KJELL Å. SALVESEN

ROUTINE ULTRASOUND IN UTERO AND DEVELOPMENT IN CHILDHOOD - A RANDOMIZED CONTROLLED FOLLOW-UP STUDY

Department of Community Medicine & Family Practice
National Center for Fetal Medicine Department of Ob. & Gyn.

University of Trondheim
Faculty of Medicine
KJELL Ä. SALVESEN

ROUTINE ULTRASONOGRAPHY IN UTERO AND DEVELOPMENT IN CHILDHOOD - A RANDOMIZED CONTROLLED FOLLOW-UP STUDY

TAPIR
1993
Contents

Acknowledgments 5
Summary 7
Papers included in the thesis 9
Introduction 11
   Ultrasound physics 11
   Bioeffect mechanisms 12
   Theoretical considerations 13
   Experimental studies 13
   Human studies 14
   Recommendations 16
Study objectives and hypotheses 18
Study population 19
Methodological considerations 21
   Ultrasound exposure 21
   Procedures 22
   Outcome variables 23
   Blinding 25
   Statistical analysis 25
Main results 26
Discussion 28
   Chance 28
   Internal validity 29
   External validity 31
   Causality 31
   Further research 34
Conclusions 35
References 36
Errata 41
Papers I – V
Acknowledgments

I wish to express my gratitude to a number of people who made important contributions to the accomplishment of this thesis. Professor Leiv S. Bakkevig introduced me to the field of epidemiology during my student years. He was the initiator of this work and my principal supervisor, and I could not have done this work without his support and encouragement. During the progress of the study my daily mentor has been Professor Sturla H. Eik-Nes. He is one of the pioneers in the field of ultrasound in obstetrics and gynecology, and he has been a gatekeeper to national and international fora where I have met the foremost experts in the world. His never-failing enthusiasm has been invaluable to my work. Dr. med. Lars J. Vatten returned from US as a highly qualified epidemiologist just when I needed him most. During the last two years he has been my tutor and friend, and his help during the statistical analysis and the writing of reports has been of utmost importance to the quality of my work.

Professor Kåre Molne, head of the Department of Obstetrics and Gynecology, was always encouraging and of great help in making the study progress possible. Within the field of psychology I had need to challenge the competence of Professor Johan Olav Undheim, whose inspiring discussions and advice have been of great help. Furthermore, I want to thank my co-authors, Dr. PH. Geir Jacobsen, Dr. Ove Økland, and Professor Kenneth Hugdahl for their important contributions to this thesis.

I also like to thank my colleagues at the Department of Community Medicine and Family Practice for providing me with excellent working conditions and for many stimulating discussions. “The Epidemiology Group” was of particular value to my education and research.

I am also grateful to Ms. Åslaug Gjermundrød, Ms. Sissel Vindvik, Ms. Liv Roald, and Ms. Inger Anne Pareliussen for their help in the examination of the children in Ålesund. Ms. Judith R. Johannesen has given valuable computer assistance, and secretaries Ms. Berit Eggen, Ms. Brit Fladvad, Ms. Eva Dahl, and Kurt Saake have given important technical assistance. I also want to thank the doctors and midwives at the National Center for Fetal Medicine for their support and all the fun we have had. Ms. Nancy Lea Eik-Nes has revised all my manuscripts.

This research has been carried out while I have been receiving a research fellowship from the Norwegian Research Council for Science and the Humanities (NAVF). Financial contributions have also been provided by the Norwegian Society for Ultrasound Diagnosis in Medicine, Trøndelag Medisinske Selskap, Scanmed A/S, and MEBI A/S.
The children, parents, teachers, and public health nurses participating in this research are the real back-bone of my thesis, and I am greatly indebted to all of them.

Finally, I want to thank my wife, Randi, and our children, Marianne, Eirik and Ingerid Birgitte, for their enduring patience and support during the course of this work.

Trondheim in April 1993
Kjell Å. Salvesen
Summary

Purpose

There has been concern that ultrasound exposure in utero might produce subtle changes in the developing fetal brain. This thesis is a follow-up study of children whose mothers participated in two randomized controlled trials of ultrasound screening in pregnancy. The objective of the follow-up study was to test six specified hypotheses related to growth and higher neurological functions.

Subjects

In 1979-81, 2,637 pregnant women were randomly divided into either a screening group who were offered two ultrasound scans during pregnancy, or into a control group who were examined with ultrasound when clinically indicated. In 1988, 2,428 singletons were eligible for follow-up, and 2,161 (89%) children were included in the study.

Methods

Data were collected from parents, from nurses and doctors at maternal and child health centers, and from school teachers. Parents responded to a questionnaire with 66 questions about sociodemographic data and the child's development. Included in the questionnaire were 21 questions about handedness; three about hearing; four about vision; and six about attention, motor control and perception. We collected height and weight data from maternal and child health center records of visits at three, six, and twelve months, and at two, four, and seven years. Distant visual acuity tests and pure tone audiometry were done at the visits at four and seven years. Results of a short version of the Denver Developmental Screening Test recorded during the first year of life were available for 1,657 children. In the second year of primary school, 2,011 children were evaluated by their teachers with regard to reading aptitude, spelling, arithmetic and overall performance. A subsample of 603 children was tested with specific tests for dyslexia in the third year of school.
Results

Routine ultrasonography offered in weeks 19 and 32 of pregnancy did not lower school performance, as reported by teachers, among children aged 8 or 9 years, and there was no evidence of an increased prevalence of dyslexia among children whose mothers were offered routine screening. A teacher rating scale was found to be a valuable tool in screening for learning disabilities among second grade primary school children.

Routine ultrasonography had no adverse effects on sensory functions of those children who were followed up to primary school age. We found no association between ultrasound and impaired neurological development. However, we found a possible association between ultrasound and subsequent non-right handedness in the children.

We found no statistically significant differences in mean body weight or height between screened and control children in a cross-sectional analysis of growth during childhood. However, in a repeated measures analysis of variance in a subsample of children, we found a statistically significant difference in growth curves between screened and control children among mothers who reported smoking at their first antenatal visit. Contrary to this, there was no such difference in growth among children of non-smoking mothers.

Conclusion

Routine ultrasonography in utero had no effect on higher neurological functions as measured by teacher assessments of school performance, specific tests for dyslexia, the Denver Developmental Screening Test during the first year of life, and parental assessments of attention, motor control and perception in childhood. Deficits in sensory functions in children and childhood growth are also unrelated to ultrasound exposure status in utero. The data suggest a possible association between routine ultrasonography in utero and non-right handedness among children. Since the association was weak, and the remaining five null hypotheses were not rejected, this result may be due to chance. Nonetheless, the study suggests that the association between ultrasound and non-right handedness should be tested in future studies.
Papers included in the thesis


Introduction

The introduction of diagnostic ultrasound has been one of the great medical advances in recent decades, and ultrasound has become widely used in most fields of medicine. The use of ultrasonography has in particular improved the clinical care of obstetrical patients and permitted the introduction of many new diagnostic and therapeutic invasive procedures. Furthermore, the use of ultrasound has expanded our research potential and understanding of human embryonic and fetal development and maldevelopment.

Today, most fetuses in developed countries are exposed to diagnostic ultrasound devices. Four European countries (Germany, Norway, Iceland, and Austria) have organised ultrasound screening programs for all pregnant women. However, the efficacy of routine screening has been questioned (1). Because of the exposure of vast numbers of the general population to ultrasound, any possibility of harmful effects becomes very important. Although no adverse effects arising from ultrasound examinations during human pregnancy have been identified, clinical safety is and will always remain a concern. Lack of risk has been assumed rather than clearly demonstrated (2).

It is important to be aware of the fact that no studies can unequivocally prove that diagnostic ultrasound is safe, because safety implies absence of any deleterious effect, recognized or unrecognized. Also, the meaning of the word “safe” is usually vague, since the criterion for safety is hardly ever stated. Therefore, properly planned studies should be implemented to provide data which may be helpful in the overall assessment of risk associated with ultrasound.

Ultrasound physics

Intensity is the physical concept which most often is quantitatively invoked to describe the strength of an ultrasonic field. One reason for the use of intensity is convenience, since it is relatively easy to measure. Intensity is the rate at which energy is transmitted by the ultrasonic wave over a small area (3). The units are watts/cm² (W/cm²) or milliwatts/cm² (mW/cm²). Two spatial terms are used to describe intensity, spatial peak (SP) and spatial average (SA). The spatial peak intensity is the greatest intensity value in an ultrasonic beam, whereas the spatial average intensity is the intensity value which has been spatially averaged over the cross-sectional area of the beam. For diagnostic ultrasound equipment such as B-mode imaging, ultrasonic energy is emitted intermittently in the form of pulses rather than emitted constantly. Typically, ultrasonic pulses are generated once every millisecond (ms), and the pulse lasts for about 1 microsecond. Therefore, the ratio of the pulse average intensity
to the temporal average intensity is 1,000. The intensity output level is often given as a time averaged value, and the spatial peak temporal average (SPTA) value has been most frequently used. For diagnostic equipment used around 1980, SPTA intensity values would typically be in the range 0.1 to 12 mW/cm² (4). Over the years, there has been a progressive increase in time-averaged intensities, and the average SPTA intensity from modern equipment is between two and three times higher than that of the equipment used in 1980 (5).

Diagnostic ultrasound implies many other modalities than B-mode imaging, and transducers may be operating in a variety of modes; M-mode; pulsed Doppler; colour Doppler; and continuous-wave Doppler modes. Today, the average value of SPTA intensity for B-mode imaging is typically around 10 mW/cm². The intensities for colour Doppler, M-mode, and continuous-wave Doppler applications lie, on average, about one order of magnitude above those used in imaging (SPTA intensities around 100 mW/cm²), but there is a significant amount of overlap between the modes (5). It is important to be aware that pulsed Doppler transducers operate at fairly high time-averaged intensity levels. One survey found that more than half of the measured beams had the greatest value of SPTA in excess of 1 W/cm² (5).

Dose is a quantitative measure which combines intensity and exposure time. However, no dose quantity has been determined for ultrasound. Variation in tissue properties between individuals as well as scanning conditions influence dose in an unpredictable way. Therefore, fetal dose cannot be precisely quantified, and thus, no data exists on the dose that the fetus is exposed to in the clinical setting (2).

Bioeffect mechanisms

Heat and cavitation are the two important mechanisms by which ultrasonic energy is known to alter biological media (6). Whenever ultrasonic energy propagates through biological tissues, part of its energy is converted into heat. At the specific insonation site in tissue, there is an ultrasonically induced temperature increase. However, it is clear that diagnostic ultrasound equipment as used today for simple B-mode imaging operates at acoustic outputs that are not capable of producing harmful temperatures (7). On the other hand, experiments using unperfused tissue have demonstrated that some Doppler diagnostic equipment has the potential to produce biologically significant temperature rises, specifically at interfaces between bone and soft tissue (7).

Most attention in the safety debate has focused on cavitation rather than on heating. Cavitation is the development of bubbles (cavities) in a medium exposed to ultrasound. As waves of positive and negative pressure within the ultrasound field pass a bubble, they cause
the bubble to expand and contract. These cavitation events produce mechanical forces in their vicinity which may be sufficient to disorganize biological materials. Therefore, cavitational effects in the cells cannot be ruled out as mechanisms of biological effects under the conditions of diagnostic ultrasound, either B-mode, M-mode or for Doppler devices (8).

Theoretical considerations

Abnormal development is typically related to damage at particular, critical stages of gestation. Routine ultrasonography is usually done at weeks 16–22 of pregnancy and, theoretically, the neuronal migration taking place in the fetal brain during this time period might be disturbed by ultrasound (9). From a theoretical point of view, possible adverse effects of ultrasound may take place in the fetal forebrain, the relay system between the retina and the visual cortex, the cochlea of the inner ear, and the fetal ovaries, since these are possible “target organs” for damage during this particular period of gestation (9).

Studies in vitro have shown changes in the cell membrane after exposure to therapeutic ultrasonic intensities (10), and in cell surface motility and architecture after exposure to a diagnostic ultrasound device (11). It is conceivable that ultrasound might influence neuronal migration, and it has been suggested that altered cerebral dominance, dyslexia, or impaired neurological development may be the result of a disturbed migration of neurons (12). The dominant hand may serve as an indicator of cerebral dominance. Satz (13) has postulated that handedness may change if the hemisphere dominant for handedness suffers early damage (before the age of 2 years). The normal high prevalence of right-handedness means that random damage to the hemispheres will be expected to increase left-handedness. Orsini and Satz (14) noted that signs of pathological left-handedness were seen in hemiplegic patients with very subtle motor impairment. Similarly, Carlsson and co-workers (15) found increased incidence of left-handedness in a sample of children whose left hemisphere was impaired. Thus, a shift towards non-right handedness may occur even when there are no obvious signs of brain damage.

Experimental studies

Biological effects of ultrasound demonstrated in experimental studies can be classified as morphological or functional alterations. A morphological alteration is usually permanent or irreversible, and is likely to be produced only by quite high levels of ultrasonic energy. Usually, there is no doubt that the alteration has taken place, but the question remains as to what might be the mechanism (heating, cavitation, or some other mechanism).
Functional alterations include biological changes that affect biochemical function or level of activity, or weight gain. Such changes are not necessarily permanent, and they may occur at low ultrasound exposure levels. One example of a functional alteration is fetal weight change. A large number of experimental studies have evaluated the effect of ultrasonic exposure on pregnant mice. Some reports have found that exposed fetuses, were smaller at the time of birth than unexposed controls (16). However, many of these studies were conducted with higher exposure levels than those of commercial diagnostic devices, and the association has been restricted to certain strains of mice while other strains have been unaffected by similar exposures (17). The data are, in general, conflicting as to whether functional alterations occur at exposure levels close to that of diagnostic ultrasound (18).

Based on all available information gained from bioeffect studies in experimental systems, the American Institute of Ultrasound in Medicine (AIUM) has stated: “In the low megahertz frequency range there have been (as of this date) no independently confirmed significant biological effects in mammalian tissues exposed to intensities (SPTA) below 100 mW/cm². Furthermore, for ultrasonic exposure times (total time) less than 500 seconds and greater than one second, such effects have not been demonstrated even at higher intensities when the product of intensity and exposure time is less than 50 joules/cm².” (8).

The US Consensus Report (2) concludes slightly differently by giving the following statements (p. 55):

- Diagnostic ultrasound may not be totally innocuous.
- Statements that ultrasound below a certain intensity or duration is safe should be challenged. Absence of enough knowledge to the contrary is not enough to warrant such statements at this time.
- It is likely that ultrasound effects, if they occur in human development, are subtle and may well be delayed in expression.

**Human studies**

**Birthweight**

The effect of ultrasound exposure on birthweight has been of particular concern, since it has been shown (16) that ultrasound exposure close to diagnostic levels may be associated with a weight reduction in fetal mice. Data on birthweight in relation to ultrasound exposure are presented in a large number of human studies (19–25). In one study, Moore and co-workers (19) found reduced birthweight among children who were exposed to ultrasound in utero. The authors concluded, however, that confounding with maternal and fetal risk factors rather than
ultrasound exposure, might explain the reduced birthweight. All other studies have reported no differences in birthweight between ultrasound exposed children and controls except a Swedish randomized controlled trial of routine ultrasonography in pregnancy which reported a significantly higher mean birthweight (but not length) among screened than among unscreened children (24). Babies in the screened group were, on average, 42 g heavier, and the difference was even greater (75 g) among mothers who had reported being smokers at their first antenatal visit. This lead the authors to suggest that the ultrasound examination could have influenced mothers in the screened group to reduce smoking and change to a healthier lifestyle more often than women who were not screened.

In conclusion, the available epidemiological data suggest that diagnostic ultrasound has no effect on birthweight per se. However, if maternal smoking is reduced by the influence of routine ultrasound in pregnancy, this may result in an increased birthweight in the offspring of screened mothers.

Only one study has assessed growth during childhood among children who were exposed to ultrasound in utero. Lyons and co-workers (26) found no differences in weight or height between the groups in a study ranging from birth and up to six years of age.

Neurological development

Only a few studies have addressed the neurological development following ultrasound exposure in pregnancy. Kohom and co-workers (27) examined EEG recordings among 20 newborns before and after ultrasound exposure, and reported no effects of exposure on the recordings. However, the EEG recordings were not read blinded to the exposure status of the infant, and therefore, the possibility of bias in the interpretation of the readings cannot be excluded. The study included a small number of neonates, and the possibility of committing a statistical type II error could also not be excluded.

Scheidt, Stanley and Bryla (21) compared 122 variables measured at birth and at one year of age in ultrasound exposed and unexposed infants. They found that a significantly higher percentage of ultrasound exposed infants had an abnormal grasp or tonic neck reflex, but no difference between the ultrasound exposed and unexposed children for any one of the other 120 variables. The biological importance of the abnormal reflexes is uncertain, and the number of abnormal infants was small. Also, a large number of statistical tests were carried out and some statistically significant differences may be due to chance alone.

Stark and co-workers examined 425 ultrasound exposed and 381 matched unexposed infants when they were 7 to 12 years of age (20). They found no association of exposure with a large
number of outcomes, including hearing, vision, and intelligence scores. However, they found a significantly greater proportion of ultrasound exposed children to be dyslexic based on the Gray Oral Reading Test ($p < 0.01$). In their analysis, numerous statistical comparisons were made, and thus, it is possible that the difference in dyslexia between the groups was due to chance. An imbalance in factors other than ultrasound that are related to dyslexia may not be adequately controlled for, and may thus contribute to the explanation of the finding. However, the general consensus is that further research on the subject is urgently needed (28).

**Childhood cancer**

Concern that children who were exposed to ultrasound in utero have an increased risk of cancer is derived from the analogy between ultrasound and high energy radiation. Exposure to diagnostic x-rays in utero causes about a two- to three-fold increase in the risk of childhood leukemia (29). Two large and carefully designed epidemiological studies of the possible association of in utero ultrasound exposure and childhood cancer have been published (30, 31), but neither study found evidence of any association between ultrasound and childhood cancer.

