WASHINGTON STATE HEALTH CARE AUTHORITY

Ultrasonography (Ultrasound) in Pregnancy

Health Technology Assessment

Date: September 22nd, 2010

Health Technology Assessment Program
676 Woodland Square Loop SE
P.O. Box 42712
Olympia, WA 98504-2712
http://www.hta.hca.wa.gov

Ultrasonography (Ultrasound) in Pregnancy

A Health Technology Assessment

Prepared for Washington State Healthcare Authority

FINAL REPORT – September 22, 2010

Acknowledgement

This report was prepared by:

HAYES, INC.
157 S. Broad Street Suite 200
Lansdale, PA 19446
P: 215.855.0615 F: 215.855.5218

TABLE OF CONTENTS

ULTRASONOGRAPHY (ULTRASOUND) IN PREGNANCY

EXECUTIVE SUMMARY

Medical Background

Ultrasonography, or simply ultrasound (US), is used in prenatal care for monitoring normal fetal development and maintenance of maternal well being. During the first trimester (6 days of gestation up to 13 weeks) an US may be performed for a variety of reasons, including estimation of gestational age diagnosis, evaluation of multiple gestations, or measurement of markers for fetal aneuploidy (abnormal chromosome number). In the second trimester (between 16 weeks and 22 weeks), US is performed to assess anatomical fetal growth and development (fetal anatomical survey), screen for markers for fetal aneuploidy, estimate fetal weight, detect and evaluate gynecological abnormalities, and detect fetal anatomical abnormalities. In the United States, routine US is not typically performed in the third trimester unless the pregnancy is considered a high-risk pregnancy or a specific indication has developed.

There are several risk factors that impact pregnancy and its management. Although low- and high-risk pregnancies are not precisely defined, conditions including age ≥ 35 years at delivery, diabetes mellitus, asthma, hypertension, or previous pregnancy loss are commonly considered risk factors. Additionally, several conditions that may arise during pregnancy such as preeclampsia, fetal intrauterine growth restriction (IUGR), premature rupture of membranes, multiple pregnancy, preterm labor, and postterm pregnancy increase maternal and perinatal morbidity and mortality and thus require accurate evaluation.

An important objective in pregnancy management is prevention of preterm birth. Preterm birth is a leading cause of neonatal mortality and morbidity and has been increasing. Because clinical criteria such as obstetric history have not been found to reliably predict preterm birth, assessment of cervical length by transvaginal US (TVU) has been tested as an alternative screening method, either in women showing signs and symptoms of preterm labor or as a means of surveillance in women who have a history of preterm birth. If short cervix is confirmed, the clinician can administer treatment to delay birth and to prevent perinatal respiratory distress or, in the case of surveillance based on history, to reinforce the cervix.

Policy Context

Increasing use of US in pregnancy, and questions about its actual clinical utility, are of concern to healthcare decision makers. According to data collected in the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey (NHAMCS) of the Centers for Disease Control (CDC) for the years 1995 to 2000, 2005, and 2006, the use of US in pregnancy has grown substantially. The Food and Drug Administration (FDA) has approved US for evaluating and monitoring pregnancy, fetal growth, and fetal health. The FDA considers US to be a safe technology. The FDA does, however, consider “keepsake videos” to be an unapproved use of US (CDRH, 2010). The Centers for Medicare & Medicaid Services (CMS) has approved US for these uses related to management of pregnancy: pregnancy sonography (details not provided) pregnancy diagnosis, fetal age determination, fetal growth rate determination, placenta localization, molar pregnancy diagnosis, ectopic pregnancy, passive testing (antepartum monitoring of fetal heart rate in resting fetus), and guidance of

amniocentesis for purposes of testing for chromosomal abnormality. CMS has also approved Doppler US for arterial flow study.

Practice Guidelines

Fair-quality guidelines from ACOG, ACR, and ICSI are consistent with each other and with the literature in describing US as a reasonably safe procedure that accurately provides a wealth of information about pregnancy status and fetal health. Although the guidelines from ACOG allude to the questionable relationship between routine use of US and maternal and fetal outcomes, recommendations were not formed with this in mind. The ICSI guidelines take into consideration the lack of evidence supporting routine use of US in low-risk pregnancy, especially in late pregnancy, but do not fully address the use of US in high-risk pregnancy. None of the guidelines addresses the use of US to monitor cervical length. None of the guidelines considers evidence pertaining to the long-term effects on child growth and development, differential effectiveness and safety, or cost-effectiveness.

