience or studies in humans, but potential benefits may still warrant use of the drug. Category X is the lowest category, used when use of the drug is not recommended for any pregnant women, as the risks clearly outweigh any benefits. One SSRI, Paxil, is a Category D drug, while all other SSRIs, including Zoloft, are Category C drugs.

possible sources of bias or confounding which may make the study a weak indicator of causation.

The best observational studies are designed and powered to test the outcome of interest (here, cardiovascular birth defects).<sup>2</sup> For rare defects, it is necessary to include very large samples of women in both the exposed and the unexposed groups, or one will not be able to reliably measure any true increase in risk that exists. Where individual studies are underpowered to detect rare outcomes, studies can sometimes be combined using a meta-analysis, to increase the power to detect increases in risk.

Epidemiological studies examining the effects of medication taken during pregnancy on birth defects calculate a relative risk or odds ratio (the Court will use the term “odds ratio” in this opinion). Simply speaking, the odds ratio is calculated by dividing the odds of a particular birth defect occurring in children who were exposed to a medication in utero by the risk or odds of finding that birth defect in children born without prenatal exposure. Where the incidence of birth defects is approximately the same in medication-exposed and unexposed women, the odds ratio will be close to one. The odds ratio is interpreted as the increase in the risk of the outcome of interest (here, cardiac birth defect) associated with the exposure of interest (Zoloft) that is above and beyond the baseline risk.

Researchers often statistically control for certain suspected and measurable confounding factors (*e.g.*, factors such as maternal age, weight, smoking, alcohol use, folic acid use, etc., which are correlated with exposure to the medication, and which may themselves contribute to an increased risk of the birth defect at issue), when information about those factors is available in the data set. When this is done, the researchers will report an “adjusted” odds ratio. The authors

---

<sup>2</sup> *In re Diet Drugs Prods. Liab. Litig.*, MDL No. 1203, 2001 WL 454586 at \*13 (E.D. Pa. Feb. 1, 2001).

of the studies Dr. Jewell has reviewed rely upon adjusted odds ratios, where available, when drawing conclusions.

Because an odds ratio calculation is only an estimate, the precision of which may be affected by general or study-specific factors (including confounders and biases, sample sizes, study methods, etc.), researchers also use statistical formulas to calculate a 95% confidence interval (sometimes abbreviated as CI), which is an estimated range of plausible odds ratio values. A 95% confidence interval means that there is a 95% chance that the “true” odds ratio value falls within the confidence interval range. Some confidence intervals are narrow, indicating that the calculated rate ratio is fairly precise, and some are wide, indicating that it is not and that additional research is warranted. If the lower bound of the confidence interval is greater than one, researchers say that the ratio is “statistically significant” (*i.e.*, there is only a 5% chance that the increased risk reflected in the odds ratio is the result of chance alone), and will report finding a statistically significant *correlation* or association between the medication exposure and the birth defect at issue.<sup>3</sup> A statistically significant result does not necessarily indicate a large increase in risk; it simply indicates that the increased risk found is unlikely to result from chance alone.

Even where the confidence interval is narrow and the increased risk is statistically significant, scientists will not draw firm conclusions from a single study, as apparent associations may reflect random error, bias, confounding, or some weakness in the study design, or they may be incongruous with existing scientific knowledge about biological mechanisms. When specific design flaws or potential confounders or biases are identified, researchers will attempt to design studies in such a way that they can determine the degree to which those factors contributed to an

---

<sup>3</sup> A factor may also be protective. For example, prenatal exposure to folic acid is associated with a decrease in neural tube defects. If a factor is protective, the ratio estimate will be less than one, and if the confidence interval’s upper bound is less than one, that protection is statistically significant.

outcome. In general, before concluding that there is a “true” association between maternal medication use and birth defects, the teratology community requires repeated, consistent, statistically significant human epidemiological findings, and studies which address suspected confounders and biases.<sup>4</sup>

Epidemiological studies alone can only inform scientists that two events (*e.g.*, medication exposure and a birth defect) are associated. For this litigation, Dr. Jewell has been asked to opine as to whether Zolofit *causes* the birth defects at issue, which requires analysis beyond the identification of statistical correlations reported in published epidemiological studies. To infer a causal relationship from an association, scientists look at well-established factors sometimes referred to as the Bradford-Hill criteria. These include: the strength of the association between the exposure and the outcome; the temporal relationship between the exposure and the outcome; the dose-response relationship; replication of findings; the biological plausibility of such an association; alternative explanations for the association; the specificity of the association (*i.e.*, does an outcome have only one cause, or several); and the consistency with other scientific knowledge.

Pfizer’s challenge to Dr. Jewell’s testimony focuses on the methods Dr. Jewell used to determine whether there is a true association between maternal Zolofit use and cardiac birth defects, his assessment of the replicability and consistency of study results, and his efforts to address alternative explanations (chance, bias, confounding) for detected associations and the specificity of the association.

---

<sup>4</sup> “Absent consistent, repeated human epidemiological studies showing a statistically significant increased risk of particular birth defects associated with exposure to a specific agent, the community of teratologists does not conclude that the agent is a human teratogen.” *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1453 (D.V.I. 1994), *aff’d* 46 F.3d 1120 (3d Cir. 1994).

## II. STANDARD OF REVIEW

Federal Rule of Evidence 702 reads:

[I]f scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient fact or data, (2) the testimony is the product of reliable principles and methods; and (3) the witness has applied the principles and methods reliably to the facts.

The Third Circuit has distilled this rule to two essential inquiries: 1) is the proffered expert qualified to express an expert opinion; and 2) is the expert opinion reliable?<sup>5</sup> In this case, Pfizer primarily challenges the reliability of the opinions.

Under the Third Circuit framework, the focus of the Court's inquiry must be on the experts' methods, not their conclusions. Therefore, the fact that Plaintiffs' experts and defendants' experts reach different conclusions does not factor into the Court's assessment of the reliability of their methods.<sup>6</sup> The experts must use good grounds to reach their conclusions, but not necessarily the best grounds or unflawed methods.<sup>7</sup> Expert evidence must be relevant and reliable to be admissible. The Court must consider: 1) whether the expert's theory can be tested; 2) whether studies have been subject to peer review and publication; 3) the potential for error in a technique used; and 4) the degree to which a technique or theory (but not necessarily a conclusion) is generally accepted in the scientific community.<sup>8</sup>

---

<sup>5</sup> *In re TMI Litig.*, 193 F.3d 613, 664 (3d Cir. 1999).

<sup>6</sup> However, where the scientific community considers the evidence to be inconclusive, a difference of opinion may sometimes undermine the reliability of an expert's conclusion that there is a causal link, and may justify excluding that expert. *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 607 (D.N.J. 2002), *aff'd* 68 F. App'x 356 (3d Cir. 2003).

<sup>7</sup> *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 745 (3d Cir. 1994); *Holbrook v. Lykes Bros. S.S. Co.*, 80 F.3d 777, 784 (3d Cir. 1996).