**Recommendations**

The European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) has a Safety Committee called "The Watchdogs". In October 1992, the Watchdog Committee reendorsed their Clinical Safety Statement (32):

> Ultrasound imaging for diagnostic purposes in obstetrics has been in extensive clinical use for more than 25 years. Numerous investigations, of various degrees of sophistication, have been undertaken in an endeavour to detect adverse effects. None of these studies has proved that ultrasound at diagnostic intensities as used today has led to any deleterious effect to the fetus or mother.

> In view of the current lack of well-designed, controlled long term prospective epidemiological studies, it is necessary to resort to evidence culled from laboratory studies in vitro and in vivo.

The Watchdog Committee then argues that the use of real-time B mode imaging is "not contraindicated by the evidence currently available". However, they are reluctant to recommend an extensive use of pulsed Doppler devices:
In view of the possibility of significant temperature elevations in utero, routine examinations of every embryo and fetus using pulsed Doppler devices is considered inadvisable at present.

As new instrumentation using higher acoustic outputs, and new clinical procedures become more widespread, thus giving the potential for higher tissue exposures, it will be necessary continually to re-evaluate the safety of these diagnostic procedures.

The NIH Consensus Conference in 1984 stated that a long term follow-up of infants in a randomized clinical trial was needed to clarify questions about the effect of ultrasound on human development (2), and a need for long-term follow-up studies has also been recommended by others (8, 33).
Study objectives and hypotheses

The main objective of the present study was to investigate possible long-term effects in childhood of a screening program with diagnostic ultrasound in pregnancy. It is important to emphasize that the study does not evaluate possible long-term effects of ultrasound exposure in pregnancy, but rather the effects of routine use as opposed to selective use of ultrasound examination in pregnancy.

Six hypotheses were stated in the study protocol, which was submitted to the National Perinatal Epidemiology Unit in Oxford, England, in 1988.

**Hypothesis 1**

There is no difference in the prevalence of dyslexia among screened and control children.

**Hypothesis 2**

There is no difference in anomalous cerebral dominance as measured by the prevalence of non-right-handedness or left-handedness among screened and control children.

**Hypothesis 3**

There is no difference in the prevalence of Deficits in Attention, Motor control and Perception (DAMP) between the two groups of children.

**Hypothesis 4**

There is no difference between the two groups of children with regard to hearing ability at four and seven years of age.

**Hypothesis 5**

There is no difference between the two groups in visual acuity at four and seven years of age.

**Hypothesis 6**

There is no difference between the groups of children in growth and development assessed by curves of height and weight or by the Denver Developmental Scale.
Study population

The present study is a long term follow-up of randomized cohorts of children with differential exposure to ultrasound in utero. Their mothers were randomized to a study or to a control group in two previous studies conducted in 1979-81 in Trondheim and Ålesund (23, 34).

The participants in the Trondheim trial were women whose first antenatal care visit was to the clinics of 25 general practitioners (GPs) in the Trondheim area between May, 1979, and September, 1980 (23). Women who were referred to the obstetric outpatient clinic at the Trondheim University Hospital were not included in the study. Thus, the study population in Trondheim included mostly low-risk pregnancies.

In Trondheim, 510 women were randomly allocated to ultrasound screening and 499 were controls. Among the screened women, there were 6 pairs of twins, 13 abortions, 4 perinatal deaths among singletons, and 6 women who could not be traced 8 years after the original study. In the control group, there were 4 pairs of twins, 20 abortions, 3 perinatal deaths among singletons and 8 women who could not be found. Since we restricted the study to singletons, 481 screened and 464 control children were eligible for follow-up.

In the Ålesund trial, participants were women whose first antenatal care visit was to the clinics of 35 GPs in the Ålesund area between May, 1979, and September, 1981 (34). Nearly all pregnant women in the area were enrolled in the study, including those with high-risk pregnancies who were referred to the obstetric outpatient clinic at the Ålesund Central Hospital.

The preliminary results from the Ålesund trial were reported as a letter to the editor, and the report stated that 809 women were screened and 819 were controls (34). By re-examining the data, the numbers were corrected to 825 women randomized to screening and 803 women as controls. Among the screened women, there were 9 pairs of twins, 42 abortions, 4 perinatal deaths among singletons, and 7 women who could not be traced in 1988, and in the control group there were 6 pairs of twins, 1 set of triplets, 46 abortions, 8 perinatal deaths among singletons, 3 late neonatal deaths, and 19 women lost to follow-up. Thus, 763 screened children and 720 controls were eligible for follow-up.

We pooled the two trials into one follow-up study because the study design and methods of randomization were identical. We later checked that pooling was appropriate by doing analyses stratified by center. Since the results were homogeneous between centers, the procedure of pooling was considered to be appropriate, and pooled estimates are used throughout the thesis.
The mothers of the 2,428 eligible children were contacted in 1988. Of 1,244 children in the screened group, 1,115 were studied; 5 had died (2 due to congenital heart disease and 3 from sudden unexplained death in infancy), and the parents of 124 did not respond to the questionnaire. Of the 1,184 control children, 1 had died (sudden unexplained death in infancy), and the parents of 137 did not respond to the questionnaire; this left 1,046 children to be studied. A comparison between children who were included in the study and those not available has been made in Paper I. We found no obvious differences between screened and control children for any of the collected family or social variables (see Papers III and IV).

Data from maternal and child health centers were available for 1,107 children in the screened group and for 1,033 controls. In all, 2,011 children were assessed by their teachers in the second grade of primary school. Response from the teachers was not evenly distributed between screened and control children, since 100 screened children and 50 controls were not assessed by their teachers (p = 0.0001). The possible bias introduced in the study by different response rates among teachers between screened and control children has been discussed in Paper I.

A subsample of 32% of the children was selected to be examined with specific tests for dyslexia. The selection criteria were year of birth (1980) and residence in the vicinity of Ålesund or Trondheim in 1989-90. In all, 603 children were examined, and a comparison between those children and all 2,161 children has been made in Paper II. The comparison indicated that the examined children constituted a representative sample of all children even if they were not randomly selected.
Methodological considerations

The available sample size was 1,000 children in each group. Calculations of statistical power were performed before the study for five of the six study hypotheses, using a two-sided alpha of 0.05 and a beta of 0.10. The calculations showed that we would be able to detect a 50% increase in the frequency of dyslexia from a base prevalence of 7% to 11%; a 25% increase in the prevalence of non-right handedness from 15% to 21%; a 75% increase in the prevalence of DAMP from 5% to 9%; a 50% increase in the occurrence of impaired hearing from 7% to 11%, and to detect a 75% increase in the occurrence of impaired visual acuity from a prevalence of 4% to 7%.

Ultrasound exposure

In the initial studies, ultrasound examinations were offered in weeks 19 and 32 of pregnancy. Examinations were done with identical ultrasonic devices at both sites (ADR 2130, Tempe, Arizona, USA). The real-time scanners were calibrated to a sound velocity of 1540 m/s and had 3.5 MHz transducers. The manual for ADR 2130 lists output intensity levels of less than 0.2 mW/cm² (SATA) and less than 1.5 W/cm² (SPTP). Thus, the ADR scanners produced lower intensities than do most obstetrical scanners in use today.

Exposure duration for the first routine scan were recorded for 155 women in Ålesund. The median exposure was 3 minutes (range 2-10 minutes), which was 1.7 times longer than the median exposure time reported by others (35). Our exposure times represent minimum estimates, however, since the ADR-scanner emits ultrasound energy when the image is “frozen”, a fact which the operators were unaware of at the time of recording. Still, it is reasonable to believe that today's fetal ultrasound examinations have a longer duration, since the booking intervals in the randomized trials were 15 minutes, compared to the present intervals in routine ultrasound examinations in Trondheim of 30 minutes.

In the screened group, 34 children (3 %) had not been exposed to ultrasound screening in utero, 31 children (3 %) had been exposed once, and 950 children (94 %) had been exposed twice or more. In all, 935 (92 %) children had been exposed to ultrasound between weeks 16 and 22 of gestation. In the control group, 808 children (81 %) had never been exposed to ultrasound, 124 children (13 %) had been exposed once (42 in the month before birth), and 64 children (6 %) had been exposed twice or more. Only 46 of the control children (5 %) had been exposed to ultrasound between weeks 16 and 22 of gestation. The overlap in exposure between the randomized groups is small, which is important for the interpretation of the results.
Most fetuses are exposed to ultrasound from other sources, such as fetal heart rate detectors and electronic fetal monitoring. Fetal heart rate detectors emit continuous-wave ultrasound and are in common use in the second and third trimester. In a survey in 1989, 21 (42%) of 50 general practitioners (GPs) who recruited women to our study, reported having used fetal heart rate detectors in 1979-81 (36). Eighteen of the 21 GPs used their detectors on most pregnant women, and 11 usually used them more than three times during pregnancy. Only 3 of the GPs used their detectors more than 30 seconds each time. We have no indication that exposure to these additional ultrasound sources was unevenly distributed between the screened and control groups. The use of fetal heart rate detectors during pregnancy has increased over the years, and 32 (73%) of 44 GPs in Trondheim and Ålesund had such devices in 1989 (36).

The fetal ultrasound exposure of the children in the study is regarded to be low compared to today's practice, since modern equipment produces two to three times higher intensities (5); the current exposure duration is presumably longer; the use of fetal heart rate detectors has increased; and other Doppler devices have been introduced in the surveillance of high-risk pregnancies.

**Procedures**

Mothers of all the 2,428 eligible children were sent a questionnaire with an information letter and a postage paid return envelope. The questionnaire consisted of 66 questions about socio-demographic data and the child's health (hearing, vision, and neurological development). We asked questions about family history of dyslexia, left-handedness and allergy, and 21 questions on dominant hand. Return of the questionnaire was taken as informed consent for the child to take part in the follow-up study. Two reminders were sent to non-responders.

Much work was invested in tracing children eight years after the original studies. As a part of this effort we also examined the effect of a newspaper article on the response rate to a postal questionnaire, and found that such an incentive may enhance the response (37).

Norwegian children are regularly examined by physicians and specially trained public health nurses at maternal and child health centers. Visits to the centers take place when a child is six weeks, at three, six, and twelve months, and at two, four and seven years of age. The prospectively recorded data from these examinations were collected for each child in the study. The data included information on height, weight, neurological development, visual acuity, and audiograms.
In the children's second year of primary school, their teachers were asked to evaluate reading achievements (reading comprehension and oral reading), spelling, and arithmetic, and to give a general impression of the children's school performances. The teachers were also asked to report if the children needed extra teaching assistance in school. A subsample of the children was examined with specific tests for dyslexia in the third grade. The methods have been used in previous Norwegian studies of dyslexia (38, 39), and details are given in Paper II.

**Outcome variables**

**Dyslexia**

Dyslexia is a neurologically based disorder characterized by an unexpected failure to read. The condition is said to be present in 3% to 6% of school-aged children (40). It has been assumed that dyslexia represents a specific syndrome that is distinct from the normal distribution of poor readers (41), but this view has recently been challenged (42).

We wanted to assess whether routine ultrasonography offered in pregnancy was associated with dyslexia among primary school children. The methods are presented and discussed in Papers I and II. We have assessed the school performances of 2,011 children by a teacher rating scale, and examined 603 children with specific tests for dyslexia. The general question of whether a teacher rating scale can measure dyslexia has been discussed in Papers I and II.

Dyslexia is usually expressed as a discrepancy between the level of reading achievement predicted on the basis of intelligence and the actual level of reading achievement. In the European research tradition, spelling has been incorporated in the definition of dyslexia, but today, most investigators in the US define dyslexia on the basis of ability to decode single real or nonsense words. We have used slightly different definitions of dyslexia in Paper I and II. In Paper I the definition of dyslexia was based on a discrepancy score between two reading tests, one spelling test, and an intelligence score, and 47 (8%) of the children were classified as dyslexic. In Paper II the definition of dyslexia was based on a discrepancy score between a reading speed test score and an intelligence score, and 26 (4%) children were classified as dyslexic. Among these 26, 21 (81%) were classified as dyslexic also by the definition used in Paper I. The Kappa coefficient (see Paper II) between the two definitions of dyslexia was 0.6, which indicates a fairly good agreement between the definitions (43).

Since the results presented in Paper II are important for the validation of a teacher rating scale as a measure of dyslexia, we also used the definition of dyslexia in Paper II in an analysis of the association between routine ultrasound and dyslexia. Thus, we classified 13 (4%) of 309 screened children as dyslexic, compared to 13 (4%) of 294 control children (odds ratio = 0.95, 95% c.i.: 0.44, 2.06). The results are similar to those presented in Paper I.
Handedness

The dominant hand of the child was assessed with 21 questions taken from a modified version of a questionnaire developed by Rackzowski and co-workers (44). The parents answered specific questions about which hand the child preferred to use while performing various tasks in daily life activities. They were instructed not to respond if they had never observed the child do the task in question. Response options were the left hand, equally often with either hand, or the right hand.

We included ten of 21 handedness questions in the analysis. The ten questions were chosen because they represented various aspects of daily life activities of a child, and received a relatively high response rate. Ideally, the ten questions should have been selected prior to the study, to avoid a possible bias between screened and control children due to the selection of questions. This methodological problem is discussed in Paper IV.

Deficits in attention, motor control and perception (DAMP)

The term “Deficits in attention, motor control and perception” (DAMP) has been recommended to replace the initial description; minimal brain dysfunction (MBD). This may be a sign of impaired neurological development which is first detectable in pre-school children. It has been shown that DAMP has a prevalence of 7% among pre-school children in Sweden (45). Five questions from a Swedish questionnaire that was specifically developed to identify children with DAMP (46), were used to assess DAMP in the present study.

We classified 15% of the children as having DAMP, and the observed prevalence was clearly above an assumed base prevalence of 7%. Thus, the instrument may be inaccurate for measuring this condition among Norwegian children. In a study of Swedish pre-school children, the reported specificity was 92%, which indicates that the false positive rate may be rather high (46). The consequences of a possible misclassification of DAMP among the children have been discussed in Paper IV.

Vision and hearing

Two hypotheses addressed any associations between ultrasound and impaired vision or hearing. We used data from the parental questionnaire and from visual acuity tests and pure tone audiometry at the age of seven years to assess vision and hearing. The methods are described and discussed in Paper III. High-frequency audiometry (threshold audiometry at frequencies above 8 kHz) is not routine in the assessment of child hearing, and was therefore not done in the present study. However, since ultrasound consists of high-frequency
mechanical waves above 20 kHz, and the cochlea of the inner ear is a postulated target organ for adverse effects of ultrasound, we may have missed possible adverse effects in the higher frequencies of hearing.

Growth and development

Neurological development during the first year of life was assessed by a modified version of the Denver Developmental Scale (47) which has been in common use at maternal and child health centers for the last 20 years. Growth was assessed by the recorded height and weight data from examinations at the health centers at three, six, and twelve months, and at two, four, and seven years of age. The methods used are described and discussed in Papers IV and V.

Blinding

Blinding was considered to be important because knowledge of ultrasound exposure might influence the measures of outcome (48). None of the teachers, public health nurses, or the examiners of dyslexia were aware of the children's exposure status when the examinations took place. However, parents could not be blinded to the children's exposure status. Thus, the outcome measures addressed solely by the questionnaire, such as handedness and DAMP, may therefore be biased by the knowledge of exposure status.

Statistical analysis

Analyses were done with SPSS (Statistical Package for Social Sciences). Proportions of missing data between groups were compared using the Mantel-Haenshel chi-square statistics. We used non-parametric Mann-Whitney tests to compare teacher evaluations and handedness scores, and Student's t-tests to compare intelligence and language test results between groups. The associations between routine ultrasonography and subsequent non-right handedness, DAMP, impaired neurological development, vision, or hearing were assessed by using the odds ratio as a measure of relative risk. The precision of the odds ratio is presented as 95% confidence intervals, calculated from Mantel-Haenshel chi-square statistics (49).

A repeated measures analysis of variance which employs the least square means method by use of the GLM procedure in SAS (50), was used to model growth curves in a subsample of children (see Paper V). The evaluation of teacher rating scales (see Paper II) was done with ROC curves and Kappa indices (43).
Main results

Paper I

We found no association between routine ultrasonography in utero and teacher-reported school performance of children in the second year of primary school. Furthermore, there was no statistically significant difference in the reported need for extra teaching assistance in school between screened and control children.

The third grade test results classified 21 (7%) of the 309 screened children as dyslexic, compared with 26 (9%) of the 294 control children (odds ratio = 0.75, 95% c.i.; 0.41, 1.37). Among screened children, 19 (6%) had low achievement in reading or spelling, compared with 23 (8%) in the control group (odds ratio = 0.77, 95% c.i.; 0.41, 1.45).

Paper II

Teacher rating scales correlated well with tests performed one year later. The best cut-off point of the rating scale to separate children with low achievement or dyslexia from normally achieving children was a score value ≤ 3. Using this score value to identify children with low reading achievement, teachers were fairly accurate (sensitivity of 89% and specificity of 88%). In screening for dyslexia, which was defined by discrepancy between the reading speed of single words and ability-predicted achievement, the assessment of oral reading was somewhat less efficient (sensitivity of 72% and specificity of 84%).

Paper III

There was no significant difference in impairment of sensory functions between the groups of children as assessed by the parental questionnaire. According to the questionnaire, 68 (6%) of 1,113 screened children and 72 (7%) of 1,039 control children had abnormal vision (odds ratio = 0.87, 95% c.i.; 0.61, 1.24). Parents also reported that 32 (3%) of 1,112 screened and 38 (4%) of 1,041 control children had abnormal hearing (odds ratio = 0.78, 95% c.i.; 0.48, 1.26).