Findings

The following selections were made from databases of systematic reviews and from a systematic search of MEDLINE and EMBASE: six systematic reviews of the effectiveness of routine US in high-risk or low-risk pregnancy, a systematic review of the safety of routine US in low-risk pregnancy, a systematic review of US performed in the emergency department for evaluation of possible ectopic pregnancy, and five additional primary studies published after the search time frame observed by the systematic reviews. Five of the systematic reviews included meta-analysis. Information about accuracy was obtained primarily from narrative and systematic review articles and was not critically appraised.

Accuracy: The selected literature suggests that US has variable accuracy, depending on the target condition. As a screening tool, it is often combined with other tests. Review articles report sensitivities of 40% to 99%, but information about specificity, positive predictive value, and negative predictive value was not readily available.

Effectiveness in High-Risk Pregnancy: The evidence provides some support for the use of Doppler US to monitor high-risk patients and the use of TVU to identify patients in need of prophylactic treatment because of imminent risk of preterm birth. The evidence does not support the use of TVU surveillance in women with a history of preterm birth. In general, evidence pertaining to these indications is of very low to low quality. Lack of standard protocols for intervention when DUS is abnormal or TVU shows short cervical length hampers the generalizability of pooled results to particular settings.

Effectiveness in Low-Risk Pregnancy, Early Screening: According to moderate- to high-quality evidence, routine US in early pregnancy (< 24 weeks) does not change patient management, substantially alter delivery modes, or improve health outcomes, at least not in high-resource settings. These various findings might not apply to low-resource settings where perinatal mortality is high and not as likely to be attributable to fetal abnormality and might not apply to all strategies for US timing and follow-up intervention. High-quality evidence shows that routine US doubles the rate of abortion for fetal anomaly, but the estimated absolute increase is 0.10 percentage point.

Effectiveness in Low-Risk Pregnancy, Late Screening: Low- to moderate-quality evidence has not shown routine US in late pregnancy (> 24 weeks) to change patient management, affect delivery mode, or improve health outcomes.

Safety: Moderate-quality evidence for major outcomes has shown US to be a reasonably safe procedure with no serious short-term adverse effects. Evidence of mixed quality suggests no general impact on developmental outcomes after birth but further research, particularly with respect to neurological development, is needed to allow definite conclusions. The applicability of most of the safety evidence is diminished by the fact that most studies were using older, weaker machines. There is also very little evidence on the safety of US performed in the first and third trimesters.

Differential Effectiveness and Safety: The evidence pertaining to differential effectiveness and safety does not address all potentially useful comparisons and is of variable quality. Routine US performed between 14 weeks and 24 weeks (second trimester) is most likely to detect multiple births (low-quality evidence) and to reduce the frequency of induction of labor (moderate quality), compared with US at other gestational ages. However, high-quality evidence shows no differential effect by gestational age on perinatal mortality. Other comparisons were derived from low or very low evidence and generally showed no effect.

Cost Implications and Cost-Effectiveness: No definitive statements can be made about the cost or cost-effectiveness of US in pregnancy because assumptions of impact on outcomes have not been based on comprehensive reviews of current evidence, costs were not trial-based, and three of four cost-effectiveness studies made no comparison with a strategy in which there was no routine US screening. There is preliminary evidence that routine use of second-trimester US to screen for fetal anomaly may reduce short- and long-term costs. There is also preliminary evidence that universal TVU screening for short cervix may prevent preterm birth and save direct short- and long-term costs but only in low-risk pregnancies (no previous preterm birth) and only as an add-on to routine second trimester US fetal assessment. There have been no economic evaluations of US in other types of high-risk pregnancy. Performance of US to rule out ectopic pregnancy may be less costly if performed in the emergency department.

