

# Conclusions Regarding Epidemiology for Obstetric Ultrasound

Jacques S. Abramowicz, MD, J. Brian Fowlkes, PhD,  
Andrea C. Skelly, PhD, MPH, Melvin E. Stratmeyer, PhD,  
Marvin C. Ziskin, MD

## Statement

Based on the epidemiologic data available and on current knowledge of interactive mechanisms, there is insufficient justification to warrant a conclusion of a causal relationship between diagnostic ultrasound and recognized adverse effects in humans. Some studies have reported effects of exposure to diagnostic ultrasound during pregnancy, such as low birth weight, delayed speech, dyslexia, and non-right-handedness. Other studies have not demonstrated such effects. The epidemiologic evidence is based on exposure conditions prior to 1992, the year in which acoustic limits of ultrasound machines were substantially increased for fetal/obstetric applications.<sup>1</sup>

## Rationale and Discussion

Epidemiology is the study of health and disease among groups or populations. This includes analysis of causes and outcomes of diseases and effects thereof among many individuals. Applied to ultrasound, it is the study of its effects on human populations as a result of ultrasound scanning. In the case of obstetric ultrasound, this should include the pregnant patient as well as her fetus. Laboratory animal experiments under diagnostic exposure levels have shown some effects from ultrasound, under certain conditions. However, a definitive statement regarding risk should, ideally, include direct analysis of the effects in human populations. Basic concepts necessary to analyze epidemiologic (or other) studies adequately include measures of disease frequency (prevalence and incidence), study design (observational or experimental), measures of association (relative risk [RR] and odds ratio [OR]), bias, and statistical precision (power and errors). For a more complete discussion, see "Appendix."

## Abbreviations

CI, confidence interval; OR, odds ratio; RR, relative risk

*Received September 18, 2007, from the Department of Obstetrics and Gynecology, Rush University Medical Center, Chicago, Illinois USA (J.S.A.); Department of Radiology, University of Michigan Health System, Ann Arbor, Michigan USA (J.B.F.); Spectrum Research, Inc, Tacoma, Washington USA (A.C.S.); Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US Food and Drug Administration, Rockville, Maryland USA (M.E.S.); and Center for Biomedical Physics, Temple University Medical School, Philadelphia, Pennsylvania USA (M.C.Z.). Revision requested September 20, 2007. Revised manuscript accepted for publication November 20, 2007.*

*Address correspondence to Jacques S. Abramowicz, MD, Department of Obstetrics and Gynecology, Rush University Medical Center, 1653 W Congress Pkwy, Chicago, IL 60612 USA.*

*E-mail: jacques\_abramowicz@rush.edu*

There is a paucity of rigorously conducted epidemiologic studies to evaluate adverse outcomes of ultrasound. A number of studies of the use of ultrasound in pregnancy, including some case-control and prospective randomized control studies, have been performed over the last 30 years. Occasional studies report an association between diagnostic ultrasound and some specific abnormalities such as low birth weight,<sup>2</sup> delayed speech,<sup>3</sup> dyslexia,<sup>4</sup> and non-right-handedness.<sup>5</sup> With the exception of low birth weight (also shown in monkeys<sup>6</sup>), these findings have never been duplicated, and most studies have had negative findings for any association. Caution in interpreting case-control studies is essential because the effect being studied (eg, low estimated weight) may be the reason for performing the ultrasound examination and may thus be found to be associated with it but not through a causal relationship. There have been several major reviews published of epidemiologic studies conducted over more than 25 years.<sup>7-11</sup> Some of the studies had serious limitations, such as small samples, poorly matched controls, and absence of information on acoustic output and quantification of exposure (number of episodes, duration, and "dose" delivered to a particular target). This is particularly relevant in today's clinical situations because of the addition of new modalities with potentially very high energy levels (such as spectral Doppler imaging) and the expansion of diagnostic studies to the first trimester, which is known to be a period of high sensitivity of the fetus to teratologic insults. More recent studies suggest an association between ultrasound exposure during pregnancy and non-right-handedness in male neonates. However, methodological questions exist.<sup>12,13</sup>