The visual acuity tests at the age of seven years, classified 31 (3%) of 1,023 screened and 34 (4%) of 948 control children as having impaired visual acuity (odds ratio = 0.84, 95% c.i.; 0.51, 1.38). Also, 77 (7%) of 1,042 screened children had reduced hearing according to the pure tone audiometry at age seven years, compared to 69 (7%) of 966 controls (odds ratio = 1.04, 95% c.i.; 0.72, 1.49). Among the children with reduced hearing, 2 of 1,042 screened children and 7 of 966 controls were deaf or dependent on a hearing aid (p = 0.07).
Paper IV

With the use of ten questionnaire items, we classified 162 (19%) of 861 screened children as non-right handed, compared to 120 (15%) of 802 controls (odds ratio = 1.32, 95% c.i.; 1.02, 1.71). Of these, 62 (7%) children in the screened group were classified as left-handed compared to 44 (5%) control children (odds ratio = 1.34, 95% c.i.; 0.90, 2.00).

A total of 1,654 children could be included in the analyses of impaired neurological development during the first year of life. According to the short version of the Denver Developmental Screening Test 75 (9%) of 859 children in the screened group and 73 (9%) of 798 children in the control group had impaired neurological development (odds ratio = 0.95, 95% c.i.; 0.68, 1.33). In all, 2,128 mothers reported the age when their child started to walk. The mean age for walking was twelve months in both groups.

A total of 2,100 children were included in the analyses of DAMP. Of the 1,081 children in the screened group, 147 (14%) were classified as having DAMP, compared to 163 (16%) of 1,019 control children (odds ratio = 0.83, 95% c.i.; 0.66, 1.05).

Paper V

We found no statistically significant differences in a cross-sectional analysis of body weight or height at birth; at three, six, and twelve months; and at two, four, and seven years of age.

In all, 39% of the screened and 36% of the control women reported smoking at their first antenatal visit (p = 0.07). Babies of screened women who smoked were on average 39 g heavier at birth than babies of control women who also smoked. Moreover, a weight difference in favor of the screened group was present at each subsequent visit throughout the follow-up up to and including age seven. Among non-smokers, babies of screened women were, on average, 23 g heavier at birth than babies of controls. Throughout the follow-up there was no systematic difference in weight between the groups. Weight differences appeared to fluctuate around no difference between screened and control children in an unsystematic fashion.

In a repeated measures analysis of variance in a subsample of children, we found a statistically significant difference in growth curves (p = 0.02) between screened and control children among mothers who reported smoking at their first antenatal visit. Contrary to this, there was no such difference in growth among children of non-smoking mothers (p = 0.7).
Discussion

The discussion of results will address four areas in which controversies in epidemiologic research tend to concentrate (51):

1. Can the results be attributed to chance, and could adverse effects of ultrasonography have been overlooked?
2. Can bias in the selection of study subjects, or bias in the information from the participants be responsible for the results?
3. Can the results be due to measured or unmeasured confounding factors?
4. How do the results relate to general criteria of causality?

Chance

In this randomized controlled follow-up study, we found an association between routine ultrasonography in utero and subsequent non-right handedness among children in primary school. The study question on handedness was one of six initially specified hypotheses, which indicates that the probability of one or more of them being significant (p < 0.05) in the predicted direction by chance is about 1 in 7 (1 - 0.9756 = 0.14). Thus, the association (p = 0.04) between ultrasonography and non-right handedness could well be due to chance alone.

However, the above calculation depends on the assumption that all six hypotheses were equally sensitive indices of subtle changes in the development of the brain or parts of the brain. Many neuropsychologists would argue that non-right handedness was in fact the most sensitive index of a change in brain development among the hypotheses in our study. Thus, it may be appropriate to attribute more weight to this hypothesis, and consequently, less appropriate to consider the association between ultrasound and non-right handedness as a pure chance finding.

We found no association between routine ultrasonography in utero and the remaining five hypotheses in the study. The possibility of doing a type II error has been assessed through the power calculations before the study. In the computations we used a beta-value of 0.1 which is more conservative than the conventionally used beta-value of 0.2. In the present study, we should have been able to detect a 50% increase in the occurrence of dyslexia, a 75% increase of DAMP, a 50% increase of impaired hearing, and a 75% increase of impaired vision. However, the issues of dose-response, confounding, misclassification, and interaction are generally ignored in power calculations, even though such issues can greatly affect the informativeness of a study, rendering the judgements based on the power calculations
superficial (52). Also, the relatively wide confidence intervals around some of the observed odds ratios in our study indicate that a 30% increase in some of the outcome measures may easily have been over-looked. Thus, even a fairly large study like this one has insufficient power to detect small differences in outcome between screened and control children.

**Internal validity**

The validity of a study is usually separated into two components: the validity of the inferences drawn as they pertain to the actual subjects in the study (internal validity); and the validity of the inferences as they pertain to people outside the study population (external validity or generalizability). Internal validity is clearly a prerequisite for external validity.

Various types of biases can detract from internal validity, and three potential sources of bias should always be considered in the interpretation of epidemiologic studies (49, 52). These include bias in the selection of study subjects; information or misclassification bias of the variable characteristics of study subjects; and bias due to confounding between the exposure and other measured or unmeasured covariables. A strong feature of the present study is that a randomized controlled design may reduce the likelihood of bias which may influence the internal validity of the study.

**Selection bias**

Missing data may introduce bias in a randomized trial if non-response is unevenly distributed between study and control group, and the reasons for non-response are associated with the outcome under study. One possible source of bias in our study was introduced by the uneven response from 150 primary-school teachers. If teachers find it difficult to assess children with dyslexia, the uneven response between screened and control children from the teachers may have biased the results. We have discussed this possible bias in Paper I, and given calculations for the worst possible case, to estimate how such a bias may have influenced the results.

**Misclassification bias**

Blinding is one important method of ensuring that classification of the outcome measures is not influenced by the knowledge of exposure status. As was pointed out in the methods section, blinding was successfully done for all outcome measures except for handedness and DAMP. Even if handedness was not assessed blinded to exposure status, we do not find it likely that the classification of non-right handedness, which was based on ten out of 21 questions, was influenced by the mother's knowledge of screening group status. No previous study has ever indicated a relation between ultrasound exposure in utero and handedness of
the child, and the mothers were not aware of the study hypotheses. However, in future replications, authors must consider the possibility of misclassification bias in the assessment of an association between ultrasound and non-right handedness.

The potential for misclassifying children because of imprecise measurements may always be a threat to the validity of the results. However, if the outcome measures in our study were subject to misclassification due to imprecision, it seems unlikely that the bias is differential, depending on ultrasound exposure. Instead, a nondifferential misclassification may be expected. The bias of a nondifferential misclassification will always attenuate any estimated association and produce a more conservative result, which will be biased towards the null hypothesis (52). Thus, poor classification of non-right handedness among the children will not invalidate our results. It is true that such misclassification can introduce a bias, but the bias will be in the direction of underestimating the effect.

In this study, the results from five of six hypotheses indicate the absence of an effect of ultrasound. It is crucial to consider the problem of nondifferential misclassification to determine to what extent a real effect might have been obscured, and we have discussed this problem in Papers III and IV. Also, the purpose of paper II was to examine the precision of teacher assessments in classification of low achievement and dyslexia. In the assessment of DAMP, an association that could have been detected with more precise instruments, may have been overlooked.

Confounding

To be confounding, an extraneous variable must have the following three characteristics (52):

1. It must be a risk factor for the disease.
2. It must be associated with the exposure under study.
3. It must not be an intermediate step in the causal path between the exposure and the disease.

Ideally in a randomized experiment, every risk factor should be evenly distributed in relation to the exposure. Nevertheless, it happens that an identified risk factor, despite the randomization, is associated with exposure. Then this risk factor may have a confounding effect in the experiment. In our study, maternal smoking in pregnancy may be considered a possible confounding factor, despite the randomization, since 39% of the screened and 36% of the control women reported smoking at their first antenatal visit (p = 0.07).
It is well known that smoking reduces birthweight (53, 54), and it has been shown (55) that a mother is strongly affected by watching her fetus on the screen. In fact, it has been suggested that the procedure may stimulate changes to a healthier lifestyle (56). Waldenström and co-workers found mean birthweight to be 42 g higher ($p = 0.008$) and that fewer babies had a birthweight below 2,500 g (2.5% vs 4.0%, $p = 0.005$) among screened newborns in their randomized controlled trial (24). The difference was even greater (75 g) among mothers who had reported to be smokers at their first antenatal visit. The authors hypothesized that the experience of the ultrasound examination might have induced mothers in the study group to reduce smoking and change to a healthier lifestyle more often than unscreened women (24). Thus, smoking, or in particular smoking cessation, may be associated with ultrasound exposure, and through this, with birthweight. We have dealt with maternal smoking as a possible confounding factor in Paper V, where the results have been presented as stratum specific estimates, separating smokers and non-smokers.

**External validity**

The study population of the Trondheim study tended towards including low-risk pregnancies (23), whereas the study population in Ålesund was representative of the general population of pregnant women in this area (34). We assume that the study results may be generalizable to children exposed to routine ultrasound in utero a decade ago. As pointed out earlier, the exposure levels of ultrasound have increased over the years. We have not assessed any dose-response relation between ultrasound and the outcome measures. Thus, it is difficult to know whether or not the study results may be generalizable to today's obstetrical practice.

**Causality**

If results of a study cannot be attributed to chance or to problems in study design resulting in any one of the major sources of bias, the results should be evaluated in light of being causal associations. Bradford Hill (57) has provided a list of criteria which characterize causal relationships. Rothman (52) has evaluated these criteria, and found most of them helpful in assessment of statistical associations. In the following, an attempt will be made to relate the association between ultrasound and non-right handedness to the most important of Bradford Hill's criteria of causality.
Strength of the association

The stronger the association is between an exposure factor and the outcome under study, the more likely the relation is to be causal. We found a weak association between ultrasound and non-right handedness, and this does not argue strongly in favour of a causal relationship.

Consistency

A causal association should be consistent with previous research. No previous study has indicated a relation between ultrasound and non-right handedness. However, the criterion of consistency may also be used to judge the results of an isolated study through a repeated observation of the association among subgroups (51). We found that a positive association between ultrasound screening and non-right handedness was present for 20 of the 21 items in the handedness questionnaire. This indicates a within-study consistency. The weakness of within-study consistency derives from the possibility that an undetected error could be so pervasive that it invalidates the findings altogether (51).

Specificity

Specificity means that exposure is linked to a particular form of disease or to the particular site at which it occurs. However, while specificity is often regarded as supporting a causal interpretation, lack of specificity cannot negate one. In this thesis an association was found between ultrasound and non-right handedness, but not for the other five hypotheses under study.

Relationship in time

It is an absolute criterion that the exposure must precede the outcome. The prospective design of a randomized, controlled trial indicates that this point is satisfied in our study.

Biological gradient

A dose-response relationship between the exposure factor and the outcome measure(s) would strengthen the evidence for a causal relationship, although a causal association does not need to take the form of a dose-response curve. The absence of a dose-response relationship, due to a threshold effect, can occur in cause-effect relationships.

A dose-response relationship was not assessed in this thesis, because such an analysis would require that the randomized groups be divided into different levels of ultrasound exposure,
and would further complicate the interpretation of any relation between ultrasound exposure and handedness of the child.

**Biological plausibility**

Biological plausibility is probably the criterion that is used most often, either to dismiss some unexpected finding or to support an association from a study based on suspect methods (51). Although biological plausibility is an intellectually appealing and logically sound criterion, it is clearly limited by existing knowledge and adherence to a particular theory.

The current understanding of biological effects caused by diagnostic ultrasonic devices suggests that any harmful effects on the fetus are unlikely. Nevertheless, we should not yet exclude the possibility that ultrasound has some effect on the biological development of the fetus. Since most ultrasound examinations are done at weeks 16–22, when the fetal brain is developing rapidly, effects on neuronal migration are possible. Increased incidence of non-right handed-ness may be a sensitive index of subtle changes in the development of the brain.

**Experimental evidence**

Certainly one would prefer that experimental evidence support findings based on observational methods. However, absence of experimental support, does not, in itself, weaken an inference of cause and effect (51). The fact that we have done the first experimental study considering a possible association between ultrasound in utero and non-right handedness among children excludes the possibility of relating our findings to other experimental evidence.

**Judgment of causality**

The Bradford Hill criteria may be helpful, but in the end, the interpretation of the association between postulated risk factors and disease must be left to judgment. Davey Smith and colleagues revisit this fundamental epidemiological issue considering a possible association between smoking and suicide, and they ask “How are we to judge whether we are being led down another blind alley?” (58).

In a reply to Smith and colleagues, Bhopal (59) has compared the search for causal associations with that of searching for gold: “the nuggets are rare, the specks indicative of nuggets are commoner, but mostly we sift dirt”. He also has proposed a way to see if a blind alley lurks. His advice is to carefully consider whether the hypothesis and the rationale for expecting it to be true (plausibility) were stated in advance. Then one must ask if the
association could be an artifact, or if there are any indications of a non-causal pathway. He also recommends the application of the Bradford Hill causality criteria. Finally, “in a debate with colleagues, a cautious, considered judgment should be made about whether we have a gold nugget, gold flake, fool’s gold or dirt” (59). We must await a forthcoming debate and future research to pinpoint which of the four classifications best applies to the association between ultrasound and non-right handedness. In the mean time, it may be wise to remember the advice of workers in fireworks factories: “It is better to curse the darkness, than to light the wrong candle”.

**Further research**

We would like to emphasize the need for replication of the association between ultrasound and non-right handedness before it is interpreted as more than a chance finding. A follow-up of the children from the Swedish randomized controlled trial of ultrasonography in pregnancy (24) may help clarify this issue. Plans are currently being made to undertake a follow-up of these children when they reach primary school age (Ove Axelsson, personal communication).

We will also advocate the need for more epidemiological studies of ultrasound safety in obstetrics. Observational studies of long-term effects of pulsed Doppler ultrasound will be important in the future, since it will be almost impossible to do new follow-up studies using a randomized controlled design, because most women assigned to a control group will be exposed to ultrasound some way or another during their pregnancy.
Conclusions

Paper I

Routine ultrasonography offered in weeks 19 and 32 of pregnancy did not lower teacher-reported school performances among children aged 8 or 9 years. There was no evidence of an increased prevalence of dyslexia among children whose mothers were offered routine ultrasonography in pregnancy.

Paper II

A teacher rating scale may be a feasible method to screen for learning disabilities among second grade primary school children.

Paper III

Routine ultrasonography in utero had no adverse effects on sensory functions of those children who were followed up to primary school age.

Paper IV

We found no association between ultrasound and impaired neurological development. However, we found a possible association between ultrasound and subsequent non-right handedness in the children. This may indicate that we should not yet exclude the possibility that ultrasonographic examinations in pregnancy have biological effects on fetal development.

Paper V

We found no statistically significant differences in mean body weight or height between screened and control children in a cross-sectional analysis of growth during childhood.
References


Errata

Paper I

Methods section Name of ultrasonic device is ADR 2130 (not ADR 3120)

Ref 9. Published in volume 2: 207 – 11 (not 1: 1179 – 82)

Paper III

Methods section Name of ultrasonic device is ADR 2130 (not ADR 3120)

Paper IV

Table I Years of education (mother) should be grouped 6 – 9, 10 – 12, > 12 (not 6 – 9, 9 – 12, > 12)
Paper I

Routine ultrasonography in utero and school performance at age 8–9 years

K. Å. Salvesen  L. S. Bakketeig  S. H. Eik-Nes  J. O. Undheim  O. Økland

Most fetuses in developed countries are exposed in utero to diagnostic ultrasound examination. Many pregnant women express concern about whether the procedure harms the fetus. Since most routine ultrasound examinations are done at weeks 16–22, when the fetal brain is developing rapidly, effects on neuronal migration are possible. We have sought an association between routine ultrasonography in utero and reading and writing skills among children in primary school.

At the age of 8 or 9 years, children of women who had taken part in two randomised, controlled trials of routine ultrasonography during pregnancy were followed-up. The women had attended the clinics of 60 general practitioners in central Norway during 1979–81. The analysis of outcome was by intention to treat: 92% of the “screened” group had been exposed to ultrasound screening at weeks 16–22, and 95% of controls had not been so exposed, but there was some overlap. 2428 singletons were eligible for follow-up, and the school performance of 2011 children (83%) was assessed by their teachers on a scale of 1–7; the teachers were unaware of ultrasound exposure status. A subgroup of 603 children underwent specific tests for dyslexia. There were no statistically significant differences between screened children and controls in reading, spelling, and intelligence scores, or in discrepancy scores between intelligence and reading or spelling. The test results classified 21 of the 309 screened children (7% [95% confidence interval 3–10%]) and 26 of the 294 controls (9% [4–12%]) as dyslexic.

The risk of having poor skills in reading and writing was no greater for children whose mothers had been offered routine ultrasonography than for those whose mothers had not been offered the procedure.

Introduction

Because the indications for diagnostic ultrasound in pregnancy are common, and because ultrasonographic screening is routine in many countries, most pregnant women in developed countries are exposed to the procedure. During ultrasonographic examinations, pregnant women commonly express concern that the procedure may harm the fetus. Ultrasound is used therapeutically to disintegrate renal calculi and in surgery of the brain, which implies that

ADDRESSES: Department of Community Medicine and General Practice (K. Å. Salvesen, MD) and Department of Psychology (Prof J. O. Undheim, PhD), University of Trondheim; Department of Epidemiology, National Institute of Public Health, Oslo (Prof L. S. Bakketeig, MD); National Centre for Fetal Diagnosis and Therapy, Trondheim University Hospital (Prof S. H. Eik-Nes, MD); and Department of Pediatrics, Ålesund Central Hospital, Norway (O. Økland, MD). Correspondence to Dr K. Å. Salvesen, University of Trondheim, Department of Community Medicine and General Practice, Medical Technical Centre, N-7005 Trondheim, Norway.
it has the potential to damage biological tissue. No adverse effects of diagnostic ultrasound in pregnancy have been reported. However, possible long-term effects among children exposed to ultrasound in utero have been examined in only a few studies. The general consensus is that further research on this topic is warranted. Abnormal development is typically related to damage at particular, critical stages of gestation. Routine ultrasonography is usually done at weeks 16–22 of pregnancy, which is an important phase of brain development. At this point, neurons migrate towards their destination in the fetal brain. Studies in vitro have shown changes in the cell membrane after exposure to therapeutic ultrasound and in cell surface motility and architecture after exposure to a diagnostic ultrasound device. It is conceivable that ultrasound might influence neuronal migration, and it has been suggested that dyslexia may be a result of disturbed migration of neurons. A study of children who had been exposed to diagnostic ultrasound found no biologically significant differences in physical and mental development or cognitive, behavioural, or neurological function between them and matched controls at age 7–12 years; there were, however, more dyslexic children in the exposed group.