MEDICAL BACKGROUND

Ultrasonography, or simply ultrasound (US), is used in prenatal care for monitoring normal fetal development and maintenance of maternal well being. During the first trimester (6 days of gestation up to 13 weeks) an US may be performed for a variety of reasons. US can help confirm intrauterine pregnancy, placental location, or fetal viability in terms of fetal cardiac activity; estimate gestational age in order to set an expected delivery date; diagnose and evaluate multiple gestations; evaluate any gynecologic abnormalities; evaluate pelvic pain; or measure nuchal translucency, which is an ultrasonographic marker for fetal aneuploidy (abnormal chromosome number). Other uses of US include evaluation of suspected gestational trophoblastic disease, and screen for any uterine, adnexal, and cervical malformations that could influence prenatal management (ACOG, 2004; ACOG, 2009).

In the first trimester, US can be performed either transabdominally, transvaginally, or both. If a transabdominal examination is not definitive, a transvaginal US or transperineal scan is indicated (ACOG, 2004; Bahado-Singh and Cheng, 2004; AIUM, 2007; Cargill et al., 2009). For the estimation of gestational age in the first trimester, the crown–rump length is a more accurate indicator than mean gestational sac diameter (Lynch and Zhang, 2007).

In the second trimester (between 16 weeks and 22 weeks), US is useful for an expanded set of purposes: to assess anatomical fetal growth and development (fetal anatomical survey); screen for markers for chromosomal trisomies, including Down ’s syndrome, Patau syndrome, and Edward syndrome; ascertain fetal presentation, amount of amniotic fluid, and placental location and its relationship to the internal cervical os; estimate fetal weight; and evaluate uterine, adnexal, and cervical abnormalities. During the second trimester US, the gestational age is estimated by comparing biometric measures to reference data. Fetal biometric measurements that can be obtained from second trimester US include biparietal diameter, head circumference, abdominal circumference, and femur diaphysis length; these allow estimation of gestational age and fetal weight (ACOG, 2004; ACOG, 2009; Cargill et al., 2009).

The sonographic fetal anatomical survey, a key element of a second trimester US, includes assessment or detection of these anatomic structures or malformations (ACOG, 2004; Cargill et al., 2009):

- Skull and brain (skull shape and cranial ossification, lateral ventricles, choroid plexus, cerebellum)
- Face (visualization of orbit and viewing fetal profile, looking for nasal bone), neck (nuchal translucency measurement)
- Spine (examination of overlying skin and neural tube in longitudinal and transverse planes)
- Heart (heart rhythm, position, axis, four-chamber view, and examination of great vessels)
- Stomach (existence in left upper abdomen)
- Abdominal wall (examining abdominal wall and insertion of umbilical vessels)
- Kidney (existence, size and shape, tissue texture)
- Urinary bladder (existence, size, shape)
- Extremities (examining proximal and distal long bones, looking for posture of extremities)
- Cystic hygroma (lymphatic tumor)

In the United States, routine US is not typically performed in the third trimester in low-risk pregnant women. Third-trimester US may be performed in patients with a high-risk history or in patients who

develop a specific indication that requires investigation. However, US protocols and standards vary between continents; in Europe, it is a common practice to perform third trimester US during routine prenatal care (Le Ray and Morin, 2009). There are several risk factors that impact pregnancy and its management. Although low- and high-risk pregnancies are not precisely defined, maternal conditions such as age ≥ 35 years at delivery, prior classical C-section, cervical insufficiency, systemic disease, psychiatric illness, and substance abuse can have significant detrimental effects on pregnancy outcomes. Women who have experienced recurrent pregnancy loss are also considered to be at high risk. Additionally, several conditions that may arise during pregnancy such as preeclampsia, eclampsia, pyelonephritis, fetal intrauterine growth restriction (IUGR), polyhydramnios, oligohydramnios, antepartum hemorrhage, premature rupture of membranes, placental placenta accerata, multiple pregnancy, preterm labor, and postterm pregnancy increase maternal and perinatal morbidity and mortality (ACOG, 2007; Alfirevic et al., 2010).