### **Low Birth Weight**

Because several animal studies have described, albeit not always convincingly, that ultrasound may cause birth weight reduction, the question of possible effects in humans has been raised. The mechanism whereby a fall in somatic growth would be induced by ultrasound is unclear, but tissue heating and effects on insulin growth factors and heat shock proteins have been considered.<sup>14</sup> Decreased birth weight after prenatal exposure to ultrasound has been reported in the

monkey<sup>6</sup> and the mouse<sup>15,16</sup> but not the rat.<sup>17</sup> In situ intensities were higher than what is considered routine in clinical obstetric imaging in the human. High-level exposures were associated with decreased body weight at birth in exposed animals compared with controls, but all showed catch-up growth when examined at 3 months of age.<sup>6</sup> Among retrospective studies in humans, one<sup>8</sup> showed birth weight reduction (116 g at term). Another retrospective study, with Moore as a coauthor, reported a 2.0 greater risk of low birth weight after 4 or more exposures to diagnostic ultrasound.<sup>18</sup> These results were not reproduced in other retrospective studies.<sup>7,19</sup> Several randomized studies did not show any ill effect of 1 or 2 prenatal scans on growth.<sup>20,21</sup> In fact, in several studies, birth weight was slightly higher in the scanned group, but not significantly so,<sup>20,22</sup> except in one.<sup>23</sup> The only randomized study that reports a difference with a higher proportion of small-for-gestational-age neonates suggests that multiple scans (5, including Doppler flow studies) may produce some decrease in birth weight, possibly via an effect on bone growth.<sup>2</sup> However, when children from the last-mentioned study were examined at 1 year of age, there were no differences between the study and control groups.<sup>24</sup> Overall, the available evidence suggests that routine diagnostic prenatal scanning does not affect birth weight. It is important to remember that all existing epidemiologic data are based on studies using 94 mW/cm<sup>2</sup> as the upper limit.

### **Delayed Speech**

In an attempt to determine whether there is an association between prenatal ultrasound exposure and delayed speech in children, Campbell et al<sup>3</sup> studied the clinical records of 72 children who had undergone a formal speech-language evaluation and were found to have delayed speech of unknown cause. For each subject, 2 control subjects were matched for sex, date of birth, sibling birth order, and associated health problems. The 72 children with delayed speech were found to have a higher rate of ultrasound exposure in utero than the 144 control subjects (OR, 2.8; 95% confidence interval [CI], 1.5-5.3; *P* = .001). There was neither a dose-response effect nor any relationship to time of exposure. Many of the records were more than 5 years old. Confounding is a

serious problem in such a study, in that the conditions prompting the performance of an ultrasound examination during pregnancy may affect the likelihood of delayed speech.

In a later study, Salvesen et al<sup>25</sup> compared the incidence of delayed speech in 1107 children who had been exposed in utero with 1033 controls. They found no significant differences in delayed speech, limited vocabulary, or stuttering. Children who were exposed to ultrasound in utero were less likely to be referred to a speech therapist.

Because of the many weaknesses in the study by Campbell et al,<sup>3</sup> the absence of any credible mechanism for an association of in utero ultrasound exposures, and the finding of no association by Salvesen et al<sup>25</sup> in a much larger sample of children, we conclude that the evidence is not sufficient to claim any association of in utero ultrasound exposure and delayed speech.

### **Dyslexia**

In one study,<sup>4</sup> 425 children, ages 7 to 12 years, exposed to ultrasound in utero were used as a selection group to obtain 381 matched control children to study the appearance of adverse effects. A total of 17 outcome measures were examined. Several of these were determined at birth, including Apgar scores, gestational age, head circumference, birth weight, length, congenital abnormalities, neonatal infection, and congenital infection. The tests performed subsequently in the children included hearing, visual acuity and color vision, cognitive function, and behavior. The authors concluded, "No biologically significant differences between exposed and unexposed children were found."<sup>4</sup> However, of the 17 measures, the authors did note a significantly greater proportion ( $P < .01$ ) of those children exposed to ultrasound to be dyslexic on the basis of a Gray Oral Reading Test. Statistical analysis by the authors indicated that given the design of the study and the numerous end points evaluated, it was possible that this was an incidental finding. In addition, there were several factors that could have contributed to the possible dyslexia finding. The exposure conditions may not have reflected the output of modern ultrasound systems, given that the fetal examinations were performed from 1968 through 1972. However, the finding of potential dyslexia