Long-term follow-up of children in randomised clinical trials has been recommended to answer questions about the effect of ultrasound on human development. We report here on the children of women who took part in two Norwegian randomised trials of ultrasound screening in pregnancy. We aimed to find out whether routine ultrasonography in utero is associated with dyslexia among children at primary school.

Subjects and methods

Two randomised, controlled trials of ultrasonic screening in pregnancy were carried out in Trondheim and Alesund, Norway, in 1979–81. The study design and methods of randomisation were identical. The participants in the Trondheim trial were women whose first antenatal care visits were to the clinics of 35 general practitioners (GPs) in the Trondheim area between May, 1979, and September, 1981. Women who were referred to the obstetric outpatient clinic at the Trondheim University Hospital were not included in the study. In the Alesund trial, participants were women whose first antenatal care visits were to the clinics of 35 GPs in the Alesund area between May, 1979, and September, 1981; nearly all pregnant women in the area were enrolled in the study, including those with high-risk pregnancies who were referred to the obstetric outpatient clinic at the Alesund Central Hospital.

The study women were offered ultrasonicographic examinations in weeks 19 and 32 of pregnancy. The same ultrasonic devices (ADR 3120, Tempe, Arizona, USA) were used in Trondheim and Alesund. The real-time scanners were calibrated to a sound velocity of 1540 m s⁻¹ and had 3.5 MHz transducers. Exposure times for the first routine scan were recorded for 155 women in Alesund. These exposure times represent minimum estimates, however, since the ADR-scanner emits ultrasound energy when the image is "frozen", a fact that the operators did not know when they recorded the exposure times.

In Trondheim, 510 women were randomly allocated screening and 499 were controls. Among the screened women, there were 6 pairs of twins, 13 abortions, and 4 perinatal deaths among singletons, and 6 women could not be traced 8 years after the original study. In the control group, there were 4 pairs of twins, 20 abortions, 5 perinatal deaths among singletons, and 9 women who could not be found. Thus, there were 481 screened children and 464 control children eligible for our follow-up study.

The report of the Alesund trial stated that 899 women were screened and 819 were controls. However, we found that the correct numbers were 825 women randomised to be screened and 763 screened children and 720 controls were eligible for our study.

We wanted to test a null hypothesis of no difference in the frequency of dyslexia between ultrasound-screened children and their controls. With a sample size of more than 1000 children in each group, and a two-sided alpha of 0.05 and a beta of 0.10, power calculations before the study showed that we should be able to detect a 5% increase in the frequency of dyslexia (from a base prevalence of 7–11%). Mothers of all the 2428 eligible children were sent a questionnaire with an information letter and a postage-paid return envelope. The questionnaire consisted of 66 questions about sociodemographic data and the child's health (hearing, vision, and development). We asked 21 questions on dominant hand and 4 on immune disorders, including allergies among the children and their relatives, because of the commonly reported association of learning difficulties, left-handedness, and (auto)immune disorders. Return of the questionnaire was taken as informed consent for the child to take part in the follow-up study. Two reminders were sent to non-responders.

During the children's second year of primary school (age 8 or 9 years), their teachers were asked to assess reading achievement (reading comprehension and oral reading), spelling, and arithmetic, and to give a score for the child's overall school performance. The teachers were not aware which children had been exposed to ultrasound in utero. The assessments used a scale of 1 to 7 (1 = very poor performance; 4 = medium; 7 = very high). The teachers were also asked to report whether the children needed extra teaching assistance in school. The assessment forms had previously been used in Norwegian studies of dyslexia.

Teacher assessments may not be an adequate measure of dyslexia. To validate their use, a subsample of 32% of the children was selected for examination with specific tests for dyslexia during their third year of primary school (age 9 or 10 years). The selection criteria were year of birth (1980) and residence in the vicinity of Alesund or Trondheim in 1989–90. 296 children in Trondheim were examined by K. A. S. and 307 children in Alesund by K. A. S. in cooperation with four primary-school teachers who had a special interest in dyslexia. None of the examiners was aware of the children's exposure status when the examinations took place. The

**Table I**—Comparison of children in study with those not available

<table>
<thead>
<tr>
<th></th>
<th>Mother lost to follow-up</th>
<th>Non-responders to questionnaire assessment</th>
<th>No teacher at follow-up</th>
<th>Final study group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean maternal age (yr) at pregnancy</td>
<td>26 (n = 40)</td>
<td>26 (n = 261)</td>
<td>25 (n = 150)</td>
<td>26 (n = 201)</td>
</tr>
<tr>
<td>% of mothers with education of 6–9 yr</td>
<td>38 (0.4)</td>
<td>46 (0.8)</td>
<td>38 (0.8)</td>
<td>39 (0.8)</td>
</tr>
<tr>
<td>9–12 yr</td>
<td>47 (0.4)</td>
<td>41 (0.8)</td>
<td>49 (0.8)</td>
<td>48 (0.8)</td>
</tr>
<tr>
<td>% of non-smoking mothers</td>
<td>76 (0.8)</td>
<td>54 (0.8)</td>
<td>65 (0.8)</td>
<td>63 (0.8)</td>
</tr>
<tr>
<td>Mean (SD) no of ultrasound examinations</td>
<td>13 (1.6)</td>
<td>20 (1.0)</td>
<td>23 (1.8)</td>
<td>23 (1.9)</td>
</tr>
<tr>
<td>% of children male</td>
<td>43 (0.5)</td>
<td>55 (0.4)</td>
<td>61 (0.5)</td>
<td>60 (0.5)</td>
</tr>
<tr>
<td>% of children left-handed</td>
<td>13 (0.9)</td>
<td>13 (0.9)</td>
<td>13 (0.9)</td>
<td>13 (0.9)</td>
</tr>
<tr>
<td>% with family history of dyslexia</td>
<td>23 (1.3)</td>
<td>23 (1.3)</td>
<td>23 (1.3)</td>
<td>23 (1.3)</td>
</tr>
<tr>
<td>% of children with history of allergies</td>
<td>14 (2.1)</td>
<td>14 (2.1)</td>
<td>14 (2.1)</td>
<td>14 (2.1)</td>
</tr>
</tbody>
</table>

803 women to be controls. The Alesund data are being reanalysed and the results will be reported elsewhere. Among the screened women in that study, there were 9 pairs of twins, 42 abortions, 4 perinatal deaths among singletons, and 7 women who could not be traced in 1988, and in the control group there were 4 pairs of twins, 1 set of triplets, 46 abortions, 8 perinatal deaths among singletons, 3 late neonatal deaths, and 18 women lost to follow-up. Thus, 763 screened children and 720 controls were eligible for our study.
children were examined in groups of 2-10 in their own schools with tests for intelligence level, reading comprehension, reading speed, and spelling. 13,14 The examinations lasted 1.5 h. The children were classified into three groups: "normal", "poor reading or spelling", or "dyslexic". The poor reading or spelling group was defined as the 5% of tested children with the lowest scores in reading and the 5% of tested children with the lowest scores in spelling, irrespective of their intelligence scores. Thus, a total of 7% of children in the test sample were put in this group.

A discrepancy score between the observed reading score and a predicted reading score estimated from the intelligence score was computed for each child in the test sample (discrepancy reading intelligence x Z-score reading intellige11ce). A similar discrepancy score for spelling (DSI) was calculated. The 5% of children with the most negative DR! and the 5% with the most negative DSI formed the dyslexic group (8% of the test sample). These children were characterised by underachievement in reading or spelling compared with the prediction from their intelligence scores.

Analyses were done with SPSS (Statistical Package for Social Sciences). We used the non-parametric Mann-Whitney test to compare teacher assessments between groups, and Student’s t test to compare results between groups in the test sample.

Results

Of 1244 children in the screened group, 1015 were studied; 5 had died (2 congenital heart disease and 3 sudden unexpected deaths in infancy); the parents of 124 did not respond to the questionnaire; and the teachers of 100 did not respond. Of the 1184 control children, 1 had died (sudden unexpected death in infancy), the parents of 137 did not respond to the questionnaire, and 50 were not assessed by their teachers, which left 996 children to be studied.

In the screened group, 34 children (3%) had not been exposed to ultrasound screening in utero, 31 children (3%) had been exposed once, and 950 children (94%) had been exposed twice or more. 935 (92%) children had been exposed to ultrasound between weeks 16 and 22 of gestation. In the control group, 808 children (81%) had never been exposed to ultrasound, 124 children (13%) had been exposed once (42 in the month before birth), and 64 children (6%) had been exposed twice or more. Only 46 of the control children (5%) had been exposed to ultrasound between weeks 16 and 22 of gestation.

The rate of non-response by parents to the questionnaire did not differ significantly between the screened and control groups, and non-responders were similar to responders in maternal age and level of education, maternal smoking status, ultrasound exposure status in pregnancy, and sex of the child (table I).

The percentages not evaluated by their teachers differed significantly between the screened and control groups (9% vs 5%, p = 0.0002). There were significantly more boys among the group not assessed by their teachers than among the final study group (p = 0.02) and also a slightly greater proportion were left-handed (p = 0.08). The parents of the children who were not assessed also reported a significantly higher proportion of first and second degree relatives with dyslexia (p = 0.002). Allergies were, by contrast, more prevalent among the final study group (p = 0.04).

684 teachers assessed 2011 children in the study (mean 2.9 range, 1-21), but 57% of teachers assessed only 1 child (median 1). There were no significant differences between the screened and control groups in any index of school performance (table II; p values on Mann-Whitney test ranged from 0.08 to 0.40) or in the reported need for extra teaching assistance in school (8.5% screened vs 8.7% controls).

In the validation study, 682 children were offered specific tests for dyslexia. 30 children had moved or refused the examination, 22 did not attend for the examinations, and 27 did not complete all the test procedures. Thus, 603 children had complete test results. There were no significant differences between screened children and controls (table III). The test results classified 21 (7%); 95% confidence interval (CI) 3-10%) of the 309 screened children as dyslectic, compared with 26 (9% [4-12%]) of the 294 control children. 19 screened children (6% [2-9%]) were classified as poor readers or spellers compared with 23 (8% [3-12%]) of the control group.

Discussion

Thus, we found no association between routine ultrasonography in utero and teacher-reported school performance of children in the second and third years of primary school. If anything, the screened children performed slightly better in school than did their controls. Our findings do not support the possible association between ultrasound exposure in utero and dyslexia reported by Stark and colleagues. However, the tests for dyslexia used in the two studies were not identical and we did not evaluate long-term effects of ultrasound exposure per se, but the effects of routine as opposed to selective ultrasound examination in pregnancy; the analyses were done according to
to the intention-to-treat principle, irrespective of the children's actual ultrasound exposure in utero. Nevertheless, the design of our randomised, controlled trial rules out many of the possible biases that might have led to the association between ultrasound exposure and dyslexia reported by Stark and colleagues.7

Some of the screened children had not been exposed to ultrasound screening during pregnancy, whereas some of the controls were scanned several times. The overlap of exposure between the randomised groups is small, which is important for the interpretation of the results. However, we acknowledge that the ability to reject the null hypothesis may be affected by protocol deviations.15

Most fetuses are exposed to ultrasound from other sources, such as fetal heart rate detectors and electronic fetal monitoring. Fetal heart rate detectors emit continuous-wave ultrasound and are in common use during the second and third trimester. In a survey in 1989, 21 (42%) of 50 GPs who recruited women to our study said that they had had fetal heart rate detectors in 1979-81.18 Of the 21 GPs used their detectors on most pregnant women, and 11 usually used them more than three times during a pregnancy. However, we have no indication that exposure to these additional ultrasound sources was unevenly distributed between the screened and control groups.

Ultrasound exposure of fetuses is difficult to quantify. The dose depends on the ultrasound exposure time and the intensity of the output of the ultrasonic device. The median exposure time in the Alesund trial was 180 s (range 120-600), which is 1.7 times longer than the median exposure time reported by Andrews and colleagues,10 and the recorded exposure times represent minimum estimates. The ADR scanners used in the two Norwegian trials produce lower intensities than do most obstetrical scanners in use nowadays. Diagnostic real-time scanners, however, produce intensities that are far below the recommended intensity limits of diagnostic devices,18 and in relation to these limits the differences in intensity between the ADR-scanners and modern devices are negligible. As regards safety, the increasing obstetric use of pulsed doppler and colour-flow doppler ultrasound devices is more of a cause for concern than is the use of real-time scanners, because their intensity levels may exceed the recommended limits. Thus, use of doppler devices may need long-term follow-up before they are introduced routinely in antenatal care.

The study objective was to test whether routine ultrasonography offered in pregnancy was associated with dyslexia among primary-school children, and power calculations before the study were based on a sample size of more than 1000 children in each group and a base prevalence of dyslexia of 7%. We have assessed the school performances among children aged 8 or 9 years. There was no evidence of an increased prevalence of dyslexia among children whose mothers were offered routine ultrasonography in pregnancy.

The follow-up study was supported by the Norwegian Research Council for Science and the Humanities Grant no 351.87001. We thank Ms Aslaug Gjerdmundsd, Ms Sissel Vinj, Ms Lis Roald, and Ms Inger Anne Parchuusen for examining children in Alesund; Ms Judith R. Johannesen for computer assistance; Ms Brit Flidav, Ms Eva Dahl, and Ms Herti Eggan for secretarial assistance; and Dr Lars Vatten and Dr Geir Jacobsen for valuable comments and suggestions during the writing of this report.

REFERENCES


Salvesen KÅ, Undheim JO.
Screening for learning disabilities with teacher rating scales.
Screening for Learning Disabilities with Teacher Rating Scales

Kjell Å. Salvesen and Johan O. Undheim

The purpose of the study was to investigate the use of teacher assessments in screening for learning disabilities. In a longitudinal study, 603 children were rated by their teachers in the second grade (age 8 to 9 years), and the ratings were correlated with examinations of reading, spelling, and intelligence in the third grade. The third-grade tests for reading, spelling, and intelligence classified children into groups with low achievement and dyslexia, and these two groups were compared with normally achieving children. The accuracy of teacher assessments, measured with correlation analysis, ROC curves, and kappa indices, showed that teachers were quite accurate in their judgment of low achievement, but somewhat less efficient in their judgment of specific reading difficulties.

During the 1960s and 1970s, major interest developed in the early identification of children with learning disabilities. Makins (1976) reported that 47% of the local education authorities in England were using some form of educational screening procedures. However, although the production of instruments designed for the purpose proliferated, the evaluation of the effectiveness of the procedures was not so prevalent (Lindsay & Wedell, 1982). Some experts in the field were skeptical of the screening procedures; in fact, Algazzine and Ysselfyde (1986) argued that because screening procedures for learning disabilities were so flawed and so limited, it was time to redirect resources toward corrective and preventive educational intervention. Thus, during the last decade, very few studies on educational screening instruments have been published.

Educational screening is most often done at school entry. Basic difficulties with screening tests include cost and time required for administration. One way around both of these problems is to make use of the teacher, and several teacher rating scales have been developed, for example, the Pupil Rating Scale (Myklebust, 1971); the Student Rating Scale (Adelman & Feshbach, 1971; for validation, see Feshbach, Adelman, & Fuller, 1974), and the Rhode Island Pupil Identification Scale (Novack, Bonaventura, & Merenda, 1973). The validity of the rating scales is at least as good as the best screening tests (Lindsay & Wedell, 1982). However, Lindsay and Wedell found that most tests and rating scales provide only a small amount of information on which they can be evaluated.

The usual method for evaluating the predictive validity of educational screening methods has been use of the correlation coefficient. However, although a highly significant relationship may be revealed, this is less important than correctly classifying individual children. Saiz and Fletcher (1979) reanalyzed the Feshbach et al. (1974) data, calculating the sensitivity and specificity of the Student Rating Scale. This type of statistical analysis of educational screening methods was advocated by Lindsay and Wedell (1982).
Specific reading difficulty, or dyslexia, is one learning disability that may be particularly difficult to identify. Dyslexia is often considered a neurologically based disorder characterized by an unexpected failure to read despite conventional instruction and adequate intelligence and sociocultural opportunity (Critchley, 1970). The condition is said to be present in 3% to 6% of school-age children (Hendee, 1989). It has been assumed that dyslexia represents a specific syndrome that is distinct from the normal distribution of poor reading (Rutter & Yule, 1975), but this view has recently been challenged (Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). Although some recent approaches to dyslexia define it as a retardation in coding of written language (Höien & Lundberg, 1991), the present study was based on the traditional discrepancy notion, examining teacher identification of such relative differences.

We have evaluated the use of a teacher rating scale in screening for low reading achievement and dyslexia via the use of receiver operating characteristic (ROC) curves and the kappa index. These statistical methods are particularly useful in evaluating screening procedures, but so far they have not been used to evaluate procedures for children with learning disabilities.

Method

Subjects

Two randomized controlled trials of ultrasonographic screening in pregnancy were carried out in Trondheim and Ålesund, Norway, between 1979 and 1981 (Bakketeig et al., 1984; Eik-Nes, Økland, Aure, & Ulstein, 1984). Between 1988 and 1990, a total of 2,161 children were followed up to assess possible harmful effects of routine ultrasonography in utero (Salvesen, Bakketeig, Eik-Nes, Undheim, & Økland, 1992). A subsample of 682 children, born in 1980 and still residing in the vicinity of Trondheim or Ålesund in 1988, was selected for further examination. The study protocol was approved by the regional ethics committee and the school authorities of Trondheim and Ålesund. Written, informed consent was sought from the parents before a child was included in this part of the study.