The types of US examinations commonly performed are standard (or basic), limited, and specialized (or detailed). The standard US determines gestational age, fetal number, fetal viability, and placental location. A limited examination can be performed in any trimester to evaluate interval growth, estimate amniotic fluid volume, evaluate the cervix, and assess the presence of cardiac activity. A detailed or targeted anatomic US is performed when an anomaly is suspected on the basis of history, laboratory abnormalities, or the results of either the limited or standard examination. Four-dimensional (4D) US, or real-time three-dimensional (3D) US can create many images per second; the result is such that either the effect of moving the probe or fetal motion can be observed in three dimensions. This type of US is not routinely used during pregnancy. Other types of examinations to assess fetal well being in high risk populations are fetal Doppler US (DUS), biophysical profile (BPP), amniotic fluid assessment, cardiotocograph (CTG), fetal echocardiography, or additional biometric measurements. The BPP consists of a fetal heart rate tracing and US to measure four parameters: fetal body movements, fetal breathing movements, amniotic fluid index, and nonstress test. The nonstress test involves monitoring fetal heart tracing for 20 minutes to look for normal variability. DUS of the fetal and umbilical vessels is intended to detect abnormal flow patterns in fetal circulation and thus has the potential to predict poor fetal outcome. Improvements in US devices have allowed increasingly extensive imaging of the fetal circulatory system. Utero-placental DUS in the first trimester can determine if the appropriate physiologic changes are taking place in the uterine arteries. Early intervention can reduce the risk of adverse perinatal outcomes such as respiratory distress syndrome, as well as maternal problems such as preeclampsia or diabetes mellitus. Imaging technologies such as magnetic resonance imaging (MRI) and computed tomography (CT) are also being utilized for diagnosis of maternal disorders during pregnancy. Routine use of MRI during pregnancy is questionable during early pregnancy. MRI is being used for detection and localization of neoplasms of the chest, abdomen, and pelvis in pregnancy; detection of pelvic and vena caval thrombosis, leading to pulmonary embolism; and evaluation of right lower quadrant pain, specifically for diagnosis of appendicitis (Kurjak et al., 2002; Oto et al., 2007; Pedrosa et al., 2007; Pugash et al., 2008; ACOG, 2009; Pedrosa et al., 2009; Alfirevic et al., 2010; Stampalija et al., 2010).

An important objective in pregnancy management is prevention of preterm birth, defined by the World Health Organization (WHO) as birth between 20 weeks or and 37 weeks. In the United States, preterm birth occurred in 12.8% of pregnancies in 2006 and has increased more than 20% over the previous 10 years. Preterm birth is a leading cause of neonatal mortality and morbidity. Because clinical criteria such as obstetric history have not been found to reliably predict preterm birth, assessment of cervical length by transvaginal US has been tested as an alternative screening method, either in women showing signs

and symptoms of preterm labor or as a means of surveillance in women who have a history of preterm birth. Short cervical length has been shown to be associated with greater risk of preterm birth. Cervical length measurement can be combined with measurement of fetal fibronectin (FFN) to improve accuracy. Cerclage and bed rest can be used for prevention of preterm labor. Cerclage involves reinforcing a compromised cervix with sutures; the goal is to prevent the cervix from opening prematurely. A new option for preventing preterm birth, progesterone, has become available in recent years and adds to the potential utility of US surveillance for short cervix (Ness et al., 2007; Berghella et al., 2009; Simcox et al., 2009; Cahill et al., 2010).

In addition to serving as a screening tool in asymptomatic pregnancy, transvaginal US may be used to help reduce the false positive rate for diagnosis of preterm labor (Alfirevic et al., 2007). One estimate suggests that 90% of women with symptoms suggesting preterm labor will not deliver within 7 days, regardless of treatment; 75% will deliver at term (Ness et al., 2007). Pregnant women who have clinical symptoms and/or signs of being in preterm labor are typically given tocolytic drugs to delay labor and corticosteroids to treat the fetal lung immaturity in time for birth and avoid infant respiratory distress syndrome. Inaccurate diagnosis of preterm labor increases the risk of infection among pregnant women due to repeated pelvic examinations. Additionally, the fetal exposure to unnecessary corticosteroids could have very serious consequences (Alfirevic et al., 2007; Ness et al., 2007).