was considered sufficient to prompt further study. Subsequently, a long-term follow-up study was performed on 2161 children who were part of 2 Norwegian randomized trials.<sup>22,26</sup> The end points included evaluation for dyslexia along with 5 additional hypotheses, including an examination of non-right-handedness (included in a separate discussion here) said to be associated with dyslexia. These studies<sup>27-29</sup> included the specific examination of 603 children with tests for dyslexia, including spelling, reading, and intelligence.<sup>27</sup> There were no statistically significant differences between children screened with ultrasound and controls in teacher-reported school performance in the third year for reading, spelling, arithmetic, and overall performance. Specific dyslexia tests showed no differences between screened children and controls in reading, spelling, and intelligence scores and no discrepancy between intelligence and reading or spelling. The rates of occurrence for dyslexia were 21 of the 309 screened children (7%; 95% CI, 3%–10%) and 26 of the 294 controls (9%; 95% CI, 4%–12%). Given that the original finding of dyslexia was not confirmed in subsequent randomized controlled trials, it is considered unlikely that dyslexia results from routine ultrasound screening examinations. However, these studies did raise the issue of laterality (in terms of handedness), which is also discussed here.

### **Non-Right-Handedness**

The first report of a possible link between prenatal exposure to ultrasound and subsequent non-right-handedness in insonated children was published in 1993 by Salvesen et al.<sup>29</sup> The acoustic outputs of the ultrasound instrumentation used are thought to have been around 1 mW/cm<sup>2</sup>. The authors stressed in their analysis that the association was "only barely significant at the 5% level, and the possibility of a chance finding should be kept in mind." Unknown confounding factors may have been present. The authors recommended that no clinical conclusions be made. When looking further at their data, as a response to a letter to the editor, they described that the association was restricted to boys.<sup>30</sup> A second group of researchers (which included Salvesen, the main author of the first study, but with the new study performed in

Sweden as opposed to Norway) published similar findings of a statistically significant association between ultrasound exposure in utero and non-right-handedness in boys.<sup>5</sup> The fact that similar findings appear to be present in 2 different populations seems to add weight to the proposed finding. In 1999, Salvesen and Eik-Nes<sup>10</sup> published a meta-analysis of these 2 studies and of previously unreported results. No difference was found in general, but a mild difference was present when analyzing boys separately. A word of caution is needed: the use of the term *meta-analysis* is somewhat of a misnomer. Only 2 studies were included. It is generally accepted that a meta-analysis should include at least 6 studies. Four comparisons were reported: (1) an intention-to-treat analysis, comparing those randomly assigned to receive ultrasound versus those in the nonultrasound group; (2) a further analysis of the above, limited to male infants; (3) a comparison of those who were actually scanned versus those who were not, thus breaking the original randomization of each trial; and (4) the analysis in comparison 3 limited to male infants. When analyzing the results, a small increase in non-right-handedness appeared to be present in male infants, but if one adjusts for sex subgroups, wider CIs should be used (97.5% versus 95%), and the OR 1 should be included in the CI, therefore making the difference not statistically significant. Other important points to consider: the studies by Kieler et al<sup>5,31</sup> were population based and observational rather than randomized and controlled, as was the study by Salvesen et al.<sup>29</sup> Furthermore the authors stretched their findings to describe non-right-handedness as brain damage. No valid mechanistic explanation can explain the findings. We conclude that, although there may be a small increase in the incidence of non-right-handedness in male infants, there is not enough evidence to infer a direct effect on brain structure or function or even that non-right-handedness is an adverse effect.

### Conclusions

In addition to the above, it must once more be stressed that there has been no epidemiologic study published on populations scanned after 1992, when regulations were altered and the

acoustic output of diagnostic instruments was permitted to reach levels many times higher than previously allowed (from 94 to 720 mW/cm<sup>2</sup> spatial-peak temporal-average intensity for fetal applications). In several studies,<sup>22,27,29</sup> levels reported by the manufacturers were no more than 1.5 W/cm<sup>2</sup> spatial-peak temporal-peak intensity. There are no epidemiologic studies related to the output display standard (thermal and mechanical indices) and clinical outcomes. The safety of new technologies (harmonic imaging, Doppler imaging [spectral and color], 3-dimensional imaging, and the use of ultrasound contrast agents) as well as that of probe self-heating<sup>32</sup> needs to be investigated.