<sup>8</sup> *Daubert v. Merrell Dow Pharma., Inc.*, 509 U.S. 579, 593-94 (1993).

The Court must also consider Rule 403, which provides that “[t]he court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, [or] misleading the jury.”

### **III. DISCUSSION**

Dr. Jewell has asserted that it is his opinion, supported by a reasonable degree of scientific certainty, that maternal use of Zoloft during the first trimester of pregnancy can cause cardiac defects in newborns. In his testimony, Dr. Jewell noted that the research on this issue has produced equivocal results; this, he explained, is why experts can be helpful to the jury as it assesses the scientific literature.

Dr. Jewell’s report includes a detailed review of 11 published, peer-reviewed epidemiological studies which he opines are the “core studies” examining whether the use of SSRIs during pregnancy are associated with an increased risk in birth defects and which provided Zoloft-specific cardiovascular birth defects odds ratios or relative risk ratios.<sup>9</sup> Dr. Jewell noted that several studies found a statistically significant association between Zoloft use during pregnancy and cardiac birth defects, but subsequent studies suggested that those associations might be the result of confounding by indication<sup>10</sup> or an unknown confounding factor. Dr. Jewell reviewed peer-reviewed, published studies which were designed to examine the issue of confounding by indication, but also performed his own re-analysis of some of the

---

<sup>9</sup> Dr. Jewell focused on the following studies: Ban (2014); Huybrechts (2014); Jimenez-Solem (2012); Colvin (2011); Malm (2011); Kornum (2010); Reis-Kallen (2010); Pederson (2009); Alwan (2007); Louik (2007); and Kallen (2007). In addition to these core studies, he reviewed additional studies and meta-analyses, including meta-analyses by Myles (2013) and McDonagh (2014) which are peer-reviewed, published studies but which do not contain original data.

<sup>10</sup> Researchers acknowledge that those who are prescribed a medication may be inherently or systematically different from those who do not take a medication, by virtue of having the medical condition the drug is prescribed to treat (i.e., the “indication” for the medication), and these differences, not the drug itself, may account for the increased risk for the outcomes of interest. For example, maternal depression may itself increase the risk of certain health outcomes in children, or it may be associated with other behaviors which are related to health outcomes (e.g. smoking, obesity). This concept is referred to as confounding by indication.

data from the Jimenez-Solem (2012) and Huybrechts (2014) studies (alone and in combination). Dr. Jewell's reanalysis is discussed in his expert report, but it has not been subject to peer review or been published.

Pfizer does not challenge Dr. Jewell's expertise in statistics, but does challenge Dr. Jewell's qualifications to make certain assumptions about embryological development, heart defects, and categories of antidepressant medications, as it argues that he does not have the necessary expertise to make assumptions in these areas.

Pfizer also argues that Dr. Jewell has employed flawed methodology in forming his opinion, and thus has formed an opinion which is not shared by the relevant scientific community, and which is not supported by the published research. Specifically, Pfizer contends, *inter alia*, that: 1) Dr. Jewell relies upon data from overlapping study populations in asserting that there are replicated, statistically significant studies demonstrating an increase in risk; 2) Dr. Jewell's opinion improperly relies on observed trends in data, and downplays the importance of statistical significance; and 3) Dr. Jewell employs inconsistent standards when reviewing and relying upon studies and types of studies, depending upon whether or not they support his *a priori* opinion. Finally, Pfizer argues that, in assessing the role of confounding factors, Dr. Jewell improperly conducted post-hoc re-analyses and a meta-analysis of certain data. The Court will address each argument in turn.

A. Qualifications to Make Certain Assumptions

Dr. Jewell holds a doctorate degree in mathematics from the University of Edinburgh, Scotland, and he has been a professor of biostatistics at the University of California, Berkeley for more than 30 years. He has authored a textbook entitled *Statistics for Epidemiology*, has published hundreds of peer reviewed papers on biostatistics, and has served as an expert witness

in several pharmaceutical mass tort cases. Pfizer does not challenge Dr. Jewell's expertise in biostatistics.

However, with regard to his opinion in this case, Pfizer challenges Dr. Jewell's qualifications to make certain assumptions about embryological development and the etiology of heart defects, specifically challenging his assumptions that it is appropriate to group certain heart defects in epidemiological studies. While the Court (and Dr. Jewell himself) agrees with Pfizer that Dr. Jewell is not an expert in these areas, the record indicates that, for the most part, Dr. Jewell's opinion does not rely upon assumptions he has independently reached regarding these issues, but rather is based upon the assumptions, definitions, and classifications made by the epidemiological study authors.<sup>11</sup>

For example, Dr. Jewell examined cardiovascular outcomes as they were grouped by study investigators, and "did not independently modify or pool the study endpoints."<sup>12</sup> Dr. Jewell noted:

[the] statistical trade-off when grouping outcome categories for the purpose of reporting and analyses. If an exposure such as Zoloft only increases the risk of *some* of the outcomes in an outcome category, then looking at the whole category will *understate* the adverse effect of Zoloft. . . . On the other hand, with very fine categorizations, one is left with very few defects to assess in a rare sub-category. This means that there will be very little statistical power to definitively address an association with an exposure.<sup>13</sup>

Because of his understanding of the potential benefits and drawbacks of grouping outcomes, Dr. Jewell looked at the data and results for broader categories first (e.g. all cardiac defects), and then for narrower categories and individual defects when the study authors provided information on subcategories. However, in all instances, he used the outcome categories determined by the study authors, and Dr. Jewell clearly considered the benefits and drawbacks of looking at both

---

<sup>11</sup> An exception will be discussed in Section D, below.

<sup>12</sup> Report at 17.

<sup>13</sup> Report at 18.

broad and narrow categories of injuries. Because Dr. Jewell’s opinion does not depend on his making independent assumptions about how to categorize cardiac defects, the Court concludes that he was not making assumptions outside of his area of expertise and finds no fault with his methodology in this regard.<sup>14</sup>

B. Interpretation of the Relevant Research

1. *No Replicated Statistically Significant Findings from Non-Overlapping Data*

The Court has previously noted the importance of replicated, statistically significant findings in the field of teratology, noting that scientists “who are examining potential teratogens generally will not draw causal conclusions in the absence of replicated statistically significant epidemiological findings . . . .”<sup>15</sup>

Dr. Jewell acknowledges the importance of replication of statistically significant results. For example, he noted that random error may account for statistically significant associations (i.e. produce a false positive) when researchers have performed multiple comparisons within the same study, as in most of the core studies regarding SSRIs and birth defects. Although there are statistical methods for correcting for multiple comparisons within a single study (e.g. the Bonferroni correction), these methods are very conservative. Therefore, Dr. Jewell wrote that

the preferred method to assess random variation is to look for replication of risk across different studies and different populations. If similar results for a given association (here, maternal Zolofit use and cardiovascular birth defects) are reported across multiple studies, the probability that a spurious [statistically significant] finding in one study is repeated by chance in another is very low.<sup>16</sup>

---

<sup>14</sup> Similarly, in examining confounding by indication, Dr. Jewell assumes that users of SSRI medications and non-SSRI antidepressants are all being treated for depression. While this assumption may be subject to challenge, Dr. Jewell merely adopts this assumption from the study authors, and does not purport to have the expertise to arrive at it independently.