A total of 677 children were assessed by 108 teachers in the second grade (median per teacher = 5, range = 1 to 21). Five children were not assessed because their teachers refused to participate in the study. In the third grade, 30 children had moved or refused further examinations, 22 did not attend the examinations, and 27 did not complete the test procedures. Thus, the population of the present study consisted of 603 white 9- or 10-year-old children (304 boys and 299 girls). Except for 2 children who were using hearing aids, the children were free of the sensory problems traditionally used as exclusionary criteria for dyslexia. Some characteristics of the study children and the population from which they were drawn are given in Table 1.

Instruments and Procedures

Parents responded to a questionnaire that consisted of 66 closed-end questions concerning sociodemographic data and the child's health, hearing, vision, dominant hand, and neurological development. We specifically included questions about family history of dyslexia, allergies, and left-handedness, and about social variables, such as parental years of education, parental occupation, and family income. We asked 21 questions on dominant hand and 4 questions on immune diseases, due to a reported triadic association among immune disease, left-handedness, and dyslexia (Geschwind & Galaburda, 1985).
During the children’s second year of primary school (age 8 or 9 years), the teacher assessed their oral reading, reading comprehension, spelling, and arithmetic, and gave a score for the child’s overall school performance. The assessment used a performance scale from 1 to 7 (1 = very poor; 4 = medium; 7 = very high). The teachers also reported whether the children needed extra teaching assistance in school. The assessment forms had previously been used in a Norwegian study of dyslexia (Undheim, 1984). The children were given reading, spelling, and intelligence tests during their third year of primary school (age 9 or 10 years). They were examined in groups of 2 to 10 in their own schools; an examination lasted 1.5 years. The children in Trondheim (n = 296) were examined by K.A.S., and the children in Ålesund (n = 307) were examined by K.A.S. in cooperation with four primary-school teachers. The tests had previously been used in Norwegian studies of dyslexia (Bjørgen, Undheim, Romslo, & Nordvik, 1987; Undheim, 1984) and included a nonverbal intelligence test, a reading-comprehension test, a reading-speed test of single words, and a spelling test. The intelligence test was used to measure fluid intelligence (Cattell, 1971), a broad dimension or factor thought to be central in general intelligence (Gustafsson & Undheim, 1992; Undheim & Gustafsson, 1987). The test consisted of three parts: matrices, picture analogies, and number series. The reading-comprehension test included 18 short stories with four content questions each. A Norwegian version of the OS 400 test (Høien & Trana, 1978) was used to assess reading speed of single words, and Asheim’s (1964) spelling test (50 words) for third grade was used to assess spelling. All tests were done on paper. Test descriptions and psychometric properties are presented in Table 2.

### TABLE 1
Characteristics Among Study Children and Children From the Background Population

<table>
<thead>
<tr>
<th>Study children (n = 603) (%)</th>
<th>Background population* (n = 2,161) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>50</td>
</tr>
<tr>
<td>With family history(^b) of Allergy</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Left-handedness</td>
</tr>
<tr>
<td></td>
<td>Dyslexia</td>
</tr>
<tr>
<td>With questionnaire reported Allergy</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Left-handedness</td>
</tr>
<tr>
<td>Lived with both parents during childhood</td>
<td>87</td>
</tr>
<tr>
<td>Having fathers with education of 6–9 years</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>10–12 years</td>
</tr>
<tr>
<td></td>
<td>&gt; 12 years</td>
</tr>
<tr>
<td>Having mothers with education of 6–9 years</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>10–12 years</td>
</tr>
<tr>
<td></td>
<td>&gt; 12 years</td>
</tr>
</tbody>
</table>

*The study population of the previous randomized studies (Salvesen et al., 1992). \(^b\)One or more with the condition, among first- and second-order relatives.
TABLE 2
Test Descriptions and Psychometric Characteristics

<table>
<thead>
<tr>
<th>Description</th>
<th>No. Items</th>
<th>Duration of test (min)</th>
<th>M</th>
<th>SD</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading speed (OS 400)</td>
<td>400</td>
<td>12</td>
<td>221</td>
<td>63</td>
<td>-0.1</td>
</tr>
<tr>
<td>Speed of reading single words. A word is presented, followed by four figure drawings, one of which is to be selected. Score is number of choices correct within time limit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>54</td>
<td>10</td>
<td>39</td>
<td>14</td>
<td>-0.3</td>
</tr>
<tr>
<td>Eighteen short stories, each followed by four multiple-choice questions about content. Score is number of content markings correct within time limit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>50</td>
<td>—</td>
<td>36</td>
<td>9</td>
<td>-0.8</td>
</tr>
<tr>
<td>Written spelling of 50 single words. Sentences containing the word are read aloud. The word is then repeated twice. Score is number of words spelled correctly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrices</td>
<td>28</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>-0.4</td>
</tr>
<tr>
<td>Twenty-eight figure matrices with one empty cell in each. Multiple choice among five or six alternative figures. Score is number of choices correct.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture analogies</td>
<td>30</td>
<td>6</td>
<td>20</td>
<td>5</td>
<td>-0.5</td>
</tr>
<tr>
<td>Thirty analogies using meaningful picture drawings. Multiple choice among five alternative drawings. Score is number of choices correct.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number series</td>
<td>20</td>
<td>6</td>
<td>11</td>
<td>5</td>
<td>-0.2</td>
</tr>
<tr>
<td>Twenty series of numbers. One or two empty cells at the end must be filled in with the next number(s) in the series. Score is number of series correct.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Definitions

Scores from the third-grade tests classified children into groups with low achievement, dyslexia, or neither of the two conditions. Low reading achievement was identified on the basis of the two reading tests, with the reading score being a sum of standardized scores from the two tests. Because reading norms for each age group do not exist for Norwegian children, slow readers were identified by a reading score below -1.5 SD. A total of 49 (8%) children (32 boys and 17 girls) were identified.

Dyslexia was identified by the use of discrepancy scores, whereby the obtained score in reading single words was subtracted from the score that was predicted from the regression of reading on intelligence (Rutter & Yule, 1975; Shaywitz et al., 1992). We used the OS 400 to assess reading speed of single, real words. This reading test is widely used in Scandinavia. The words are all considered to be within the vocabulary knowledge of children this age. A discrepancy score between the observed reading speed score and a predicted reading score based on the regression of reading on intelligence, was computed for each child in the study (Salvesen et al., 1992; Undheim, 1984). Children scoring 1.5 SDs or more below their predicted reading achievement were classified as dyslexic. A total of 26 (4%) children (17 boys and 9 girls) were identified.

Analysis

Teacher ratings in the second grade were correlated with the test results in the third grade using Pearson’s correlation coefficients. We also evaluated the teacher ratings as screening tests to identify children with low achievement (LA) or dyslexia (D). This evaluation was made in terms of detection rate and false-positive rate of the test.
To evaluate the detection rate and the false-positive rate of the teacher ratings, we adopted a method referred to as receiver operating characteristic (ROC) curves (Sackett, Haynes, & Tugwell, 1985). ROC curves characterize the effectiveness of a test when different cutoff points are used to distinguish between normal and abnormal. For each cutoff point, the detection rate and the false-positive rate are plotted on a graph, with the detection rate on the y axis and the false-positive rate on the x axis. A test is ideal if the plotted curve ascends along the y axis. When a test has no diagnostic value, the plot ascends along the x = y line.

Another way of comparing the efficiency of tests is to calculate the kappa index, which is the ratio of the observed accuracy beyond chance to the maximum achievable accuracy beyond chance (Sackett et al., 1985). The kappa index has a maximum value of +1. Values close to 1 reflect almost perfect agreement between tests, values between .2 and .8 indicate a fairly good agreement only by chance, and values below zero show disagreement. We calculated the kappa index for agreement between the reported need for extra assistance in school and low achievement and dyslexia, and between teacher assessment of oral reading (cutoff point at score-value ≤ 3) and low achievement and dyslexia.

Results

Teacher assessments given on 7-point rating scales showed skewed frequency distributions, with medians of 5. Teacher assessments of reading, spelling, and overall performance were highly correlated with each other (r = .8 to .9, p < .01). The arithmetic assessment showed lower correlations with the reading and spelling assessments (r = .7, p < .01), but similar correlation with the overall performance score (r = .8, p < .01).

Test scores were normally distributed. The reading scores were highly correlated (r = .8, p < .01); the spelling scores showed lower correlations with reading scores (r = .6 to .7, p < .01). The reading and spelling scores were only modestly correlated with the intelligence test score (r = .5, p < .01).

Despite the skewness of the teacher assessments, the ratings of oral reading and reading comprehension correlated well with the reading and spelling test scores (r = .6 to .7, p < .01). Teacher assessment of spelling was also highly correlated with the spelling test scores (r = .7, p < .01). Teacher assessments were only modestly correlated with the intelligence test score (r = .5, p < .01), and the correlations between teacher assessments and the discrepancy score were low (r = .3 to .4, p < .01).

TABLE 3
Teacher Assessments Among Children with Low Achievement (LA) and Normal Achievement (NA) and Among Children with Dyslexia (D) and No Dyslexia (ND)

<table>
<thead>
<tr>
<th>Score</th>
<th>Oral reading</th>
<th>Reading comprehension</th>
<th>Overall performance</th>
<th>Oral reading</th>
<th>Reading comprehension</th>
<th>Overall performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LA</td>
<td>NA</td>
<td>LA</td>
<td>NA</td>
<td>LA</td>
<td>NA</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>17</td>
<td>17</td>
<td>6</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>48</td>
<td>9</td>
<td>40</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>97</td>
<td>9</td>
<td>84</td>
<td>5</td>
<td>117</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>99</td>
<td>2</td>
<td>111</td>
<td>5</td>
<td>113</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>128</td>
<td>0</td>
<td>129</td>
<td>2</td>
<td>154</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>151</td>
<td>1</td>
<td>166</td>
<td>0</td>
<td>113</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>13</td>
<td>4</td>
<td>17</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>

Note: Score values: 1 = very poor performance; 4 = medium; 7 = very high performance. For LA, n = 49; for NA, n = 554; for D, n = 26; for ND,
The third-grade tests classified 49 (8%) children with low achievement and 26 (4%) with dyslexia. Teacher assessments of reading and overall performance among subgroups of the study children are given in Table 3. Seventeen (65%) children with dyslexia also had low achievement. In all, 15 (94%) of 16 children with both dyslexia and low achievement had a score value ≤ 3 on the teacher assessment of oral reading, whereas only 3 (33%) of 9 children with dyslexia and normal achievements had a score value this low.

Figure 1 illustrates ROC curves of teacher assessments of oral reading. The detection rate and false-positive rate for different cutoff points for children with low achievement and dyslexia have been computed from Table 3. The best cutoff point to separate children with low achievement or dyslexia from normally achieving children was the score value ≤ 3. This refers to the point on the graph (see Figure 1) that is closest to the coordinates \((x = 0, y = 1)\). At this point, low achievement was separated from normal achievement with a detection rate of 89% and a false-positive rate of 12%, and dyslexia was separated from no dyslexia with a detection rate of 72% and a false-positive rate of 16%.

We do not present ROC curves for other teacher assessments, but analogous graphs for reading comprehension and overall performance may be constructed from Table 3. Thus, it may be shown that, as screening methods, the reading comprehension score and the overall performance score are not as efficient as the oral reading score. We found a detection rate of 74% and a false-positive rate of 9% for children with low achievement, and detection rates from 60% to 64% and a false positive rate of 8% for dyslexia (cutoff score value ≤ 3).

Figure 1. Receiver operating characteristic (ROC) curves showing the detection rate and false positive rate for teacher ratings of oral reading in the prediction of children with low achievement and dyslexia.
The teachers reported that 36 children (22 boys and 14 girls) needed extra teaching assistance in school. However, there was poor agreement between the need for extra teaching assistance in school and the third-grade test results. Only 4 (15%) children with dyslexia had a reported need for extra teaching assistance in school in the second grade (kappa = .1, indicating agreement by chance alone). The agreement between extra assistance and low achievement was slightly better, as 15 (31%) children with low achievement had been identified as needing extra assistance (kappa = .3). In comparison, the kappa index for agreement between oral reading assessment (cutoff score value ≤ 3) and dyslexia was .2, and between oral reading assessment and low achievement was .5.

Discussion

In the present study, teacher ratings correlated well with tests performed 1 year later. In identifying children with low reading achievement, teachers were quite accurate, because 89% of those children were identified. In screening for dyslexia, the assessment of oral reading was somewhat less efficient. Almost all (94%) children with both dyslexia and low achievement were identified by way of scale assessment, and one third of the children with dyslexia and relatively high intelligence test scores were also identified. Thus, teacher assessments of oral reading included a fair proportion of the children with dyslexia and higher IQ in addition to almost all children with both dyslexia and low achievement. Our results indicate that teachers are quite accurate in selecting children with low achievement, and fairly accurate in selecting children with dyslexia, by way of scale assessment.

Using ROC curves to compare teacher assessments and specific tests constitutes a simple method of applying different cutoff points and observing how this may influence detection rates and false-positive rates. In our study we found that the best way to screen for learning disabilities was to test all children scoring ≤ 3 on the teacher ratings of oral reading. In all 18% of the children had oral reading scores ≤ 3; among them were 89% of the children with low achievement and 72% of the children with dyslexia. When selecting cutoff scores on the screening instrument, the usual aim is to identify 15% to 20% of children as failing (Lindsay & Wedell, 1982), as this is the level to be expected from epidemiological surveys. However, if special education services are restricted to examining no more than 8% of all children, only 60% of children with low achievement and 56% of children with dyslexia will be identified by way of scale assessment.

In the present study, scale assessment was better than the teacher selection of children needing extra assistance in school, as the kappa indices for children with dyslexia and low achievement were higher for the teacher assessment of oral reading (.2 and .5, respectively) than for the reported need for extra teaching assistance (.1 and .3, respectively). In a previous study (Undheim, 1984), teachers were asked to “nominate” children with dyslexia. They nominated 14% of the children, which is close to the proportion identified by scale assessments at cutoff values ≤ 3 in the present study. The nomination procedure in the Undheim study resulted in a detection rate of 58% and a false-positive rate of 11% for children with dyslexia. Thus, the results of the present study indicate that teacher rating scales may be a better method of screening for dyslexia than referral for special education services based on a global impression of the child. Teachers tend to refer students who bother them (Shaywitz, Shaywitz, Fletcher, & Escobar, 1989), but this may not be the case when screening with teacher rating scales.
Teacher ratings may be influenced by three types of rater errors: range restriction errors, leniency errors, and halo errors (Anastasi, 1988). Range restriction errors were not a major problem in our study, because the teachers used the full scale in their assessments. Leniency errors may have been a problem, however, because the teachers were more likely to use the top end of the scale instead of the full width of the scale (oral reading score: median = 5, mode = 7), indicating that most pupils were doing very well in school. Thus, leniency errors may influence the ability of the teacher rating scale to distinguish top performers from children with average scores. However, we were preoccupied with the lower tail of the assessment scores, and leniency errors are unlikely to influence the discrimination between children with low achievement and normally achieving children. Halo errors reflect raters' tendencies to allow their overall evaluation of a child to affect the evaluation of each specific aspect of the child's performance. Correlations among teacher assessments relative to test correlations indicated a rather strong halo effect in our study. However, reading is a dominant activity in primary school, and performance in reading probably influences other evaluations, rather than being influenced by them. Thus, ratings of reading may be quite accurate despite such halo effects.

The present study has certain limitations. The study population comprised a sociodemographically homogenous group of 9- and 10-year-old Norwegian children, and the results may not be valid to other populations of children. Also, the results from the present study will be valid only if the tests used can distinguish between children with low achievement and dyslexia and nondisabled children.

Reading disability was recently defined on the basis of ability to decode nonsense words as well as real words (Höien & Lundberg, 1991); no measure of nonsense-word decoding was used in the present study. However, the test for word reading used herein to identify children with dyslexia was similar to other tests often used. A definition of dyslexia based on discrepancy scores for such tests relative to intelligence measures has been used in many other studies (Rutter & Yule, 1975; Shaywitz et al., 1992). Thus, we believe that the children identified with low achievement or dyslexia in the present study are similar to children with reading disabilities identified in many other studies.

ABOUT THE AUTHORS

Kjell A. Salvesen is a research scientist in the Department of Community Medicine and Family Practice, University of Trondheim, Norway. He works in the fields of epidemiology and obstetrical ultrasonography. He received his medical degree from the University of Trondheim. Johan O. Undheim is professor of psychology in the Department of Psychology, AVH, University of Trondheim, Norway. He works in both cognitive psychology and psychometry. His research interests focus on human intellectual abilities and on multivariate analysis of individual differences. Professor Undheim is currently head of the department. Address: Kjell A. Salvesen, Department of Community Medicine and Family Practice, University Medical Center, N-7005 Trondheim, Norway.

AUTHOR'S NOTE

The study was financed by the Norwegian Research Council for Science and the Humanities Grant No. 351.87/001. We thank Ms. Áslaug Gjerdumrød, Ms. Sissel Vindvåg, Ms. Liv Raud, and Ms. Inger Anne Parelussen for their participation in the examinations of children in Ålesund.
REFERENCES


Paper III

Routine ultrasonography in utero and subsequent vision and hearing at primary school age

K. Á. Salvesen*, L. J. Vatten†, G. Jacobsen†, S. H. Eik-Nes*, O. Økland**, K. Molne* and L. S. Bakketeig†

*National Center for Fetal Diagnosis & Therapy, Department of Gynecology and Obstetrics, University Medical Center, Trondheim; †Department of Community Medicine and General Practice, University of Trondheim; ‡Department of Pediatrics, Ålesund Central Hospital; and **Department of Epidemiology, National Institute of Public Health, Oslo, Norway

Key words: Ultrasonography, Randomized Controlled Trial, Vision, Hearing

ABSTRACT

The study was performed to investigate any associations between routine ultrasonography during pregnancy and subsequent reduced vision and/or hearing among children. A follow-up was carried out of primary school children born to women who took part in two randomized, controlled trials of routine ultrasonography during pregnancy. Of 2428 eligible singletons, 2161 (89%) were followed up with a parental questionnaire and with information from maternal and child health centers. Parents assessed vision and hearing in their children and reported their observations in questionnaires. At the age of 7 years, the children were also screened with bilateral visual acuity tests and pure tone audiometry at public maternal and child health centers.