Policy Context

Increasing use of US in pregnancy, and questions about its actual clinical utility, are of concern to healthcare decision makers. According to data collected in the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey (NHAMCS) of the Centers for Disease Control (CDC) for the years 1995 to 2000, 2005, and 2006, the use of US in pregnancy has grown substantially. These trends have been observed (Siddique et al., 2009):

- The average number of US exams per pregnancy increased from 1.5 (1995 to 1997) to 2.7 (2005 to 2006). Statistical analysis ruled out higher prevalence of high-risk pregnancy as an explanation, and increases occurred in both low-risk pregnancy (from 1.3 to 2.1) and in high-risk pregnancy (from 2.2 to 4.2).
- After adjusting for demographic and risk category, Medicaid status as opposed to private insurance was not associated with the odds of receiving an US exam.
- US exams were more common in the Northeast than in the Midwest, South, or West. When the Northeast was considered the reference, the odds of an US were significantly reduced for patients in the Midwest and the South but not for patients in the West.

The Food and Drug Administration (FDA) has approved US for evaluating and monitoring pregnancy, fetal growth, and fetal health. The FDA considers US to be a safe technology. The FDA does not, however, consider “keepsake videos” to be an unapproved use of US (CDRH, 2010). The Centers for Medicare & Medicaid Services (CMS) has approved US for these uses related to management of pregnancy: pregnancy sonography (details not provided), pregnancy diagnosis, fetal age determination, fetal growth rate determination, placenta localization, molar pregnancy diagnosis, ectopic pregnancy, passive testing (antepartum monitoring of fetal heart rate in resting fetus), and guidance of amniocentesis for purposes of testing for chromosomal abnormality. CMS has also approved Doppler US for arterial flow study.

Washington State Agency Data

The following data is provided by the Washington State agencies on their utilization and cost information.

Figure 1: UMP /PEP/DSHS combined data

Year	2006	2007	2008	2009	Overall*
Ultrasound Count	50,584	52,654	56,356	40,713	**200,307**
Pregnancy Count	19,111	19,064	19,547	12,647	**59,653**
Avg # of Ultrasounds/ Pregnancy/Year	2.65	2.76	2.88	3.22	**3.36**
Total Cost of Ultrasounds	$5,049,915	$5,146,072	$5,676,992	$4,193,088	**$20,066,067**
Average Cost of Ultrasounds/ Pregnancy/Year	$264	$270	$290	$332	**$336**

*Pregnancies are double counted when they extend into a second year. Overall costs and counts (last column) reflect the number of individual pregnancies, so are more accurate.

Figure 2a –Ultrasound counts by Pregnancy, UMP 2006-2009 (Aetna excluded)