<sup>15</sup> Memorandum Opinion and Order dated January 23, 2015 at 3; see also Memorandum Opinion and Order dated June 27, 2014.

<sup>16</sup> Report at 6-7.

Although Dr. Jewell articulates the importance of replication of statistically significant results across studies and independent populations, Pfizer argues that Dr. Jewell's opinion is unreliable because it is not, in fact, supported by replicated, statistically significant findings from studies using non-overlapping (independent) populations.

The Court acknowledges that the relevant literature did, a year ago, include some replication of statistically significant results for cardiac defects, which is one reason the Court allowed the PSC to submit a new expert report focusing on cardiac claims. Specifically, a statistically significant positive association between Zolofit use and "all cardiac defects" was reported by both Kornum (2010) and Jimenez-Solem (2012). Louik (2007), Pederson (2009), Kornum (2010), and Jimenez-Solem (2012) all reported statistically significant positive associations between Zolofit use and septal defects. A recent study, Bérard (2015), has also reported a statistically significant association between Zolofit use and certain septal defects.

However, Pfizer argues that there is no true replication of results because: 1) Pederson (2009), Kornum (2010), and Jimenez-Solem (2012) all use overlapping data from the Danish population; 2) the Bérard (2015) study is unreliable; and 3) the Louik (2007) study was corrected and no longer reports a statistically significant association between Zolofit use and septal defects.<sup>17</sup>

With regard to the Danish population studies, Pfizer argues that Pederson (2009), Kornum (2010), and Jimenez-Solem (2012) cannot be treated as replications because they use overlapping data from the Danish population. Moreover, Pfizer argues, the data from all three studies is included in a much larger study by Furu (2015), which did not replicate the results of

---

<sup>17</sup> Pfizer also argues that one of the Danish studies Dr. Jewell relies upon, Jimenez-Solem (2012), concluded that the statistically significant results it obtained for Zolofit were the result of confounding, and did not reflect a true increase in risk associated with Zolofit use. The Court addresses Dr. Jewell's approach to the issue of confounding later in this opinion.

the earlier studies, indicating that some unknown factor, confounder, or bias may have been at play in the earlier studies (as Jimenez-Solem (2012) concluded).

In his report and at the *Daubert* hearing, Dr. Jewell addressed the issue of overlapping study populations, noting that when studies are not completely overlapping in time or population, “the studies in fact provide important evidence of replication of findings. . .”<sup>18</sup> However, Pfizer pointed out at the *Daubert* hearing that the overlap between the studies may be more extensive than Dr. Jewell understood when he was preparing his report,<sup>19</sup> and thus the value of later studies as replications for earlier studies may be diminished.<sup>20</sup>

Shortly after Dr. Jewell produced his expert report (which is dated March 6, 2015), the *British Medical Journal* published Furu (2015). Furu (2015) is a large study, including 7,245 babies exposed to Zoloft in the first trimester. The study population overlapped almost completely with the earlier Danish studies, but also incorporated thousands of additional subjects from other countries.<sup>21</sup> According to Dr. Jewell’s own explanation of the replication value of partially overlapping studies, Furu (2015) should provide important information about the reliability of the odds ratios found in the earlier studies.

Despite including virtually all the data from the earlier Danish studies, Furu (2015) *did not* replicate the findings from those studies. The earlier Danish studies reported that the risk of all cardiac defects and septal defects was approximately tripled in the exposed group (i.e. reported odds ratios at or near 3.0). In contrast, the Furu (2015) study authors found an adjusted

---

<sup>18</sup> Report at 23.

<sup>19</sup> For example, Dr. Jewell learned at the *Daubert* hearing that Zoloft was not approved for use in Denmark during the early years of the Kornum (2010) study. Therefore, with regard to Zoloft, only two years of data from one county in Denmark used by Kornum (2010) was not included in later Danish population studies.

<sup>20</sup> See Doc. No. 1483, Tr. at 3-21.

<sup>21</sup> Jimenez-Solem (2012) and Pederson (2009) are complete subsets of Furu, and Kornum (2010) contained an additional two years worth of data from a single county. In addition to the overlapping data from Denmark, Furu (2015) included data from other countries, and studied a total of 2.3 million pregnancies. Doc. No. 1483, Tr. at 21.

odds ratio of 1.13 (with a narrow CI of .93-1.38, indicating that 1.13 is a fairly precise estimate) for Zoloft and any cardiac defects, and 1.05 (CI of .82-1.35) for Zoloft and septal defects, indicating that there is no association between Zoloft use and cardiac birth defects.

Dr. Jewell testified that he was familiar with Furu (2015), although it was published after his report was prepared. However, he failed to provide any methodological or statistical explanation for why this larger, later study failed to replicate the findings of the earlier studies, or why the earlier studies should be considered more reliable than Furu (2015). Scientists are expected to address and reconcile data that does not support their opinions, and not simply rely upon data which does. Was Furu (2015) better designed, addressing design flaws, biases, and confounders which inflated associations reported in earlier studies,<sup>22</sup> or was it poorly designed and thus unable to detect a true association? Dr. Jewell fails to explain his reasoning, leaving the Court to speculate as to why Dr. Jewell continues to rely upon the earlier Danish studies notwithstanding the very different results of Furu (2015).

Assuming, *arguendo*, that the earlier Danish studies contain reliable measures of increased risk, the Court then turns to the question of whether any studies using independent data sets have replicated the results of the earlier Danish studies. It is not clear the extent to which Dr. Jewell relies upon Bérard (2015) for his conclusion that there are independent, replicated, statistically significant findings regarding an association between Zoloft exposure and congenital septal defects, as Bérard (2015) was not yet published at the time Dr. Jewell submitted his report. Nevertheless, as Dr. Jewell's testimony revealed that he is familiar with the study, and as one odds ratio from the Bérard (2015) study was included on his chart illustrating the data, the Court will address Pfizer's argument regarding the study's reliability.

---

<sup>22</sup> For example, Furu (2015), unlike the earlier Danish studies, included a sibling analysis, using data from women who had had multiple pregnancies.

Typically, the Court would find a statistician's reliance on the results of a peer-reviewed, published epidemiological study to be methodologically sound. However, in this instance, Pfizer's expert (Dr. Kimmel), Dr. Jewell, and at least one member of the PSC's legal team all tried and failed to duplicate the statistically significant results reported by Bérard (2015), using the data published by the study authors, even though they were able to duplicate all the non-significant results she reported. As Pfizer pointed out, this calls into question the reliability of the results reported as statistically significant in the publication.