No significant differences between ultrasound-screened children and their controls were found in the parental assessment of vision or hearing. The objective tests did not show any significant differences between children in the two groups with regard to visual acuity or hearing.

The risk of reduced vision or hearing was no greater for children of mothers who had been offered routine ultrasonography during pregnancy than for those whose mothers had not received that offer.

INTRODUCTION

It seems unlikely that the low energy levels emitted from diagnostic ultrasound devices can cause damage to the fetus. This is supported by the fact that no adverse effects of ultrasound exposure during pregnancy have been reported. Still, it has been suggested that potential harm caused by cavitation effects in the cells cannot be excluded. Studies in vitro have shown changes in the cell membrane after exposure to therapeutic ultrasound, and diagnostic ultrasound may affect cell surface motility and architecture. Despite general agreement on the necessity for further research on possible long-term effects among children exposed to ultrasound in utero, only a few studies have been reported.

Routine ultrasonography is usually performed between the 16th and the 22nd week of pregnancy, which is an important phase in the development of the human brain. Mole has proposed that the relay system between the retina and the visual cortex, and the cochlea of the inner ear are two possible sites for localized irreparable damage from ultrasound in utero. Thus, loss of visual acuity over a small area of the visual field or of hearing over a narrow range of sound frequencies are possible adverse effects of ultrasound. However, in a study by Stark and co-workers, no association between antenatal ultrasound exposure and impaired visual acuity or hearing among 7-12-year-old children was demonstrated.

Long-term follow-up of infants in randomized clinical trials has been recommended in order to answer questions about the effect of ultrasound on human development. We have previously reported that there was no association between routine ultrasonography in utero and school performance or dyslexia among 8- and 9-year-old children in a randomized controlled follow-up. The present report on the same children considers possible effects of routine ultrasonography on sensory functions, where the aim was to find out whether routine ultrasonography was associated with impaired vision or hearing.

MATERIAL AND METHODS

Two randomized, controlled trials of ultrasonographic screening in pregnancy were carried out in Trondheim.
The women were offered ultra­

chart of letters under standard conditions. Each eye was

of each child was tested with the Sheridan Gardiner's

aware of the objectives of the present study.

and at

data from these examinations were collected for each

child is 6 weeks

was taken as informed consent for the child to take part

nal and child health centers. Visits take place when a

questions addressed the child's vision. Return of the questionaire

addressed the child's hearing, and four

questionnaire consisted of 66 closed-end questions. Three

information and a postage-paid return envelope. The

were sent a questionnaire together with a letter of

reported elsewhere 8. The mothers of the eligible children

and details about the study population have been re­

and Alesund, Norway, in 1979–819,10. The study design

and methods of randomization (sealed envelope method)

were identical. The pregnant women in Alesund were

and Alesund, Norway, in

The study design

of pregnancy. The same ultrasonic devices (ADR

stated before the study. With a given sample of

possible to detect a 7.5% increase in the occurrence of

impaired distant visual acuity (from a prevalence of

p

0.10), power calculations showed that it would be

possible to detect a 75% increase in the occurrence of

impaired distant visual acuity (from a prevalence of

4–7%), and to detect a 50% increase in the occurrence of

impaired hearing from a prevalence of 7–11%.

Data analyses were carried out with the SPSS (Statisti­

Package for the Social Sciences). Proportions of

missing data between groups were compared using the

Mantel–Haenschen χ² statistics or the Fisher exact test.

The associations between routine ultrasonography in

utero and subsequent impaired vision or hearing were

assessed by using the odds ratio as a measure of relative

risk. The precision of the odds ratio is presented as 95%

confidence intervals (CI), calculated from Mantel–Haenschen

χ² statistics11.

RESULTS

Of 1244 children in the screened group, five had died,

and the parents of 124 did not return the questionnaire.

This left 1115 children to be studied. In the control group

of 1184 children, one had died, and 137 parents did not

respond, leaving 1046 children to be studied. Data from

the maternal and child health centers were available for

1107 screened and 1033 control children. Information on

tested separately. In children who used glasses, testing

was done with the aid of glasses. A test was regarded as

normal if a child was able to interpret correctly at least

two of the pictures corresponding to a visual acuity of

6/9 with its poorest eye, and with no more than a

difference of one line between the eyes. If a child had an

abnormal test, re-examination was performed after one

month, before referral to an ophthalmologist was made.

If a child was re-examined, data from the last visit were

used. In this study, impaired vision was defined as visual

acuity less than 6/9 (= 0.83) in one of the eyes at the age of

7 years.

Hearing tests were carried out with the use of pure

tone audiometry according to a standard procedure11. At

4 years, the child was tested routinely with at least three

frequencies in the range 500–4000 Hz in both ears. At 7

years, at least four, but occasionally up to seven frequen­
cies (250, 500, 1000, 2000, 4000, 6000 and 8000 Hz) were

tested. We used the frequencies of 500, 1000, 2000 and

4000 Hz from both ears at the age of 7 in the analysis.

Re-examination was usually performed when a child had

an abnormal first screening: we used the latest test

performance at the age of 7 years. Impaired hearing was

defined as a hearing reduction of 20 dB at three or more

frequencies or a reduction of 30 dB at one frequency11.

Information on childhood history of a middle ear

effusion was occasionally reported by the parents in the

questionnaire and even more consistently documented in

the medical records at the maternal and child health

centers.

A null hypothesis of no difference in visual acuity or

hearing between screened children and their controls was

stated before the study. With a given sample of 1000

children in each group, and a two-sided α of 0.05 and a

β of 0.10, power calculations showed that it would be

possible to detect a 75% increase in the occurrence of

impaired distant visual acuity (from a prevalence of

4–7%), and to detect a 50% increase in the occurrence of

impaired hearing from a prevalence of 7–11%.

Table 1 Background variables among screened (n = 1115) and

control children (n = 1046)

<table>
<thead>
<tr>
<th></th>
<th>Screened group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of children male</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>Mean (SD) age (years) at the time of vision and hearing tests</td>
<td>6.9 (0.4)</td>
<td>6.9 (0.4)</td>
</tr>
<tr>
<td>% of children with family economy reported as good</td>
<td>59</td>
<td>56</td>
</tr>
<tr>
<td>medium</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>poor</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>% of children lived with both parents during childhood</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>% of mothers with education of 6-9 years</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>10-12 years</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>&gt;12 years</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>% of fathers with education of 6-9 years</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>10-12 years</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>&gt;12 years</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

Norwegian children are regularly examined by physi­
cians and specially trained public health nurses at materi­

The study design and methods of randomization (sealed envelope method)

were identical. The pregnant women in Ålesund were representative of the general population, whereas the study population in Trondheim tended towards including low-risk pregnancies. The women were offered ultra­sonographic examinations in the 19th and 32nd weeks of pregnancy. The same ultrasonic devices (ADR 3120, Tempe, Arizona, USA) were used in Trondheim and Ålesund. These scanners produced lower intensities than do most obstetrical scanners in use today.

In the two previous studies combined, 2637 women were randomized into a screening group of 1335 women and 1302 controls8,10. In all, 1244 screened children and 1184 controls were eligible for follow-up after 8 years, and details about the study population have been reported elsewhere. The mothers of the eligible children were sent a questionnaire together with a letter of information and a postage-paid return envelope. The questionnaire consisted of 66 closed-end questions. Three questions addressed the child’s hearing, and four addressed the child’s vision. Return of the questionnaire was taken as informed consent for the child to take part in the follow-up study. Two reminders were sent to non-responders.

Vision and hearing

Table 1 Background variables among screened (n = 1115) and control children (n = 1046)
Table 2 Questionnaire items regarding vision and hearing among screened (n = 1115) and control children (n = 1046)

<table>
<thead>
<tr>
<th>Questionnaire reported</th>
<th>Screened group</th>
<th>Control group</th>
<th>Odds ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced hearing</td>
<td>32</td>
<td>280</td>
<td>0.78</td>
<td>0.48-1.26</td>
</tr>
<tr>
<td>Otitis &gt; 2 episodes per year</td>
<td>214</td>
<td>199</td>
<td>0.87</td>
<td>0.51-1.50</td>
</tr>
<tr>
<td>Visit to an otorhinolaryngologist</td>
<td>314</td>
<td>300</td>
<td>0.87</td>
<td>0.61-1.24</td>
</tr>
<tr>
<td>Reduced vision</td>
<td>68</td>
<td>72</td>
<td>0.87</td>
<td>0.61-1.24</td>
</tr>
<tr>
<td>Use of glasses</td>
<td>82</td>
<td>95</td>
<td>0.78</td>
<td>0.58-1.07</td>
</tr>
<tr>
<td>Ever observed squinting</td>
<td>110</td>
<td>101</td>
<td>0.80</td>
<td>0.60-1.29</td>
</tr>
<tr>
<td>Treated for squinting</td>
<td>47</td>
<td>46</td>
<td>0.96</td>
<td>0.66-1.41</td>
</tr>
</tbody>
</table>

Table 3 Screened (n = 1042) and control children (n = 966) with hearing acuity thresholds above 20 dB at frequencies 500-4000 Hz

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Screened group</th>
<th>Control group</th>
<th>Odds ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right ear</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>59</td>
<td>70</td>
<td>0.77</td>
<td>0.54-1.10</td>
</tr>
<tr>
<td>1000</td>
<td>32</td>
<td>46</td>
<td>0.63</td>
<td>0.40-1.00</td>
</tr>
<tr>
<td>2000</td>
<td>24</td>
<td>26</td>
<td>0.85</td>
<td>0.48-1.50</td>
</tr>
<tr>
<td>4000</td>
<td>40</td>
<td>38</td>
<td>0.97</td>
<td>0.56-1.67</td>
</tr>
<tr>
<td>Left ear</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>67</td>
<td>60</td>
<td>1.04</td>
<td>0.71-1.55</td>
</tr>
<tr>
<td>1000</td>
<td>40</td>
<td>38</td>
<td>0.97</td>
<td>0.56-1.67</td>
</tr>
<tr>
<td>2000</td>
<td>32</td>
<td>29</td>
<td>1.02</td>
<td>0.66-1.57</td>
</tr>
<tr>
<td>4000</td>
<td>35</td>
<td>41</td>
<td>0.78</td>
<td>0.49-1.25</td>
</tr>
</tbody>
</table>

DISCUSSION

In this randomized controlled follow-up, we found no association between routine ultrasonography in utero and subsequent impaired vision or hearing, which is in agreement with the results of others. Thus, our data do not support the hypothesis of possible damage to the relay system between retina and visual cortex and to the cochlea of the inner ear, after exposure to diagnostic ultrasonography in utero.

A strong feature of this study is that a randomized controlled design rules out many of the possible biases that might influence an association between routine ultrasonography and the measurements of outcome. Nonetheless, the potential for misclassifying children due to imprecise measurements may be a threat to the validity of the results. However, if the tests of vision and hearing at the maternal and child health centers were subject to misclassification, it seems unlikely that the bias is differential, depending on ultrasound exposure. Instead, a non-differential misclassification may be expected. This would attenuate any estimated association (the odds ratio) and produce a more conservative result, which would be biased towards the null hypothesis. Thus, an association that might have been detected with more precise instruments may have been overlooked.

Parents were aware of the ultrasound exposure status of their child, and this may have caused a systematic bias in their assessment of the child's vision and hearing. Thus, the consistent, although not statistically significant, tendency for screened children to have lower odds of sensory impairment may be due to a possible under-reporting among their parents.

A visual acuity test is likely to detect most clinically significant disorders in children since visual acuity at the age of 7 years was available for 1023 screened children and 948 controls, while information from pure tone audiometry at 7 years was available for 1042 screened and 966 control children.

Background data for the two groups of children are presented in Table 1. There was no obvious difference between the groups. Data on vision and hearing from the parental questionnaire are presented for screened and control children in Table 2. There was no significant difference in impairment of sensory functions between the groups of children, as assessed by their parents. According to the parental reports, 68 (6%) of 1113 screened children and 72 (7%) of 1039 control children had abnormal vision (odds ratio = 0.87; 95% CI, 0.61-1.24). Parents also reported that 32 (3%) of 1112 screened and 38 (4%) of 1041 control children had abnormal hearing (odds ratio = 0.78; 95% CI, 0.48-1.26).

The visual acuity tests at the age of 7 years classified 31 (3%) of 1023 screened and 34 (4%) of 948 control children as having impaired visual acuity (odds ratio = 0.84; 95% CI, 0.51-1.38).

In all, 77 (7%) of 1042 screened children were classified as having reduced hearing according to the pure tone audiometry, compared to 69 (7%) of 966 controls (odds ratio = 1.04; 95% CI, 0.72-1.49). Among the children with reduced hearing, two of 1042 screened children and seven of 966 controls were deaf or dependent on a hearing aid (p = 0.07). The numbers of screened and control children with hearing acuity thresholds above 20 dB at the frequencies 500-4000 Hz in both ears are given in Table 3.
tests were carried out with the aid of glasses, if such were used, misclassification of vision might occur, and if the use of glasses was unevenly distributed between screened and control children, this could have influenced the results. Based on the parents’ reports (Table 2), 8% of the screened children and 10% of the controls (p = 0.14) used glasses. Controlling for the use of glasses in the analyses of visual acuity at the age of 7 years did not alter the estimated effect (odds ratio = 0.82).

There was also a potential for misclassification in the pure tone audiometry screening. In a Norwegian study, 3.5% of 17 597 children aged 7 years suffered from reduced hearing, but 7.5% of the children were identified with a suspicion of reduced hearing at the initial audiometry screening. Thus, the initial audiometry screening had a sensitivity of 86% and a specificity of 96% among 7-year-old children. Persistent middle ear effusion is the most important cause of reduced hearing in children, and may be responsible for as many as 55% of children with reduced hearing at the age of 7. There is no theoretical link between ultrasound exposure in utero and the development of a middle ear effusion during childhood, and controlling for middle ear effusion in the analysis of routine ultrasonography and reduced hearing at the pure tone audiometry did not alter the effect (the odds ratio changed from 1.04 to 1.06).

We assessed hearing ability with the use of the pure tone frequencies 500–4000 Hz, and the generalization of the results should be restricted to those frequencies. High-frequency audiometry was not performed in the present study, and threshold audiometry at frequencies above 8 kHz up to 20 kHz appears to be a valid indicator of auditory sensitivity, especially during childhood. Since ultrasound consists of high-frequency mechanical waves above 20 kHz, and the cochlea of the inner ear is a postulated target organ for adverse effects of ultrasound, high-frequency audiometry should be included in future follow-up studies of children.

Missing data are always a threat to the validity of a study. Our study had a response rate of 89% for the parental questionnaire. We obtained information from the visual acuity tests for 81% and audiometry data for 83% of the 7-year-old children. In addition, we had access to information from vision and hearing tests at the age of 4 years in 85 children (4%) who had missing visual acuity data at 7 years, and in 63 children (3%) who had missing audiometry data at 7 years. At the examination at 4 years, seven of 39 screened children and five of 46 control children had impaired visual acuity (p = 0.27). At 4 years, three of 34 screened children and four of 29 control children had reduced hearing (p = 0.41). Thus, it seems unlikely that the missing data could have distorted the test results at the age of 7 years.

If ultrasound screening during pregnancy has any biological effect on sensory development, one might expect such effects to be subtle and difficult to detect. In this study we had a statistical power (α = 0.05, β = 0.90) to detect a 75% increased frequency of reduced visual acuity and a 50% increase in hearing impairment. Thus, even a fairly large study like this one has insufficient power to detect small differences in sensory impairment between screened and control children.

Exposure status in the present study was assessed according to the principle of 'intention to treat', and details of actual ultrasound exposure of the study children have been reported elsewhere. In all, 3% of the screened children were never exposed to ultrasound in utero, whereas 7% of the controls were scanned several times. Between 16 and 22 weeks of pregnancy, only 5% of the controls had ultrasonography, while 8% of the screened children had no ultrasound. Thus, the overlap in ultrasound exposure between the randomized groups is probably of little importance in the interpretation of the results.

In conclusion, the results of the present study suggest that routine ultrasonography offered in weeks 19 and 32 of pregnancy has no adverse effects on sensory functions of children who have been followed to primary school age.

ACKNOWLEDGEMENTS

The study has been supported by the Norwegian Research Council for Science and the Humanities (NAVF), grant no 351.87/001. We also thank the Norwegian Society for Ultrasound Diagnosis in Medicine, Scanned A/S, and MEBI A/S for financial contributions. Ms Nancy Lea Eik-Nes has revised the manuscript.

REFERENCES

Vision and hearing


*Ultrasound in Obstetrics and Gynecology* 247
Paper IV

Routine ultrasonography in utero and subsequent handedness and neurological development

Kjell A Salvesen, Lars J Vatten, Sturla H Eik-Nes, Kenneth Hugdahl, Leiv S Bakketeig

Abstract
Objective—To examine any associations between routine ultrasonography in utero and subsequent brain development as indicated by non-right handedness at primary school age and neurological development during childhood.

Design—Follow up of 8 and 9 year old children of women who took part in two randomised, controlled trials of routine ultrasonography during pregnancy.


Subjects—2161 (89%) of 2428 eligible singletons were followed up, partly through a questionnaire to their parents and partly through information from health centres.

Main outcome measures—The dominant hand of the child was assessed by 10 questions. Deficits in attention, motor control, and perception were evaluated by five questions. Impaired neurological development during the first year of life was assessed by an abbreviated version of the Denver developmental screening test.

Results—The odds of non-right handedness were higher among children who had been screened in utero than among control children (odds ratio 1.32; 95% confidence interval 1.02 to 1.71). No clear differences were found between the groups with regard to deficits in attention, motor control, and perception or neurological development during the first year of life.