Major limitations of epidemiologic studies are that (1) a testable hypothesis for causation of the studied effect is not always apparent, and (2) information about exposimetry (pulse parameters and exposure duration) is almost never presented or known. These deficiencies need to be rectified in future research to arrive at truly valid and clinically meaningful results. While ultrasound may have bioeffects, based mainly on two mechanisms of action (thermal and mechanical), many uncertainties persist. These include, among others, existence of a threshold or lack thereof, importance of exposure duration (type of examination, examiner's skills, and multiple examiners), influence of repeated scans (question of cumulative doses), and biological response differences secondary to tissue properties, susceptibility (eg, fetuses at different gestational ages or brain versus bones) and environment (eg, elevated temperature in the mother or medical conditions complicating the pregnancy). A further confounding factor is the fact that major anomalies occur in 3% to 5% of the general human population. An increment of 1% to 2% over this "background" incidence would be a major clinical effect but might go undetected as an individual finding in routine clinical practice.

Epidemiologic evidence is thus not sufficient to establish a causal relationship between ultrasound and adverse effects. Biological effects have been shown using some forms of ultrasound in animal and in vitro models. Subtle or transient effects in humans are possible, but none have been consistently demonstrated so

far; therefore, diagnostic ultrasound remains without any known risk. Furthermore, while not always relevant in an analysis of epidemiologic data, risk-benefit issues are extremely important in clinical practice. Because risks of adverse effects appear so low and clinical benefits so great, there is no justification to withhold the prudent use of diagnostic ultrasound in medically indicated conditions.

## Appendix: An Epidemiology Primer

### *Measures of Disease Frequency*

The first point to consider in any analysis of a disease or a condition is its frequency. The two basic parameters are prevalence and incidence. Prevalence is the proportion of affected individuals at a specific point in time (also known as point prevalence). It is obtained by dividing the total population into the number of affected individuals. Incidence is the frequency of new cases over a particular measure of time. Further characterization of incidence can include cumulative incidence and incidence density. Cumulative incidence describes the proportion of new cases over a certain time period. This is obtained by dividing the number of new cases by the total at risk population during that time period. Incidence density adds into the denominator the amount of time each person under study has been at risk.

### *Study Design*

The next major area to discuss is the study type or design. Epidemiologic studies may be observational or experimental. Observational studies are further divided into descriptive (with cross-sectional or longitudinal designs) and analytical (with case-control or cohort designs depending on whether the researcher begins with the disease or the exposure) studies. Descriptive studies include case reports, case series, correlational studies, and cross-sectional studies. These types of studies are helpful in characterizing aspects of exposure and outcome (who gets it, where, and when), providing clues regarding etiology, assisting with diagnosis and planning, and, from a research standpoint, helping generate hypotheses. They do not, however, provide information regarding alternative explanations or compar-

isons between exposed and nonexposed. Additionally, timing of exposure and development of a disease or condition may not be clear. These studies neither provide enough information about multiple factors that may influence suspected associations nor allow testing the validity of potential associations.

Analytical studies test hypotheses including those that are generated by descriptive studies. Experimental studies include natural experiments and randomized controlled trials. Longitudinal descriptive, cohort analytical, and randomized controlled trials are prospective (although cohort studies may be retrospective), and case-control studies and natural experiments are retrospective. In case-control studies, the subjects are examined once. In cohort studies, the subjects are followed longitudinally over time and examined more than once, usually prospectively but occasionally retrospectively. Analytical studies involve comparisons between groups of study participants with different exposures or outcome statuses. They generally allow for evaluation of the timing of exposure relative to outcome. They have, therefore, the ability to evaluate the suspected presence of an association between exposure and outcome frequency and prevalence. Interventional studies provide the best opportunity to parallel the goal of the experimental laboratory studies, namely, to have participants be as similar as possible in all characteristics, except the exposure of interest (controlled by the investigator) so that an unbiased delineation of the effect of that exposure can be made. This is the most rigorous scientific approach, which can be taken to accept the null hypothesis, that is, that the exposure is not associated with the outcome being studied.