Dr. Jewell initially testified that he believed the Bérard (2015) results were reliable, despite the fact that he could not duplicate the result with the data reported in the publication, based upon Dr. Bérard's explanation that she had corrected for multiple pregnancies using data not publicly available, and her assurances that her results were correct. However, upon further questioning by Pfizer's counsel, Dr. Jewell testified that he had "no way of checking. . . if Dr. Bérard is either telling the truth or. . . whether she did the analysis correctly from raw data."<sup>23</sup> On cross examination, Dr. Jewell acknowledged that a 2013 abstract for the Bérard (2015) paper reported preliminary results, *all non-significant*, which matched the results Dr. Kimmel, Dr. Jewell, and the PSC produced when they ran the statistical analysis, and the abstract stated that the authors had arrived at those results by utilizing a statistical model which adjusted for multiple pregnancies.<sup>24</sup> After reviewing this abstract during cross examination, Dr. Jewell noted that it did not appear that a subsequent analysis to adjust for multiple pregnancies accounted for the difference between the results reported in the 2013 abstract and the 2015 paper, and testified that he was unable to explain, or even speculate as to, why there was a difference between the results

---

<sup>23</sup> Doc. No. 1483, Tr. at 52.

<sup>24</sup> Doc. No. 1483, Tr. at 53.

reported in the abstract and those reported in the published study.<sup>25</sup> Because, upon cross-examination, Dr. Jewell expressed a lack of confidence in the reliability of the Bérard (2015) paper's reported statistically significant findings, the Court will not allow Dr. Jewell to rely upon this study as evidence of replication of statistically significant findings.

Finally, at the time Dr. Jewell submitted his report, Louik (2007) (which used data from American women) reported a statistically significant association between Zolofit use and septal defects. As originally published, Louik (2007) reported a statistically significant doubling of risk of septal defect (odds ratio of 2.0, with a confidence interval of 1.2-4.0). As a result of an inquiry by Pfizer's expert, the study authors discovered a transcription error that required correction of the originally reported confidence interval. The publisher of the paper, the *New England Journal of Medicine*, has published a correction. The confidence interval is now reported as 1.0-4.0, which is not statistically significant. The highly respected *New England Journal of Medicine* believed this correction to be meaningful, and the journal required the authors to modify their discussion to reflect this change in the reported confidence interval.<sup>26</sup>

Because the Louik (2007) confidence interval remains on the borderline of being statistically significant, and because it is just one of multiple studies, Dr. Jewell testified that this correction did not change his opinion. Dr. Jewell explained: "if we were here and [Louik] was the only study that we had information on regarding Zolofit and cardiac defects and in specific,

---

<sup>25</sup> Doc. No. 1483, Tr. at 53.

<sup>26</sup> Dr. Jewell suggested that certain email correspondence between professionals, which was presented to the Court at the *Daubert* hearing, indicates that other scientists believe that it makes no meaningful difference whether the lower bound of the relevant confidence interval is 1.0 (which is not statistically significant) or 1.2 (the originally reported lower bound of the confidence interval). *See* Doc. No. 1483, Tr. at 33. However, this correspondence could also be read as addressing a different point: does it matter whether the lower bound to the confidence interval was rounded up or down (e.g. up from .99 or down from 1.01 to 1.0). *See* Doc. No. 1481, Tr. at 155-156. The Court's analysis does not require it to adopt one of these interpretations, as statistical significance retains its traditional importance according to a leading scientific publication, but is not the only factor the Court considers.

septal, that [change in the confidence interval] becomes much more important. I think it becomes important to the Court. It becomes important to the statistician. . . . [B]ut that's not the context I was in, in July. I was in the context where [Louik] was one of a dozen or so studies. . . . It's a tiny change in a sea of evidence."<sup>27</sup> Because the odds ratio Louik (2007) reported for septal defects (2.0) was consistent with findings from other studies, and the lower bound of the confidence interval was 1.0 (indicating borderline statistical significance), Dr. Jewell was comfortable ruling out the possibility that the result was obtained by chance, notwithstanding the fact that the odds ratio was no longer statistically significant.<sup>28</sup>

The Court is persuaded that Dr. Jewell did review the then-available evidence and rely upon replication of statistically significant (and borderline significant) results in forming the opinion expressed in his expert report and at the *Daubert* hearing. However, as discussed above, when he was confronted with new and contrary information at the *Daubert* hearing, such as the Furu (2015) results or the extent of data overlap, and given the opportunity to address it, he did not reconcile this information with his opinions.

Also of concern to the Court is that the more recent studies, including but not limited to Furu (2015), which have been well-powered and designed to address issues of bias and confounding, have not confirmed the positive associations between Zolof and cardiac birth defects reported in the studies Dr. Jewell relies upon as evidence of replicated positive findings. Beginning with Jimenez-Solem (2012), researchers have designed their studies to address issues of confounding and bias. With the exception of Bérard (2015) (discussed above), authors of the recent studies (e.g. Ban (2014); Huybrechts (2014); Furu (2015)) have uniformly failed to replicate the associations noted in early studies, and the study authors have concluded that the

---

<sup>27</sup> Doc. No. 1481, Tr. at 156-57. *See also*, Doc. No. 1483, Tr. at 34-35.

<sup>28</sup> Doc. No. 1481, Tr. at 160.

reported association between Zoloft and cardiac birth defects may have been the result of chance, confounding by indication, or other confounders. The Court must, then, examine whether Dr. Jewell used reliable scientific methods in concluding that the non-confirming study results, and especially the recent studies, were not incompatible with his causation opinion.

## 2. *Use of Non-Statistically Significant Study Results*

Pfizer argues that Dr. Jewell’s opinion is not based upon reliable methodology because, like Dr. Bérard, whose expert opinion was previously excluded by the Court, he downplays the importance of statistical significance and relies upon “trends” in odds ratios rather than replication of statistically significant results.

The Court agrees that Dr. Jewell’s approach to the Zoloft data de-emphasizes the traditional importance of statistical significance. Dr. Jewell notes that a non-significant result “does not tell us that the exposure has no effect—only that ‘no effect’ remains *one of the plausible explanations* for the data (but not necessarily the most plausible).”<sup>29</sup> Like Dr. Bérard, he cites to Rothman’s *Modern Epidemiology* textbook for the principle that it is “generally accepted to examine the effect estimates (i.e., Odds Ratio) without exclusion of non-significant results.”<sup>30</sup> Like Dr. Bérard, he points to no other evidence indicating that the fields of epidemiology and teratology have abandoned, or even reduced the importance of, the principle of statistical significance. *The New England Journal of Medicine*’s treatment of the correction to the Louik (2007) study provides evidence to the contrary.

In some circumstances, experts may use congruent but non-significant data to bolster inferences drawn from replicated, statistically significant data. However, in this case, as discussed above, three of the studies Dr. Jewell relies upon to show replication use overlapping

---

<sup>29</sup> Report at 27.