Conclusion—Our data suggest a possible association between routine ultrasonography in utero and subsequent non-right handedness, whereas no association with impaired neurological development was found. As the question on non-right handedness was one of six initial hypotheses, the observed results may be due to chance. None the less, the results suggest that the hypothesis may have some merit and should be tested in future studies.

Introduction
The common indications for diagnostic ultrasound scanning in pregnancy and the routine screening offered in some countries result in most pregnant women in developed countries being exposed to the procedure. No adverse effects of diagnostic ultrasound screening in pregnancy have been reported. Possible long term effects among children exposed to ultrasound in utero, however, have been examined in only a few studies. The general consensus is that further research on this topic is warranted.

Abnormal development is typically related to disturbances during critical stages of gestation. Routine ultrasonography is usually done between the 16th and the 22nd week of pregnancy, which is an important phase of brain development. At this point neurones migrate towards their destination in the fetal brain. Experimental studies in vitro have shown changes in the cell membrane and cell surface motility and architecture after exposure to ultrasound. Ultrasound might influence neuronal migration, and it has been suggested that altered cerebral dominance, dyslexia, or impaired neurological development may be the result of a disturbed migration of neurones. The dominant hand may serve as an indicator of cerebral dominance. The normal high prevalence of right handedness means that random damage to the hemispheres will increase left handedness.
Long-term follow-up of infants in randomised clinical trials has been recommended to answer questions about the effect of ultrasound on human development. In a previous report we were unable to find any association between routine ultrasonography in utero and poor performance at school or dyslexia among 8 and 9 year-old children.1 Nor did we find any differences in vision or hearing at ages 4 and 7.2 In the present report on the same children we aimed to find out whether routine ultrasonography was associated with changes in handedness patterns or with impaired neurological development.

Subjects and methods

Two randomised controlled trials of ultrasonographic screening in pregnancy were carried out in the Norwegian cities of Trondheim and Alesund in 1979-81.3,4 The study design and methods of randomisation (sealed envelope method) were identical. The pregnant women in Alesund were representative of the general population,5 whereas the study population in Trondheim included more low-risk pregnancies.6 The study women were offered ultrasonographic examinations in the 19th and 32nd weeks of pregnancy. The same ultrasonic devices (ADR 2130, Tempe, Arizona) were used in Trondheim and Alesund. Those scanners, designed to produce lower intensities than do most scanners in obstetric use today. The median exposure time for the first routine scan in Alesund was three minutes.

Altogether 2637 women were randomised into a screening group of 1335 women and 1302 controls. Among the screened women there were 15 pairs of twins, 55 abortions, and eight perinatal deaths among singleton fetuses. The whereabouts of 13 children could not be traced eight years after the original studies, leaving 1244 eligible, live-born singletons to be followed up in 1988. In the control group, there were 10 pairs of twins, one set of triplets, 55 abortions, and eight perinatal deaths among singletons, three late neonatal deaths, and 27 women who could not be found, leaving 1184 eligible singletons for the present study.

Mothers of all the 2428 eligible children were sent a questionnaire to be filled in by an information letter and a postage paid return envelope. The questionnaire consisted of 66 closed questions on sociodemographic data; the child's health, hearing, and vision; and specific questions about dominant hand and neurological development. We specifically included questions about family history of dyslexia, left handedness, and allergy and about social variables such as parental years of education, parental occupation, and family income. Immune diseases were of particular interest because of a reported triadic association among immune disease, dyslexia, and left handedness,7 and these were covered in four questions. Return of the questionnaire was taken as informed consent for the child to take part in the follow-up study. Two reminders were sent to non-responders.

Norwegian children are regularly examined by physicians and specially trained public health nurses at the maternal and child health centres. Visits take place when a child is 6 weeks old and at 3, 6, and 12 months and at 2, 4, and 7 years of age. The prospectively recorded data from these examinations were collected for each child in the study. The staff at the maternal and child health centres were not aware of whether the child had been exposed to ultrasound or the objectives of the study.

Among six stated hypotheses in the study protocol, two dealt with vision and hearing and four with possible effects of routine ultrasonography on the developing fetal brain. These included increased incidence of dyslexia; deficits in attention, motor control, and perception; impaired neurological development during the first year of life; and changes in handedness. Analyses of the association between ultrasound scanning and dyslexia have been reported elsewhere.8

HANDEDNESS

The dominant hand of the child was assessed with 21 questions taken from a modified version of a questionnaire developed by Rackzowski and coworkers.9 The parents answered specific questions about which hand the child preferred to use while performing various tasks in daily life activities. They were instructed not to respond if they had never observed the child do the task in question. Response options were the left hand, equally often with either hand, or the right hand.

Before the study we had decided to include questions on a variety of activities and to exclude questions that were not responded to with reasonable frequency. We had not, however, decided which questions to include in the analysis before the study.

Complete data on all 21 questions was available for only 1210 children (50%). In a trade-off between increasing statistical power and losing information by dropping questions, we decided to use information from 10 of the 21 questions. These included which hand the child preferred when drawing, writing, dealing cards, using a bottle opener, throwing a ball, using an eraser and a pair of scissors, eating with a spoon and a fork, and cutting with a knife. Complete data were available from 1663 children (69%). The 10 items represented various aspects of activities of the daily living. The child's handedness was classified as right handed, left handed, or ambidextrous (if there was no clear-cut preference).

Children were classified as non-right-handed if they were not right-handed, thereby including all children who were left-handed.

The data were also analysed with a quantitative approach by using a handedness score based on the 10 selected questions. If none of the 10 questions were answered as right-handed, the handedness score was 0. In all 10 questions were answered as right-handed, the handedness score was 10. The distribution of this laterality score was, of course, highly skewed towards right-handedness. Thus, the handedness score was compared between screened and control children with non-parametric statistics.

NEUROLOGICAL DEVELOPMENT DURING FIRST YEAR OF LIFE

Neurological development in infancy is closely monitored at the maternal and child health centres. The original Denver developmental screening test included 105 items which cover four developmental functions in infants and preschool children (gross motor, language, fine motor-adeptive, and personal-social functions).10 In Norway a modified version with 10 items has been used for the past 20 years, including six items for gross motor functions: prone, lifts the head up 90 degrees (should be achieved by the age of 5 months); rolls over (6 months); sits without support (9 months); pulls self to stand (11 months); walks holding on to furniture (12 months); and walks well (14 months). Four items for personal-social, language, and fine motor-adeptive functions comprise smiles responsively (should be achieved by 4 months); imitates speech sounds (7 months); thumb-finger grasp (10 months); and three words other than "mama" or "dada" (14 months).

A child was included in the analyses if information on at least one of the 10 items from the short version of the Denver test was available. Children were classified as having impaired neurological development if they had not achieved one of the 10 functions at the
expected age. In addition, mothers reported in the questionnaire at what age their child started to walk.

ATTENTION, MOTOR CONTROL, AND PERCEPTION

Deficits in attention, motor control, and perception have been replaced by the initial description minimal brain dysfunction. This may be a sign of impaired neurological development, which is first detectable in preschool children. It has been shown that this condition has a prevalence of 7% among preschool children in Sweden. *

We used a questionnaire that was specifically developed to identify children with deficits in attention, motor control, and perception with a reported sensitivity of 74% and a specificity of 92%. One of the original six questions, however, apparently did not identify children with deficits in attention, motor control, and perception in our study. Thus almost one third of parents agreed that their child moved about by "shuffling" before starting to walk implying that as many as one third of the children had signs of deficits in attention, motor control, and perception. When we restricted the analysis to the remaining five questions, 15% of children in the study were classified as having deficits in attention, motor control, and perception.

POWER CALCULATIONS AND STATISTICAL ANALYSIS

With a given sample size of 1000 children in each group and a power of 0.05 and a β of 0.10, power calculations before the study showed that we would be able to detect a 25% increase in the prevalence of non-right handedness (from a base prevalence of 15-21%).

Anallogously, a 50% increase in left handedness from a prevalence of 9-13% and a 75% increase in the prevalence of deficits in attention, motor control, and perception from 5-9% would be detected. Power calculation of the hypothesis of impaired neurological development during the first year of life had not been done before the study.

Analyses were done with the statistical package for science. We compared proportions of missing data between groups with Mantel-Haenszel χ² statistics. The associations between routine ultrasonography and subsequent handedness; neurological impairment; and deficits in attention, motor control, and perception were assessed by using the odds ratio as a measure of relative risk. The precision of the odds ratio is presented as 95% confidence intervals, calculated from Mantel-Haenszel χ² statistics. * Differences in the mean age at walking between the two groups of children was tested with Student's t test. The handedness score was compared between groups with the Mann-Whitney test. The data were collected from randomised controlled trials in two centres with identical study design and method of randomisation. The analyses were first done stratified by centre but as the results were homogeneous pooled estimates are presented.

Results

Of 1244 children in the screened group, 1115 were studied; five had died (two of congenital heart disease and three of sudden unexpected death in infancy), and the parents of 124 did not respond to the questionnaire. In the control group of 1184 children, one had died (sudden unexpected death in infancy), and the parents of 137 did not respond to the questionnaire, which left 1046 children to be studied. Data from maternal and child health centres were available for 1107 children in the screened group and for 1053 controls. Information on the Denver developmental screening test, however, was available for only 859 screened children and 798 controls. We found no obvious differences between screened and control children on any of the collected family or social variables (table I) nor between the children included in the analysis of handedness and those excluded because of missing data.

Complete data on the 10 selected questions from the handedness questionnaire were available for 1663 children (69%). In addition, we had information on which hand the child used the most before starting school and family history of left handedness for 466 of the children with missing data. Among the 239 children in the screened group for whom some data were missing, 34 (14%) were reported to be non-right handed and 44 children in the control group for whom some data were missing were non-right handed. Among these children, 91 (38%) in the screened group reported to have one or more left handers among their first and second order relatives compared with 81 (37%) in the control group. In total, 356 (34%) screened children reported having a family history of left handedness compared with 350 (36%) control children (table II).

Ultrasonographic exposure in utero of screened and control children is shown in table II. The mean number of scans in the screened group was 2.5 (SD 0.9). Table III shows the numbers of left handed children and children using either hand equally often in screened and control groups for each item in the questionnaire. With the use of the 10 selected questions we classified 162 (19%) of 861 screened children as non-right handed compared with 120 (15%) of 802 controls (odds ratio 1.32; 95% confidence interval 1.02 to 1.71). Of these, 62 (7%) children in the screened group were classified as left handed compared with 44 (5%) control children (1.34; 0.90 to 2.00). The mean handedness score was 8.70 among screened and 8.95 among control children. The median score was 10 in both groups. The distribution of the handedness score was significantly different between screened and control children (p<0.02).

A total of 1654 children could be included in the analyses of impaired neurological development during
the first year of life. According to the short version of the Denver developmental screening test 75% (95% CI 69–80) of 859 children in the screened group and 73% (90% CI 68–78) of 798 children in the control group had impaired neurological development (0.95; 0.68 to 1.33). In all 2128 mothers reported the age when their child started to walk. The mean age for walking was 12 months in both groups.

A total of 2100 children were included in the analyses of deficits in attention, motor control, and perception. Of the 1081 children in the screened group, 147 (14%) were classified as having deficits in attention, motor control, and perception compared to 163 (16%) of 1019 control children (0.83; 0.66 to 1.05).

Allergies, as reported by the mothers, were equally prevalent among the children in the two groups. About a fifth of the children had experienced one or more episodes of allergy which had been treated with prescribed medication.

Discussion

In this randomised controlled follow up we found a possible association between routine ultrasonography in utero and subsequent non-right handedness among children in primary school. No previous study has assessed the association between routine ultrasonography and non-right handedness.

Table III — Use of left hand or other hand among 1115 children who had been screened by ultrason in utero and 1046 children who had not.

Table IV — Different approaches to analysis of association between ultrasonography and non-right handedness in children who had been screened in utero and those who had not.

The problem with the two latter alternatives is the loss of power resulting from incomplete response to the items. Overall, however, the results show a consistent positive association, suggesting that ultrasonography.
screened children had a relative risk of non-right handedness of about 1.3. We found no association between ultrasonography and impaired neurological development, which agrees with the results of other studies. Neurological development during the first year of life was assessed from data collected at maternal and child health centres. The precision of the modified version of the Denver developmental screening test has not been formally evaluated, but the study design precludes that assessment of neurological development would be systematically influenced by the children's exposure to ultrasound in utero.

Analogous arguments related to misclassification would apply to the questionnaire, which was designed to measure deficits in attention, motor control, and perception. In this study as many as 15% of the children were classified as having deficits, but there was no statistical difference between the two groups. The observed prevalence was clearly above the assumed base prevalence of 7%. Thus, the instrument may be inaccurate for measuring these deficits. In a study of Swedish preschool children the reported specificity of the questionnaire was 92%, which indicates that the false positive rate may be rather high.

Data in the present study were analysed according to the "intention to treat" principle. Table II shows that 3% of the children who were offered screening were never exposed to ultrasound in utero whereas 7% of the controls were exposed several times. During the perceived critical time window at 16 to 22 weeks of pregnancy 5% of the controls were scanned and 8% of the screening group were not. Thus, the overlap in ultrasound exposure between the randomised groups was probably of little importance in the interpretation of the results.

In the results, we did exploratory analyses of the association between ultrasound exposure (at 16-22 weeks) and handedness regardless of which screening group the child had been in. By doing so the positive association with non-right handedness was strengthened for 12 of the 21 questions in table III, indifferently for five, and weakened for four questions. The estimated odds ratio of non-right handedness increased from 1.32 to 1.34. After adjusting for familial predisposition of left handedness the odds ratio was 1.42.

Many fetuses are exposed to ultrasound from additional sources during pregnancy (fetal heart rate detectors and electronic fetal monitoring). In this study such use should have been evenly distributed between screened and control children but may nevertheless represent a background influence, which may weaken the estimated association between ultrasonography and subsequent handedness.

Women who were randomly allocated to routine ultrasonographic screening were typically examined at weeks 19 and 32 of pregnancy. Although a potential biological effect of ultrasound would focus on the developing fetal brain, it does not seem plausible that the low energy levels emitted for diagnostic ultrasound would apply to the questionnaire, which was designed to measure deficits in attention, motor control, and perception. In this study as many as 15% of the children were classified as having deficits, but there was no statistical difference between the two groups. The observed prevalence was clearly above the assumed base prevalence of 7%.

Data in the present study were analysed according to the "intention to treat" principle. Table II shows that 3% of the children who were offered screening were never exposed to ultrasound in utero whereas 7% of the controls were exposed several times. During the perceived critical time window at 16 to 22 weeks of pregnancy 5% of the controls were scanned and 8% of the screening group were not. Thus, the overlap in ultrasound exposure between the randomised groups was probably of little importance in the interpretation of the results.

In the results, we did exploratory analyses of the association between ultrasound exposure (at 16-22 weeks) and handedness regardless of which screening group the child had been in. By doing so the positive association with non-right handedness was strengthened for 12 of the 21 questions in table III, indifferently for five, and weakened for four questions. The estimated odds ratio of non-right handedness increased from 1.32 to 1.34. After adjusting for familial predisposition of left handedness the odds ratio was 1.42.

Many fetuses are exposed to ultrasound from additional sources during pregnancy (fetal heart rate detectors and electronic fetal monitoring). In this study such use should have been evenly distributed between screened and control children but may nevertheless represent a background influence, which may weaken the estimated association between ultrasonography and subsequent handedness.

Women who were randomly allocated to routine ultrasonographic screening were typically examined at weeks 19 and 32 of pregnancy. Although a potential biological effect of ultrasound would focus on the developing fetal brain, it does not seem plausible that the low energy levels emitted for diagnostic ultrasound devices (such as the ADR scanners) would cause damage to the fetal brain. Nevertheless, potential harm caused by cavitation effects in the cells cannot be excluded.

Experimential studies in vitro have suggested that ultrasound may influence cell membranes. Others have hypothesised that ultrasound exposure in utero may influence neuronal migration during a critical stage, which may influence brain development and be an underlying explanation for changes in handedness patterns, dyslexia, or impaired neurological development. We have previously examined the relation between ultrasonography and dyslexia and found no evidence to support the hypothesis. Thus, the finding of a link between ultrasound exposure in utero and non-right handedness is weak and may be due to chance or the small number of study participants.


Clinical implications

- Most women in developed countries have ultrasound examinations during pregnancy.
- No problems from the use of ultrasonography have so far been detected.
- This study shows a positive association between ultrasound scanning during pregnancy and dyslexia.
- The proportion of children who are not right handed at the age of 8 and 9 years.
- This may be due to chance or it may be the result of ultrasonography's effect on the development of the fetal brain.
- This study found no association between ultrasonography during pregnancy and impaired neurological development of the child.

and laterality might seem odd, but it is often not recognised just how tenuous the association between dyslexia and lateralisisation is.

The present study does not indicate any association between ultrasound in utero and impaired neurological development. We are, however, left with an unexplained positive association between ultrasound scanning during pregnancy and non-right handedness. This is one possible chance finding among a number of non-significant findings. Theoretically, the concept of pathological left handedness implies that children with early brain damage to the left hemisphere will have an increased incidence of left handedness. A left hemisphere lesion, however subtle, may cause a shift in hand preference in otherwise genotypic right handers, thus increasing the overall percentage of non-right handedness in these children. Increased incidence of non-right handedness in a particular population may therefore be a sensitive index of subtle changes in the development of the brain or parts of the brain. We would, however, emphasise the need to replicate the positive association between ultrasound and non-right handedness before it is interpreted as more than a chance finding. Follow up of the children from the Swedish, Finnish, or other randomised controlled trials of ultrasonography in pregnancy may help clarify this issue.