### *Measure of Association*

In epidemiologic (and other) studies, several parameters are used to measure association: RR, OR, and attributable risk. The RR (also known as the risk ratio or rate ratio) is the likelihood of developing a disease or condition in an exposed group compared with a nonexposed group. It is equal to the ratio of the incidence of the outcome in the exposed group (eg, disease) divided by the incidence in the nonexposed group. It is used in prospective cohort studies. Relative risks and

ORs both provide estimates of excess risk. The OR compares the odds of having been exposed among positive cases to the odds of having been exposed among controls. An association is present if cases are more (or less) likely to have been exposed than controls. This is relevant in retrospective case-control studies. Because both the RR and OR are ratios, a value of 1 indicates no difference between subjects and controls. A ratio less than 1.0 means protection from the outcome, and a ratio greater than 1.0 indicates risk of the outcome. Both are estimates of risk, based on the sampled population. The true population risk cannot be determined. Therefore, a range is calculated. This range describes our confidence that the true risk to the entire population is within this range; thus, it is called the CI. Generally, a 95% CI is reported. This means that 95% of the time, the true risk for the entire population will be within the calculated range. The attributable risk is the difference between the incidence of the disease/condition in the exposed and unexposed. This measure is often used to describe the amount of disease that may be prevented by avoiding or removing the exposure. It is important in determining the relevance of findings to specific clinical decisions. Power is another major concept. Statistical power refers to the strength of the conclusions that can be drawn from a particular set of data. Four components can be distinguished: the sample size (large enough to represent the entire population), the effect size (significance of the factor that makes the population studied different from the control), a or significance level (odds that the observed effect is due to chance alone), and power (odds that an effect will be observed when present). Knowing 3 of the 4 elements allows for calculation of the fourth, for example, the sample size needed to observe a particular event with an appropriate certainty of correlation with the exposure. A weakness of epidemiology is that the power needed to be worthy of clinical consideration may be prohibitive. Indeed, the background of birth defects in the general population is considered to be 3% to 5% of live births. If, clinically, only 1% is added to the rate of malformation induction, would it be detectable? The problem of a large number of affected individuals combined with a low ability to detect the increment is a major epidemiologic challenge.

### **Bias**

An important element that needs to be taken into account when analyzing epidemiologic studies is bias. Is the association that is being observed truly related to the exposure, or was there an error (sometimes very difficult to discern) in the methods? Bias is an error that may lead to inaccurate estimation of the true association between exposure and the disease/condition. Bias can be of several types. The major ones to consider include selection (due to differences in properties or characteristics between cases selected or not), information recall (and its correlate, misclassification), diagnostic testing (cases are tested more often than controls), and attrition (potential study cases are eliminated, resulting in a higher number of "healthy" subjects). Furthermore, one must be cautious that confounding factors are distinguished: some findings may naturally be present because of certain characteristics such as age and sex, which cause some unbalance, potentially affecting results of the research because of spurious associations.

### **Statistical Precision**

In certain studies, too few cases are described to permit a meaningful estimate of the population risk. Meta-analysis is an attempt to remediate this problem. It is a method for combining results of a number of small trials, addressing the same questions but with different outcomes, to gain statistical power for determining significant differences that the smaller trials individually are unable to detect. Single studies often cannot detect or exclude small, although potentially clinically significant, differences in the effects of two alternatives, resulting in false-negative results, or a type II error. The other type of error is the false-positive, or type I, error, that is, showing significant differences that, in fact, are due to chance. Several independent separate studies can be combined or integrated and analyzed together, if considered similar enough, thus reducing or eliminating type I and, more particularly, type II errors. It is important to evaluate the quality of the studies included in the analysis because poor studies (such as one with poor inclusion criteria or an obvious bias) can invalidate the results of the meta-analysis. Although a meta-analysis is not about original research, it

may contain original findings. Other terms describing this process are systematic review, pooling of data, and quantitative synthesis. These are more appropriate for descriptive analysis, while meta-analysis actually performs data integration. Certain criteria need to be fulfilled, similarly to any general statistical analysis: formulation of a hypothesis, methods and criteria for inclusion and exclusion, collection of data, analysis, and reporting. Occasionally, the results are reported in a graph form, with a vertical axis representing the value 1 and ORs of the results of different studies on either side of the 1, left being less significant and right being more significant. If these results cross the 1 line, the odds are not significantly different.