<sup>30</sup> Report at 29.

data from the Danish population, and the early studies suggesting a doubling or tripling of risk have not been replicated by later, well-powered studies which attempt to control for various confounding factors and biases. Thus, the Court must examine how Dr. Jewell addresses non-significant data more closely.

Dr. Jewell's causation opinion relies, in part, upon trends he observes in odds ratios (although Dr. Jewell has assiduously avoided using the term "trend"). Dr. Jewell explained that he believes consistent odds ratios across studies are telling, even if they do not reach statistical significance. He explains that odds ratios may be non-significant because a study is underpowered, or the outcome is biased towards the null hypothesis by misclassification of exposure.<sup>31</sup> In one section of his report, he points out the unlikelihood of finding a consistent increase in risk over different data sets purely by chance.<sup>32</sup> His report provides a hypothetical example: where two independent studies<sup>33</sup> report a positive association with a non-significant p-value of .1, he multiplies the p-values together and concludes that "observing a one in ten (0.1) chance event twice occurs with a probability 0.01."<sup>34</sup> This method of multiplying p-values from independent studies is not a method used in the peer-reviewed, published meta-analyses the Court has reviewed, and the Court has been presented with no evidence that it is accepted as a scientifically reliable methodology. Moreover, although Dr. Jewell's report describes this mathematical approach to determine whether associations are true or the result of chance, Dr.

---

<sup>31</sup> Dr. Jewell points out that some women falsely report taking Zoloft when they did not; others deny taking it when, in fact, they did. Either type of misclassification of exposure can bias the observed association towards the null. Report at 48-49. Both commonly utilized methods for determining exposure (pharmacy dispensing records and maternal self-report) carry a risk of misclassification of exposure, but because researchers cannot assign women to treatment and placebo groups when studying potential teratogens, they must utilize these imperfect methods for assessing exposure.

<sup>32</sup> Report at 32-33.

<sup>33</sup> Several of the studies which report a doubling or tripling of risk rely upon overlapping patient data, and such studies cannot be considered independent.

<sup>34</sup> Report at 32.

Jewell does not then apply this mathematical approach to the p-values that can be calculated from results reported in independent studies of Zolof and cardiac birth defects.<sup>35</sup> Instead, he generalizes, concluding that “the generally consistent findings of cardiovascular birth defects associated with maternal Zolof exposure during the relevant period of pregnancy indicates that these are true, rather than false, positive findings.”<sup>36</sup> Even if the multiplication of p-values were a reliable method for ruling out the role of chance, the Court finds Dr. Jewell’s failure to apply the methodology he outlined to the studies he reviewed problematic.

The Court also notes with concern that, although Dr. Jewell relies heavily upon trends and “general consistency” in the data for his opinion, he selectively emphasizes observed consistency between studies of Zolof and cardiac birth defects only when the consistent studies support his opinion. Dr. Jewell does not squarely address why, if there is a true association between Zolof and cardiac birth defects, relatively few studies report odds ratios consistent with a doubling or tripling of risk (whether significant or non-significant), and most report data consistent with the authors’ conclusions that there is no true association. While he critiques individual studies, he does not explain why he considers generally consistent associations so telling only when they support his opinion. In addition to the small number of studies which reported odds ratios consistent with a doubling or tripling of risk of cardiac birth defects, many studies have reported odds ratios consistent with no increase in risk. Yet, Dr. Jewell does not address the *cumulative* evidence indicative of no association. For example, two recent studies from large, independent data sets, Huybrechts (2014) and Furu (2015), reported consistent odds ratios—but these odds ratios were close to one, suggesting no increased risk. Huybrechts (2014)

---

<sup>35</sup> Although the studies themselves do not report p-values, but only report the odds ratios and confidence intervals, it appears from his report that Dr. Jewell is able to calculate the p-values from the published information. *See* Report at 32.

<sup>36</sup> Report at 33.

reported an odds ratio of 1.09 for all cardiac defects and 1.04 for ventricular septal defects, and Furu (2015) reported an odds ratio of 1.13 for all cardiac defects and 1.05 for septal defects. In contrast to his treatment of cumulative evidence in support of his opinion, when he approached such cumulative evidence which is not supportive of his opinion, Dr. Jewell did not address the likelihood that two or more studies would report odds ratios approaching one if there were a true association between Zolofit and birth defects (i.e. the likelihood that both studies would fail to detect a true increase in risk). Dr. Jewell’s selective emphasis on trends and general consistency only when such concepts support his opinion is one example of “situational science” which renders his opinion unreliable. Another will be discussed in the section to follow.

### 3. *“Situational Science” and Heterogeneity*

According to Dr. Jewell, “[h]eterogeneity is the measure of the variation among the effect sizes reported in [various] studies [and] . . . where heterogeneity is significant, the source of variation should be investigated and discussed.”<sup>37</sup> When heterogeneity is statistically significant, that means the variability between studies is not mere “noise” in the data; it is greater than one would expect by chance variations between studies, and must be accounted for by the researchers.

Dr. Jewell explained to the Court that, when researchers are conducting a meta-analysis (a well-established statistical technique used to combine data from multiple studies, often used when individual studies suggest the possibility of an association between two variables, but are unable to rule out the role of chance due to a lack of statistical power), the authors must measure the degree of variability (heterogeneity) between the parallel results of the studies included in the meta-analysis. Meta-analyses are reliable, Dr. Jewell said, when the results from the individual

---

<sup>37</sup> Report at 57.

studies being combined are similar (homogeneous), but may be unreliable when the results of the individual studies are variable (heterogeneous),<sup>38</sup> if the variation between studies is not properly addressed. Heterogeneity between studies is not assessed subjectively, but is assessed using an objective statistical formula, which produces an estimate of heterogeneity and also a “p-value,” which indicates whether the estimate of heterogeneity is statistically significant.

Pfizer produced evidence that Dr. Jewell relied heavily on two meta-analyses, Myles (2013) and McDonagh (2014) (both published, peer-reviewed, meta-analyses of the effects of maternal SSRI use during pregnancy) as a plaintiffs’ expert in the Prozac birth defects litigation,<sup>39</sup> but he does not rely upon those same studies as a plaintiffs’ expert in the Zoloft litigation. Pfizer suggests that his variable reliance on the same studies in the two cases is results driven: Myles (2013) and McDonagh (2014) reported statistically significant associations between Prozac and the outcomes of interest in the Prozac MDL, but did not report statistically significant associations between Zoloft and the outcomes of interest in this MDL.

Dr. Jewell testified that his variable reliance on these two studies was driven by methodology, not results, and his explanation relied heavily upon the principle of heterogeneity.<sup>40</sup> He explained that, in the Myles (2013) study, heterogeneity was statistically significant for Zoloft and cardiac birth defects, but not for Prozac. Because heterogeneity was not an issue in the Prozac data, Dr. Jewell testified that he felt more comfortable relying upon those results. Dr. Jewell testified that “you can’t use a single statistical tool in all situations blindly as if it will give you interpretable results, ignoring what the data itself is telling you about whether

---

<sup>38</sup> See Report at 57-58.