The study was supported by the Norwegian Research Council for Science and the Humanities (NANVF), grant no. 151.87/00. We also thank the Norwegian Society for Ultrasound Diagnosis in Medicine and Trøndelag medisinsk selskap for financial contributions. Ms Nancy Lee Eik-Nes has given valuable comments to the manuscript.
References:

(Accepted 11 June 1983)

Printed in Great Britain by Meridian Print Centre Ltd., Derby.)
Paper V

Routine ultrasonography in utero and subsequent growth during childhood

K. Á. Salvesen*†, G. Jacobsen†, L. J. Vatten†, S. H. Eik-Nes* and L. S. Bakketeig†‡

*National Center for Fetal Medicine, Department of Gynecology and Obstetrics, University Medical Center, Trondheim; †Department of Community Medicine and General Practice, University of Trondheim; and ‡Department of Epidemiology, National Institute of Public Health, Oslo, Norway

Key words: Ultrasonography, Pregnancy, Randomized Controlled Trial, Childhood Growth

ABSTRACT

The study was performed to investigate any associations between routine ultrasonography during pregnancy and subsequent growth during childhood. A follow-up was carried out of children born to women who took part in two randomized, controlled trials of ultrasonic screening during pregnancy. From a total of 2428 eligible children, 2140 (88%) were followed with measurements of growth at Norwegian maternal and child health centers. The body weight and height were recorded at birth and at 3, 6 and 12 months of age; and subsequently at 2, 4 and 7 years of age. A repeated-measures analysis of growth was carried out, stratified by maternal smoking in a subsample of 1201 children. No significant differences between ultrasound-screened children, and their controls, were found in mean body weight, or height, at birth and at all the subsequent visits to the health centers. However, the repeated-measures analysis of variance in the subsample indicated that growth from birth to 7 years of age differed significantly (p = 0.02) between screened and control children of mothers who reported smoking at the first antenatal visit. We conclude that children who were routinely exposed to ultrasonography in utero showed no statistically significant differences in growth during childhood compared to control children.

INTRODUCTION

Most pregnant women in developed countries are exposed to diagnostic ultrasound during pregnancy. No adverse effect of this procedure has been reported, but such effects cannot be entirely ruled out. The general consensus is that further research is needed.

In animal experiments, the most consistent finding has been a decrease in fetal weight among mice which were exposed to ultrasound at varying times during gestation. However, this weight decrease was associated with the use of much higher ultrasonic intensity levels than the levels used by diagnostic equipment. Further, the association was restricted to certain strains of mice while other strains and rats were unaffected by similar exposures.

In an epidemiological study, Moore and co-workers found reduced birth weight among children who were exposed to ultrasound in utero. The authors concluded, however, that confounding with maternal and fetal risk factors, rather than ultrasound exposure, might have explained the reduced birth weight. A Swedish randomized controlled trial of routine ultrasonography in pregnancy reported a significantly higher mean birth weight (but not birth length) among screened children. Babies in the screened group were on average 42 g heavier, and the difference was even greater (75 g) for mothers who at their first antenatal visit had reported being smokers. The authors hypothesized that the experience of the ultrasound examination might have induced mothers in the study group to reduce smoking and change to a healthier lifestyle more often than unscreened women. Other studies have reported no differences in birth weight between ultrasound-exposed children and controls. Only one study has assessed growth among ultrasound-exposed and control children beyond birth. Lyons and co-workers found no differences in weight or height between the groups from birth and up to 6 years of age.

Long-term follow-up of infants in randomized clinical trials has been recommended to answer questions about the effect of ultrasound on human development. In a previous report, we found no association between routine ultrasonography in utero and poor school performance or dyslexia among 8 and 9-year-old children. Nor did we find any differences in vision and hearing at ages 4 and 7 years. The present report on the same children considers possible effects of routine ultrasonography on growth during childhood in a randomized controlled follow-up.
Ultrasonography and growth during childhood

MATERIALS AND METHODS

Two randomized controlled trials of ultrasonographic screening in pregnancy were carried out in the Norwegian cities Trondheim and Alesund in 1979–81. The study design and randomization method were identical. The pregnant women in Alesund were representative of the general population, whereas the study population in Trondheim included a higher proportion of low-risk pregnancies. The women were offered ultrasonographic examinations in the 19th and 32nd weeks of pregnancy. The same ultrasonic devices (ADR 2130, Tempe, Arizona, USA) were used at both sites. Those scanners produced lower intensities than do most scanners in obstetric use today. The median exposure time for the first routine scan in Alesund was 3 min.

Altogether, 2637 women were randomized into a screening group of 1335 women and 1302 controls. A total of 1244 screened and 1184 control children were eligible for follow-up after 8 years. Details about the study population have been reported elsewhere. Mothers of all eligible children were sent a questionnaire containing 66 closed-end questions together with a letter of information and a postage-paid return envelope. Return of the questionnaire was taken as informed consent for the child’s participation in the follow-up.

Two reminders were sent to non-responders. Norwegian children are regularly examined by physicians and specially trained public health nurses at maternal and child health centers. Visits take place when the child is 6 weeks, 3, 6 and 12 months old, and again at 2, 4 and 7 years of age. The recorded weight and height data from these examinations were retrieved for each child. Any discrepancies between those recorded measures were checked and corrected by use of the hospitals’ birth records. Discrepancies occurred in only 106 (5%) children.

Birth weight and height had been recorded for each child in the previous studies as well as in the records from the health centers. Any discrepancies between those recorded measures were checked and corrected by use of the hospitals’ birth records. Discrepancies were a minor problem and occurred in only 106 (5%) children.

Data were analyzed with Student’s t-test in the Statistical package for the Social Sciences, in a cross-sectional comparison between screened and control children at birth; at 3, 6 and 12 months; and at 2, 4 and 7 years of age. Analogous comparisons were done, stratifying the children according to whether their mothers reported regular smoking or not at the first antenatal examination. The aim was to test a hypothesis of no effect of routine ultrasonography in utero on growth during childhood.

Moreover, since weight and height measurements of children in the present study were correlated, a repeated-measures analysis of variance was considered to be another appropriate way to analyze the data. Since some of the measures were missing in the total population, we decided to restrict the repeated-measures analysis to a subsample which met certain specified criteria. Thus, we identified children who had been seen on all four occasions, i.e. at 1, 2, 4 and 7 years of age. They were not included if any of the visits took place outside a time window of ±2 standard deviations from the scheduled ones. Hence, children who had complete records of weight and height measures at birth; between 315 and 415 days; 1.7–2.3 years; 3.5–4.5 years; and 6.2–7.8 years of age were included in the sample, which consisted of 639 (58%) screened children and 562 (54%) controls. There were no obvious differences between all screened and control children for any of the background variables listed in Table 1. However, in the sample there was a greater proportion of children whose mothers had reported smoking at the first antenatal visit (p = 0.03). Weight and height measures were recorded at visits to the maternal and child health centers. We found no differences in the average ages of the two groups at the examination times (Table 2). A repeated-measures analysis of variance which employs the least-square-means method by use of the Generalized linear models procedure in SAS was used to model growth curves of children in the sample. To minimize the potential for confounding with smoking, which may have been introduced as a result of missing data in the sample, the repeated-measures analysis was performed separately for children of smoking and non-smoking mothers.

Table 1 Background variables in screened and control children for the whole group and subsample (see text for selection criteria)

<table>
<thead>
<tr>
<th>All children</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Screened</td>
</tr>
<tr>
<td></td>
<td>(n = 1107)</td>
</tr>
<tr>
<td>Mean maternal age at pregnancy (years)</td>
<td>26</td>
</tr>
<tr>
<td>Percentage of non-smoking mothers during pregnancy</td>
<td>61</td>
</tr>
<tr>
<td>Percentage of mothers with education of</td>
<td></td>
</tr>
<tr>
<td>6–9 years</td>
<td>28</td>
</tr>
<tr>
<td>10–12 years</td>
<td>50</td>
</tr>
<tr>
<td>&gt; 12 years</td>
<td>22</td>
</tr>
<tr>
<td>Percentage of families with</td>
<td></td>
</tr>
<tr>
<td>poor economy</td>
<td>3</td>
</tr>
<tr>
<td>medium economy</td>
<td>38</td>
</tr>
<tr>
<td>good economy</td>
<td>59</td>
</tr>
<tr>
<td>Percentage of male children</td>
<td>50</td>
</tr>
</tbody>
</table>
RESULTS

Of 1244 children in the screened group, five had died. The parents of 124 children did not return the questionnaires, while eight records from maternal and child health centers could not be traced. This left 1107 children to be studied. In the control group of 1184 children, one had died and 137 parents did not respond. Also, 13 records from health centers could not be retrieved. Hence, 1033 controls were studied.

The mean body weight and height at the time of the various health center visits are presented in Tables 3 and 4. At birth, screened children were on average 23 g heavier ($p = 0.3$) and equally as long as the controls. There were no statistically significant differences in weight or height between the two groups of children at any of the visits at the health centers.

In all, 39% of the screened and 36% of the control women reported smoking at their first antenatal visit ($p = 0.07$). Table 5 shows that babies of screened women who smoked were on average 39 g heavier at birth than babies of control women who also smoked. Moreover, a weight difference in favor of the screened group was present at each subsequent visit throughout the follow-up up to and including age 7 years. Among non-smokers, babies of screened women were on average 23 g heavier at birth than babies of controls. Throughout follow-up there was no systematic difference in weight between the groups. Weight differences appeared to fluctuate randomly around no difference between screened and control children (Table 5).

Among screened women in this study, the newborn babies of non-smoking women were 206 g heavier than babies of smoking women. In comparison, babies of non-smoking control women were 222 g heavier at birth than those of smokers.

In the repeated-measures analysis of variance in the subsample of children, we found a statistically significant difference in growth curves ($p = 0.02$) between screened and control children among mothers who reported smoking at their first antenatal visit. Contrary to this, there was no such difference in growth among children of non-smoking mothers ($p = 0.7$).

DISCUSSION

In this controlled follow-up of children whose mothers were randomly allocated to routine ultrasonographic screening during pregnancy, we found no statistically significant differences in body weight or height at birth; at 3, 6 and 12 months; and at 2, 4 and 7 years of age. Thus, the concern about ultrasound safety that has emerged from some animal studies, indicating reduced growth after ultrasound exposure, was not found valid in children in our study.
Our findings agree with the results reported by Lyons and co-workers, who found no statistically significant differences in body weight or height at birth; at 6 months; and at 1, 2, 3, 4, 5 and 6 years of age between ultrasound-exposed children and controls. Waldenström and co-workers reported a significantly higher mean birth weight among screened children in a randomized controlled trial which included nearly 5000 pregnancies. A similar Finnish trial could not confirm that finding, but a large proportion of ultrasound-exposed controls might have masked any effect that routine ultrasonography had on mean birth weight in that study.

It is well known that smoking reduces birth weight\(^{16,19}\), and it has been shown\(^9\) that mothers are strongly affected by watching their fetus on the screen. In fact, it has been suggested that the procedure may stimulate changes to a healthier lifestyle\(^9\). Waldenström and co-workers found mean birth weight to be 42 g higher \((p = 0.008)\) and fewer babies with a birth weight below 2500 g \((2.5\% \text{ versus } 4.0\%, p = 0.005)\) among the screened newborns in their randomized controlled trial\(^6\). In our study, babies of screened women had a mean birth weight 23 g above that of controls \((p = 0.34)\), and there were slightly fewer with a birth weight below 2500 g \((2.6\% \text{ versus } 3.0\%, p = 0.58)\). It is likely that the statistical power of our study was too low to unveil differences of this magnitude.

In the Swedish study, babies of screened women who smoked were on average 75 g \((p = 0.01)\) heavier at birth than babies of control women who smoked\(^6\). Newborns of screened non-smokers were in comparison 26 g \(\text{not significant}\) heavier than babies of non-smoking controls. The authors speculate that the greater differences in birth weight among newly born babies born to smokers may reflect that screened women could have reduced their smoking during pregnancy to a greater extent in response to watching their fetus on the screen\(^6\). In our study, babies of screened smokers were 39 g heavier at birth than controls, compared to a 23 g difference between screened and control babies of non-smokers. We also found that, among screened women, babies of non-smokers were on average 206 g heavier that babies of smokers, and among controls, the analogous difference was 222 g.

If screening induces smoking cessation during pregnancy, two observations might be anticipated. First, among neonates of women who reported smoking at their first antenatal visit, one would expect screened children to be heavier at birth than unscreened children. Second, in the screened group one would expect the difference in birth weight between babies of smokers and non-smokers to be smaller than among babies in the control group. Both anticipated observations were present in our study, but the small weight differences suggest that the effect of screening on smoking cessation could not have been strong, if at all present.

A possible explanation for the greater differences in birth weight in the Swedish study may be reflected in the technological development of ultrasound devices that took place during the 6-year interval between our study and the Swedish one. The newer ultrasound devices considerably improved the image of the fetus on the screen. Also, in the Swedish study, the scanning was done in a 'comfortable' atmosphere and there was 'usually time for parents to take a good look at the fetus\(^6\). Maybe these factors had a stronger impact on smoking than the routine examinations done with a booking interval of 15 min and a median ultrasound exposure time of 3 min in the Norwegian studies.

In the cross-sectional analyses of growth during childhood, we found no statistically significant differences between screened and control children. However, a separate analysis of children, stratified according to their mothers' smoking status at the first antenatal visit, revealed that, among smokers, screened children were consistently heavier than controls at each weight measurement during follow-up. No such difference was noted among children of non-smoking mothers (Table 5). The repeated-measures analysis of a subsample of children confirmed this finding. The growth curve of screened children whose mothers had reported smoking was statistically significantly different from that of control children. Thus, our data indicate that routine ultrasonography may initiate a process among smoking mothers which leads to an improved growth during pregnancy and childhood. Whether or not this increased growth is mediated through a cessation of smoking is not clear from our data.

A strong feature of our study is that the randomized controlled design rules out many of the possible biases that might influence an association between routine ultrasonography and later growth. Thus, the cross-sectional analyses of growth during childhood were apparently free from bias. However, since the sample

---

**Table 5** Mean body weight (g) for all screened and control children of mothers who reported smoking \((n = 768)\) or non-smoking \((n = 1276)\) at the first antenatal visit

<table>
<thead>
<tr>
<th>Age</th>
<th>Smokers</th>
<th>Controls</th>
<th>Mean difference</th>
<th>p value</th>
<th>Non-smokers</th>
<th>Controls</th>
<th>Mean difference</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Screened</td>
<td>(n = 423)</td>
<td></td>
<td></td>
<td>Screened</td>
<td>(n = 650)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth</td>
<td>3 471</td>
<td>3 432</td>
<td>39</td>
<td>0.31</td>
<td>3 677</td>
<td>3 654</td>
<td>23</td>
<td>0.43</td>
</tr>
<tr>
<td>3 months</td>
<td>6 123</td>
<td>6 071</td>
<td>52</td>
<td>0.48</td>
<td>6 365</td>
<td>6 356</td>
<td>9</td>
<td>0.88</td>
</tr>
<tr>
<td>6 months</td>
<td>7 998</td>
<td>7 917</td>
<td>81</td>
<td>0.25</td>
<td>8 094</td>
<td>8 165</td>
<td>-71</td>
<td>0.17</td>
</tr>
<tr>
<td>1 year</td>
<td>10 252</td>
<td>10 128</td>
<td>124</td>
<td>0.15</td>
<td>10 201</td>
<td>10 250</td>
<td>-49</td>
<td>0.42</td>
</tr>
<tr>
<td>2 years</td>
<td>12 816</td>
<td>12 703</td>
<td>113</td>
<td>0.33</td>
<td>12 935</td>
<td>12 911</td>
<td>24</td>
<td>0.78</td>
</tr>
<tr>
<td>4 years</td>
<td>17 393</td>
<td>17 284</td>
<td>109</td>
<td>0.53</td>
<td>17 415</td>
<td>17 394</td>
<td>21</td>
<td>0.86</td>
</tr>
<tr>
<td>7 years</td>
<td>24 293</td>
<td>23 808</td>
<td>485</td>
<td>0.13</td>
<td>24 097</td>
<td>24 393</td>
<td>-296</td>
<td>0.17</td>
</tr>
</tbody>
</table>
included a higher proportion of smoking mothers \((p = 0.03)\) among screened than among control children, missing data may have introduced a bias in the repeated-measures analysis of variance. Besides, differences in body weight were consistently greater in the sample than in the population as a whole at all the health center visits (Table 3). With this background, we found it particularly important to analyze the data by the mother's smoking status at their first antenatal visit. Unfortunately, we have no further information about smoking habits after the mothers' first antenatal visit. More smoking data, e.g. after birth, might have enhanced more comprehensive analysis of its impact on later growth of the children.

Data in the present study were analyzed according to the 'intention to treat' principle. Details of actual ultrasound exposure of the study children have already been reported\(^1\). In all, 3% of the children who were offered screening were never exposed to ultrasound \textit{in utero}, whereas 7% of the controls were scanned several times. Between gestational weeks 16 and 22, only 5% of the controls were scanned, while 8% of the children in the screening group were not. Thus, the overlap in ultrasound exposure between the randomized groups is not likely to have played an important role in our findings.

To conclude, we have found that children whose mothers were invited to take part in a routine ultrasonography program during pregnancy showed no differences in mean body weight or height compared to control children in a cross-sectional analysis of growth during childhood. However, a repeated-measures analysis of variance in a subsample of the children indicated that growth from birth to 7 years of age differed significantly between screened and control children born to mothers who reported smoking at the first antenatal visit.

ACKNOWLEDGEMENTS

The study has been supported by the Norwegian Research Council for Science and the Humanities (NAVF), grant no 351.87/001. Ms Nancy Lea Eik-Nes revised the manuscript.

REFERENCES


10 Ultrasound in Obstetrics and Gynecology

Salvesen et al.
59. Arne Z. Henriksen: STUDIES ON CONSERVED ANTIGENIC DOMAINS ON MAJOR OUTER MEMBRANE PROTEINS FROM ENTEROBACTERIA. 1990.
60. Steinar Westin: UNEMPLOYMENT AND HEALTH: Medical and social consequences of a factory closure in a ten-year controlled follow-up study. 1990.