A further major statistical concept to keep in mind when looking at epidemiology is causality. Epidemiology reveals associations between exposure and observed effects. These may be weak or strong, true or random, and statistically significant or not. Epidemiology does not determine causality. Causality is judged by the strength of the association, consistency, specificity, the relationship in time, the biological gradient, biological plausibility, and experimental evidence.<sup>9</sup> The last concept to consider when looking at epidemiologic studies is critical appraisal. This includes categorizing the level of evidence and weighing the risks and benefits of a given treatment or diagnostic test based on individual patient presentation and the quality of existing evidence. The level of evidence varies between level I, which is a randomized controlled trial study, the reference standard, and level III, evidence derived from reports by experts (opinions rather than actual data). The RR and OR determine the association (or lack thereof) between an exposure and an effect. It is vital to verify whether the association is clinically relevant. Specific criteria, as described by Bradford-Hill,<sup>33</sup> consist of the type of experiment, strength, consistency, gradient, biological plausibility, specificity, coherence, temporality, and analogy. All of the above form the basis of looking for, analyzing, and using research and its results, that is, evidence-based medicine. When evaluating epidemiologic studies, critical appraisal allows one to categorize the data in several levels to formulate an opinion regarding validity and clinical relevance. The

highest level, the reference standard, is the randomized controlled trial. The lowest is a report from a committee of experts.

## References

1. American Institute of Ultrasound in Medicine. Conclusions Regarding Epidemiology for Obstetric Ultrasound. Laurel, MD: American Institute of Ultrasound in Medicine; 2005. Available at: [http://www.aium.org/publications/statements/\\_statementSelected.asp?statement=16](http://www.aium.org/publications/statements/_statementSelected.asp?statement=16)
2. Newnham JP, Evans SF, Michael CA, Stanley FJ, Landau LI. Effects of frequent ultrasound during pregnancy: a randomized controlled trial. *Lancet* 1993; 343:887–891.
3. Campbell JD, Elford RW, Brant EF. Case-control study of prenatal ultrasonography exposure in children with delayed speech. *Can Med Assoc J* 1993; 149:1435–1440.
4. Stark CR, Orleans M, Haverkamp AD, Murphy J. Short- and long-term risks after exposure to diagnostic ultrasound in utero. *Obstet Gynecol* 1984; 63:194–200.
5. Kieler H, Axelsson O, Haglund B, Nilsson S, Salvesen KA. Routine ultrasound screening in pregnancy and the children's subsequent handedness. *Early Hum Dev* 1998; 50:233–245.
6. Tarantal AF, Hendrix AG. Evaluation of the bioeffects of prenatal ultrasound exposure in the cynomolgus macaque (*Macaca fascicularis*), II: growth and behavior during the first year. *Teratology* 1989; 39:149–162.
7. Lyons EA, Dyke C, Toms M, Cheang M. In utero exposure to diagnostic ultrasound: a 6-year follow-up. *Radiology* 1988; 166:687–690.
8. Moore RM, Diamond EL, Cavalieri RL. The relationship of birth weight and intrauterine diagnostic ultrasound exposure. *Obstet Gynecol* 1988; 71:513–517.
9. Ziskin MC, Pettiti DB. Epidemiology of human exposure to ultrasound. *Ultrasound Med Biol* 1988; 14:91–96.
10. Salvesen KA, Eik-Nes SH. Ultrasound during pregnancy and subsequent childhood non-right handedness: a meta-analysis. *Ultrasound Obstet Gynecol* 1999; 13:241–246.
11. National Council on Radiation Protection and Measurements. Exposure Criteria for Medical Diagnostic Ultrasound, II: Criteria Based on All Known Mechanisms. Bethesda, MD: National Council on Radiation Protection and Measurements; 2002. Report 140.
12. Abramowicz JS, Kossoff G, Marsal K, ter Haar G. Literature review by the ISUOG Bioeffects and Safety Committee. *Ultrasound Obstet Gynecol* 2002; 19:318–319.
13. Salvesen KA. EFSUMB safety tutorial: epidemiology of diagnostic ultrasound exposure during pregnancy—European committee for medical ultrasound safety (ECMUS). *Eur J Ultrasound* 2002; 15:165–171.
14. Angles JM, Walsh DA, Li K, Barnett SB, Edwards MJ. Effects of pulsed ultrasound and temperature on the development of rat embryos in culture. *Teratology* 1990; 42:285–293.