<sup>39</sup> Like Zoloft, Prozac is an SSRI used for treatment of depression.

<sup>40</sup> See Report at 57; Doc. No. 1481, Tr. at 138-140.

the tool is valid or not.”<sup>41</sup> While the Court finds that Dr. Jewell has set forth a scientific rationale for his variable reliance on relevant meta-analyses in the Prozac MDL and the Zolofit MDL, the Court is concerned that Dr. Jewell selectively relies upon the principle of heterogeneity in a results-driven manner.

That is, Dr. Jewell’s discussion of heterogeneity with regard to the meta-analyses raises another concern about “situational science,” and Dr. Jewell’s failure to *consistently* apply statistical methods he identifies as important to the data at hand. While the heterogeneity of the study results may well be an acceptable explanation for Dr. Jewell’s skepticism towards the Zolofit findings contained in the Myles (2013) and McDonough (2014) meta-analyses, what concerns the Court is that Dr. Jewell makes much of what he characterizes as the “generally consistent” results across the 11 core studies, but fails to statistically calculate the heterogeneity of those 11 studies, instead focusing on a subjective general consistency among the odds ratios in published Zolofit studies. As Dr. Jewell testified that the heterogeneity between studies can be determined statistically, and as he relies upon those statistical measures of heterogeneity in critiquing the quality of key meta-analyses, the Court finds that Dr. Jewell’s reliance on his subjective observation that the reported odds ratios from various studies are “generally consistent,” in another part of his report and in other portions of his testimony, is not based upon any reliable statistical or scientific method. As a statistician, the Court would expect Dr. Jewell to rely upon statistical methods to determine the degree of consistency (or variability), not rely upon trends and generalities, but he again places importance upon statistical principles when they support his opinion, and ignores them when they do not. This cannot be considered a valid methodology.

---

<sup>41</sup> Doc. No. 1481, Tr. at 134.

C. Use of Company Documents

Pfizer also argues that Dr. Jewell inappropriately relies upon statements set forth in Pfizer company documents which are themselves literature reviews, not studies reporting original data, and which are not typical of documents statisticians would generally rely upon in a causation analysis. For example, Pfizer argues, Dr. Jewell relies upon a preliminary draft of a Periodic Safety Update Report (PSUR) concerning growth retardation in children and adolescents, and misrepresents both the purpose of the document and the meaningfulness of the statements he quotes. Pfizer notes that the PSUR does not report on any safety studies conducted by Pfizer regarding use during pregnancy, and as such, it should not be of any interest to a statistician.

In his testimony, Dr. Jewell noted that what interested him about Pfizer's report was the methodology used by Pfizer's employees, not the opinions stated.<sup>42</sup> "[T]hey were doing the same kind of thing, albeit in a smaller set of studies as was pointed out this morning by [Pfizer's counsel] Mr. Cheffo, but the methodology was still the same. You take the studies, you look at them and you see if there's a consistent positive association."<sup>43</sup>

To the extent that Dr. Jewell does rely upon the PSUR as supportive of his conclusions, rather than his methods, the Court notes that the portion of the PSUR Dr. Jewell relies upon is in the form of a brief review of only three published studies, and all three reported statistically significant associations between Zolofit and cardiac birth defects (one of which, Louik (2007), was later corrected to report a non-significant association). It was appropriate for Dr. Jewell himself to review and assess any studies Pfizer cited in the PSUR, and he did this in his expert report. However, the Court has not been presented with any evidence from which it can find that the PSUR is the type of document statisticians such as Dr. Jewell generally rely upon to support

---

<sup>42</sup> Doc. No. 1482, Tr. at 53.

<sup>43</sup> *Id.*

their opinions. The cited studies themselves are a better source of information regarding the methods used and the results of studies of the association of interest, and it is the methods, data, and results that a statistical expert such as Dr. Jewell is called upon to interpret.

The same is true regarding the email summarizing a preliminary literature review (of five studies) conducted by a Pfizer employee in conjunction with a request from the FDA to update the Zoloft label.

The PSC has not demonstrated that Dr. Jewell is the proper witness through which to introduce these documents, as such partial literature reviews are not the kinds of information generally relied upon by statisticians, whether to support their conclusions or their methods. In addition, to the extent that the documents report Pfizer's preliminary concerns about product safety, warranting further investigation, and not final conclusions drawn by Pfizer (as Pfizer argued at the hearing), Dr. Jewell's use of them would potentially be misleading to a jury.

D. Re-Analyses on the Issue of Confounding by Indication

Because the research results regarding Zoloft and cardiac birth defects are equivocal, an expert could assist the trier of fact by opining as to whether the results showing a positive association are indicating an association that is not truly present (a false positive), or the results which show little or no association are failing to detect an effect that is present (a false negative). That opinion must, of course, be based upon reliable methodology.

With regard to Zoloft and cardiac birth defects, some researchers have hypothesized that the association detected in early studies resulted from confounding, and is not a true association. Because confounding by indication is one important source of potential bias, researchers have designed studies to test the hypothesis that confounding by indication accounts for any increased risk. For example, they may look at whether the risks are similar in women taking Zoloft and

non-SSRI antidepressants, or by examining whether the risks are similar in women who take Zoloft during the first trimester and those who are prescribed an SSRI but “pause” their use of it during the first trimester, as Jimenez-Solem (2012) did.<sup>44</sup> Dr. Jewell opines that the associations reported in early studies reflect true associations between Zoloft and cardiac birth defects, and hypothesizes that confounding by indication cannot account for the statistically significant associations reported in the early studies.

After he was retained as a Plaintiffs’ expert in this litigation, and for the purpose of supporting his testimony in this litigation, Dr. Jewell reanalyzed certain data from published, peer-reviewed studies and concluded that his own analyses support his hypothesis that confounding by indication is an insignificant factor, accounting for no more than a 10% increase in risk over the background risk for cardiac birth defects.<sup>45</sup> This Court must determine whether Dr. Jewell used reliable principles and methods in addressing the issue of confounding by indication in his re-analyses.

First, Dr. Jewell explains that if reported positive associations between Zoloft and cardiac birth defects were the result of confounding by indication, researchers would consistently detect a similar increased risk in both the “exposed” group and the “paused” group. Jimenez-Solem (2012) reported an odds ratio of 2.73 for the “exposed” group, and an odds ratio of 1.85 for the “paused” group—both statistically significant increases in risk compared to those who were unexposed—and concluded that a confounding factor present in both groups, rather than Zoloft, was accounting for most of the increase in risk reported in her study. Dr. Jewell, however, was

---

<sup>44</sup> See Jimenez-Solem (2012) (finding the SSRI paused group and the Zoloft exposed group were at similarly heightened (approximately tripled) risk of cardiac malformations). *Cf.*, Kornum (2010) (finding *neither* the Zoloft exposed nor the paused group were at heightened risk of congenital malformations overall (odds ratios 1.4 and 1.1, respectively) and not reporting regarding whether the paused group was at heightened risk of cardiac birth defects).

<sup>45</sup> Dr. Jewell cites Ban (2014) in support of his assertion that confounding by indication accounts for no more than a 10% increase in risk.

not persuaded by the analysis of confounding set forth in this peer reviewed, published study. He ran a supplemental analysis, directly comparing the increase in risk in the “exposed” group to the increase in risk in the “paused” group,<sup>46</sup> and found that although those in both groups were at significantly increased risk compared to the general population, those in the Zoloft-exposed group were at greater risk compared to those in the “paused” group. Thus, although the study authors concluded from their results that confounding was an important factor in the increased risk detected in the exposed group, Dr. Jewell rejected that conclusion, and based upon his own *post hoc* analysis, explained above, he concluded that the conclusions of the study authors were incorrect.

Next, Dr. Jewell reanalyzed data from the Huybrechts (2014) study. Huybrechts (2014), which was recently published in the *New England Journal of Medicine*, included almost one million women, 14,000 of whom the authors categorized as “exposed” to Zoloft during the first trimester. The exposed group included both: 1) women who had filled a prescription for Zoloft after their last menstrual period but prior to a positive pregnancy test and who had sufficient pills to last through at least part of the first trimester,<sup>47</sup> and 2) women who had filled or refilled a prescription at least once during their first trimester. Dr. Kimmel reported that the study had sufficient power to identify a doubling of risk of cardiac defects among the children of Zoloft exposed mother, *if* such an association existed.<sup>48</sup> The study, however, found no such association, reporting odds ratios near one (e.g., 1.09 for cardiac malformations overall) and narrow confidence intervals, indicating that the odds ratios were fairly precise estimates of relative risk.

---

<sup>46</sup> The study authors compared both groups to their sample of unexposed women, but did not directly compare the exposed group to the paused group.

<sup>47</sup> The parties agree that the critical period for cardiac development is early in the first trimester.

<sup>48</sup> Kimmel Report at 5.

Dr. Jewell opined that the subset of women who filled a Zoloft prescription after their last menstrual period and had sufficient pills to overlap with at least part of the first trimester, but who did not refill it during the first trimester, are more properly considered “paused” users. That is, he assumes that those women stopped taking Zoloft by the start of the first trimester, whereas the study authors assumed that they did not. Dr. Jewell does not indicate what information he relied upon in rejecting the assumption of the study authors and drawing his own assumption about women’s medication use. According to the study’s methods section, the study used only pharmacy dispensing records (including the date Zoloft was dispensed and the number of pills supplied) to determine first trimester Zoloft exposure; the authors did not conduct interviews, collect patient questionnaires, review doctors’ records, or compile other sources which could be used to test Dr. Jewell’s assumption that these women did not take the dispensed medication during the first trimester. Dr. Jewell is not an obstetrician, psychiatrist, or medical doctor, and therefore presumably has no first-hand experience regarding the likelihood that women will continue to take prescribed antidepressant medication during pregnancy. And he does not cite to any studies which addressed that issue. The Court finds that Dr. Jewell provides no scientific rationale for discarding the assumption of the study authors, who have greater knowledge of both the subject matter and the data itself, that women who had a prescription with enough pills to extend into the first trimester were exposed, and substituting his own assumption that those women were unexposed. Thus, it appears that Dr. Jewell formed his assumption that women in this group “paused” their Zoloft use during the first trimester without any scientific foundation.

This methodological flaw may have been of less importance if Dr. Jewell had simply opted, based upon his assumption that many of the women who had a supply overlap but did not refill their Zoloft prescription during the first trimester were unexposed, to reanalyze the risk

using the subset of Huybrechts (2014)'s exposed subjects who had filled or refilled one or more Zolof prescription during the first trimester. He did not do this, perhaps because Huybrechts (2014) recognized the potential for misclassification of exposure, and ran secondary analyses using only women who had filled or refilled a prescription once or twice during the first trimester.<sup>49</sup> The results of these secondary analyses were substantially the same, although the confidence intervals were wider due to the decreased number of subjects (decreasing statistical power).<sup>50</sup> The authors concluded from these sub-analyses that misclassification of exposure in the no-refill group was unlikely to bias the overall study results towards the null. These sub-analyses directly address Dr. Jewell's criticism of the study results as potentially biased towards the null hypothesis by misclassification of exposure.

Dr. Jewell conducted an additional statistical analysis of the Huybrechts (2014) data, to examine his hypothesis that confounding by indication did not explain the heightened risk of cardiac birth defects in earlier studies. In this analysis, Dr. Jewell compared the outcomes for women whom he considered "paused" to the women he considered "exposed." The results indicated that women in his "exposed" group were at an increased risk of giving birth to a baby with a congenital heart defect, compared to those in his "paused" group.<sup>51</sup> Because both the women in his "paused" group and those who refilled one or more Zolof prescriptions in the first trimester shared indications warranting a Zolof prescription, Dr. Jewell interpreted the difference in risk between the two groups as supportive of his hypothesis that confounding by

---

<sup>49</sup> Huybrechts (2014) at 2405.

<sup>50</sup> *Id.*

<sup>51</sup> Report at 41. In conducting and interpreting this reanalysis, Dr. Jewell makes much of the difference between his two groups, and downplays the fact that neither group is at a heightened risk of giving birth to a child with birth defects when compared to unexposed women.

indication does not fully explain the positive associations between Zolofit and cardiac birth defects noted in some studies.<sup>52</sup>

It is appropriate for a statistician to design a study and statistically analyze the data collected when testing a hypothesis. However, results-oriented, *post-hoc* re-analyses of existing epidemiological studies are disfavored by scientists and often deemed unreliable by courts, unless the expert can validate the need for reanalysis in some way.<sup>53</sup> Although Dr. Jewell reports that he ran his re-analysis of the Huybrechts (2014) data in order to address the unexamined issue of confounding by indication, study author Krista Huybrechts, Ph. D. noted, in correspondence with Dr. Jewell, that it makes no scientific sense to look, *post hoc*, at the issue of confounding by indication *when no association between Zolofit use and cardiac birth defects was found in the primary analyses*.<sup>54</sup> That is, in conducting his additional analysis, Dr. Jewell seemed to disregard the authors' reported finding that women who were exposed to Zolofit were no more likely than unexposed women to have children with cardiac birth defects. Dr. Huybrechts also challenged Dr. Jewell's definition of "paused" and indicated that she felt his approach to potential confounding variables was flawed.<sup>55</sup> Dr. Jewell did not explain to the Court why his reanalysis to assess the role of confounding was a proper one under the circumstances. These concerns, combined with the problem of Dr. Jewell's re-classification of subjects' exposure to Zolofit without an articulated scientific foundation, are methodological flaws which undermine the

---

<sup>52</sup> Because Daubert requires the Court to focus on methodology, the Court cannot consider whether Dr. Jewell has reached a correct conclusion regarding the role of confounding by indication, but may only examine his methods.