## Conclusions Regarding Epidemiology for Obstetric Ultrasound

15. Hande MP, Devi PU. Effect of in utero exposure to diagnostic ultrasound on the postnatal survival and growth of mice. *Teratology* 1993; 48:405–411.
16. O'Brien WD Jr. Dose-dependent effect of ultrasound on fetal weight in mice. *Ultrasound Med Biol* 1983; 2:1–8.
17. Vorhees CV, Acuff-Smith KD, Schilling MA, et al. Behavioral teratologic effects of prenatal exposure to continuous-wave ultrasound in unanesthetized rats. *Teratology* 1994; 50:238–249.
18. Marinac-Dabic D, Krulewitch CJ, Moore RM. Birth weight in relation to frequent prenatal exposures. *Am J Epidemiol* 1994; 139:S62.
19. Smith CB. Birth weights of fetuses exposed to diagnostic ultrasound. *J Ultrasound Med* 1984; 3:395–396.
20. Saari-Kampainen A, Karjalainen O, Ylostalo P, Heinonen OP. Ultrasound screening and perinatal mortality: controlled trial of systematic one-stage screening in pregnancy. The Helsinki Ultrasound Trial. *Lancet* 1990; 336:387–391.
21. Ewigman BG, Crane JP, Frigoletto FD, LeFevre ML, Bain RP, McNeillis D. A randomized trial of prenatal ultrasound screening in a low-risk population: impact on perinatal medicine. RADIUS Study Group. *N Engl J Med* 1993; 329:821–827.
22. Bakketeig LS, Eik-Nes SH, Jacobsen G, et al. Randomised controlled trial of ultrasonographic screening in pregnancy. *Lancet* 1984; 2:207–211.
23. Waldenstrom U, Axelson O, Nilsson S, et al. Effects of routine one-stage ultrasound screening in pregnancy: a randomized controlled trial. *Lancet* 1988; 2:585–588.
24. MacDonald W, Newnham J, Gurrin L, Evans S. Effect of frequent prenatal ultrasound on birthweight: follow-up at 1 year of age. Western Australian Pregnancy Cohort (Raine) Working Group. *Lancet* 1996; 348:482.
25. Salvesen KA, Vatten LJ, Bakketeig LS, Eik-Nes SH. Routine ultrasonography in utero and speech development. *Ultrasound Obstet Gynecol* 1994; 4:101–103.
26. Eik-Nes SH, Okland O, Aure JC, Ulstein M. Ultrasound screening in pregnancy: a randomized controlled trial. *Lancet* 1984; 2:1347.
27. Salvesen KA, Bakketeig LS, Eik-Nes SH, Undheim JO, Okland O. Routine ultrasonography in utero and school performance at age 8–9 years. *Lancet* 1992; 339:85–89.
28. Salvesen KA, Vatten LJ, Jacobsen G, et al. Routine ultrasonography in utero and subsequent vision and hearing at primary school age. *Ultrasound Obstet Gynecol* 1992; 2:243–247.
29. Salvesen KA, Vatten LJ, Eik-Ness SH, Bakketeig LS. Routine ultrasonography in utero and subsequent handedness and neurological development. *BMJ* 1993; 307:159–164.
30. Salvesen KA, Eik-Ness SH, Vatten LJ, et al. Routine ultrasound scanning in pregnancy: authors' reply. *BMJ* 1993; 307:1562.
31. Kieler H, Cnattingius S, Haglund B, Palmgren J, Axelsson O. Sinistrality—a side effect of prenatal sonography: a comparative study of young men. *Epidemiology* 2001; 12:618–623.
32. Calvert J, Duck F, Clift S, Azaime H. Surface heating by transvaginal transducers. *Ultrasound Obstet Gynecol* 2007; 29:427–432.
33. Bradford-Hill A. *Principles of Medical Statistics*. 9th ed. New York, NY: Oxford University Press; 1971.