<sup>53</sup> *In re: Lipitor (Atorvastatin Calcium) Marketing, Sales Practices and Products Liability Litigation*, 14-md-2502, Memorandum Opinion and Order dated November 20, 2015 at 32 (collecting cases).

<sup>54</sup> Email dated 7/9/15 from K. Huybrechts to N. Jewell ("Doing such an analysis would be useful if there was a concern about residual confounding by depression and its associated factors in the main analysis. Given that the overall analysis showed a null association and residual confounding is not a concern, what would be the value of such analysis?"). *See also* email dated 7/12/15 from K. Huybrechts to N. Jewell.

<sup>55</sup> Email dated 7/9/15 from K. Huybrechts to N. Jewell. *See also* email dated 7/12/15 from K. Huybrechts to N. Jewell.

reliability of his analysis. For these reasons, the Court finds that Dr. Jewell’s *post-hoc* reanalysis of the Huybrechts (2014) data is not reliable and should not be presented to the jury.

The Court also has concerns about the methodology used in a two-study meta-analysis Dr. Jewell performed, which combined data from Jimenez-Solem’s paused SSRI users (defined in her study as those who stopped taking an SSRI at least 3 months before becoming pregnant, and resumed taking it within a year after the baby was born),<sup>56</sup> with data from women in the Huybrechts (2014) study that Dr. Jewell categorized as “paused” using different criteria (women who filled a prescription for Zoloft after their last menstrual period before conception, with a pill supply overlapping with the first trimester, but who did not refill the prescription in the first trimester).

The Court identifies several problems with Dr. Jewell’s meta-analysis. First, Dr. Jewell failed to present a scientific rationale for combining these two data sets in a meta-analysis (for example, he has not demonstrated that the studies, standing alone, lacked statistical power to study the questions of interest to him). Second, he fails to explain why he limited himself to those two studies, and did not also include studies such as Kornum (2010), which also included a “paused” group. Third, the definitions of “paused” used by Jimenez-Solem and Dr. Jewell are

---

<sup>56</sup> Although Dr. Jewell relied upon “paused” group data from Jimenez-Solem (2014), he expressed concerns about the definition of “paused” used in the study. First, Dr. Jewell criticized Jimenez-Solem (2012) for combining paused users of all SSRIs in their “paused” group. However, he offers no scientific rationale for analyzing paused users of each SSRI separately. The purpose of the “paused” versus “exposed” comparison is to examine issues of confounding by indication. In Jimenez-Solem (2012), “paused” subjects have no prenatal exposure to an SSRI, but share indicators for prescription of an SSRI. Neither side has presented evidence that the indicators for the various SSRIs are different. Therefore, the Court perceives no flaw with Jimenez-Solem’s combining paused users of any SSRI, and doing so provides the advantage of greater statistical power.

Dr. Jewell also criticizes Jimenez-Solem’s definition of “paused,” speculating that if one had a child with a heart defect, one might be more likely to resume taking an antidepressant; therefore, he said that the very definition of “paused” might select into the “paused” group a disproportionate number of non-exposed women who had adverse outcomes. Dr. Jewell is not a psychiatrist or medical doctor, and therefore presumably has no first-hand experience regarding the likelihood that women will resume taking antidepressant medication after pregnancy, but the Court acknowledges the potential for bias inherent in the definition chosen by the study authors. However, Dr. Jewell did not statistically estimate for the Court how large an impact such a bias could have on the results, given the rarity of the cardiac conditions at issue.

completely different, such that women categorized as “paused” by Jimenez-Solem would be characterized as unexposed (not as “paused”) in the Huybrechts sample, using Dr. Jewell’s definition, and women Dr. Jewell categorizes as “paused” in the Huybrechts sample would be categorized as “exposed” by Jimenez-Solem (as they were by Huybrechts in the published study). Dr. Jewell has not established that it is scientifically acceptable to perform a meta-analysis combining data on “exposed” and “paused” women from two different studies when the two categories are so differently defined in the underlying studies being combined. Thus, the Court concludes that no reliable inferences or conclusions can be drawn from the meta-analysis conducted by Dr. Jewell for purposes of this litigation.<sup>57</sup>

### **III. CONCLUSION**

Dr. Jewell opines that Zolofit may cause cardiac birth defects. His opinion is based upon his review and analysis of the relevant literature, and his own reanalysis of certain data from published studies. Plaintiffs argue that Dr. Jewell employed the well-accepted methods of Dr. Jewell’s profession as a statistician in arriving at his opinion. Pfizer has argued that while Dr. Jewell outlines a valid method for assessing the literature regarding Zolofit and cardiac birth defects, he has not faithfully applied that methodology.

“The reliability analysis [required by Daubert] applies to all aspects of an expert’s testimony: the methodology, the facts underlying the expert’s opinion, [and] the link between the facts and the conclusion.”<sup>58</sup> As an expert witness in statistics, faced with equivocal evidence in

---

<sup>57</sup> In addition to the methodological problems outlined, *Daubert* instructs the Court to consider that both the reanalysis of Huybrechts (2014) data, and the meta-analysis of the Jimenez-Solem (2012) and the Huybrechts (2014) data, were conducted by Dr. Jewell solely for the purpose of this litigation, without the participation of scientists with expertise in the fields of embryology and teratology, and have not been subjected to peer review or publication. These issues also weigh against admissibility of the independent analyses.

<sup>58</sup> *ZF Meritor, LLC v. Eaton Corp.*, 696 F.3d 254, 291 (3d Cir. 2012) (internal quotation omitted).

this case, Dr. Jewell would be called upon to: 1) explain why he believes that the positive associations between maternal Zolof use and cardiac birth defects, reported in some papers, are true associations, and not the result of a study flaw, confounding, bias, or other factor; and 2) reconcile those studies which reported no increased risk of cardiac birth defects with his opinion. Because of the methodological weaknesses identified herein, the Court concludes that, in addressing both issues, Dr. Jewell, *inter alia*, has failed to consistently apply the scientific methods he articulates, has deviated from or downplayed certain well-established principles of his field, and has inconsistently applied methods and standards to the data so as to support his *a priori* opinion. It is improper for an expert to take a results-driven approach to a question, molding his methodology and selectively relying upon data so as to confirm his preconceived opinion. Because the Court finds that Dr. Jewell's opinion is based upon his failure to faithfully apply reliable scientific and statistical methods and because his testimony is likely to confuse or mislead the jury, the Court will exclude his testimony at trial under Federal Rules of Evidence 403 and 702, and the principles outlined in *Daubert*.