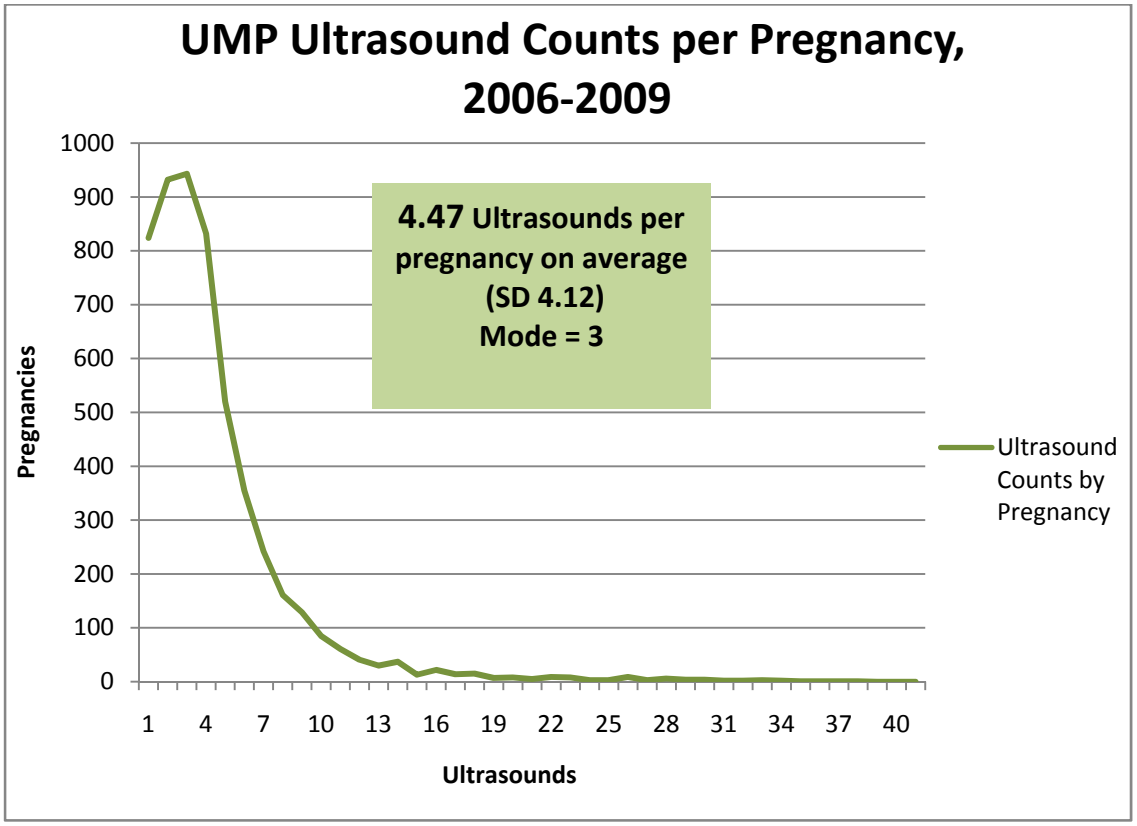


Figure 2b – DSHS Ultrasound counts by Pregnancy, 2006-2009

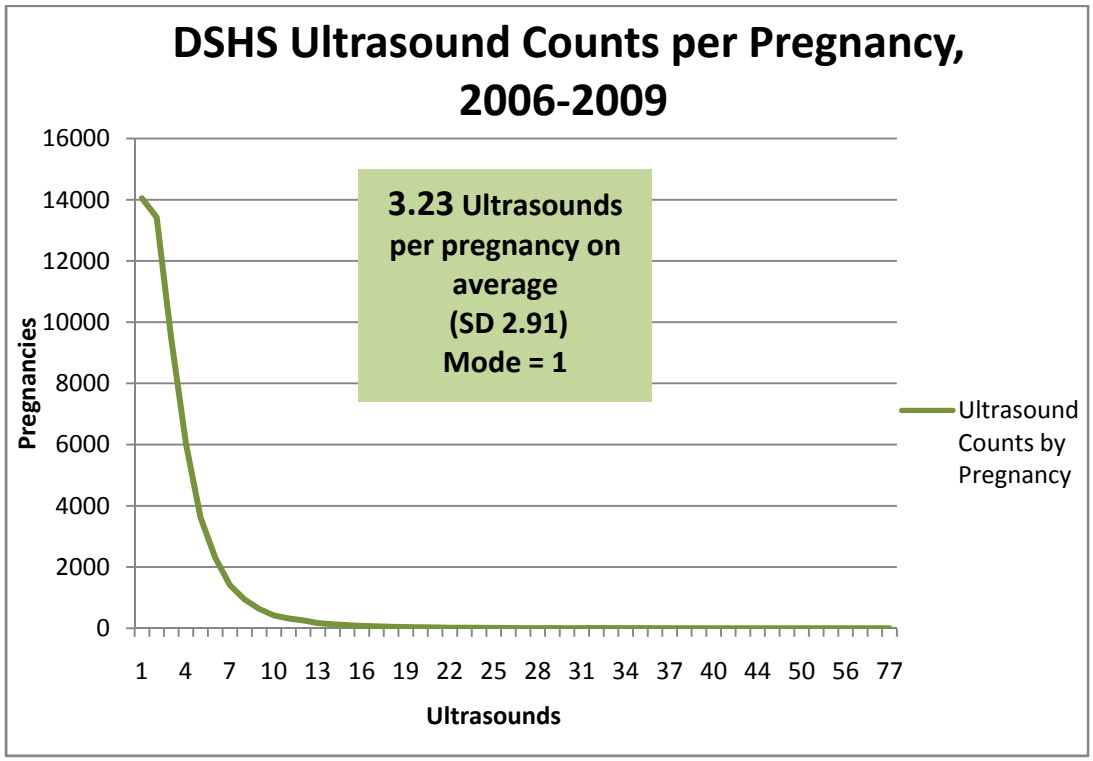


Figure 3a: Average Ultrasound Count per Pregnancy by Age Group – UMP 2006-2009

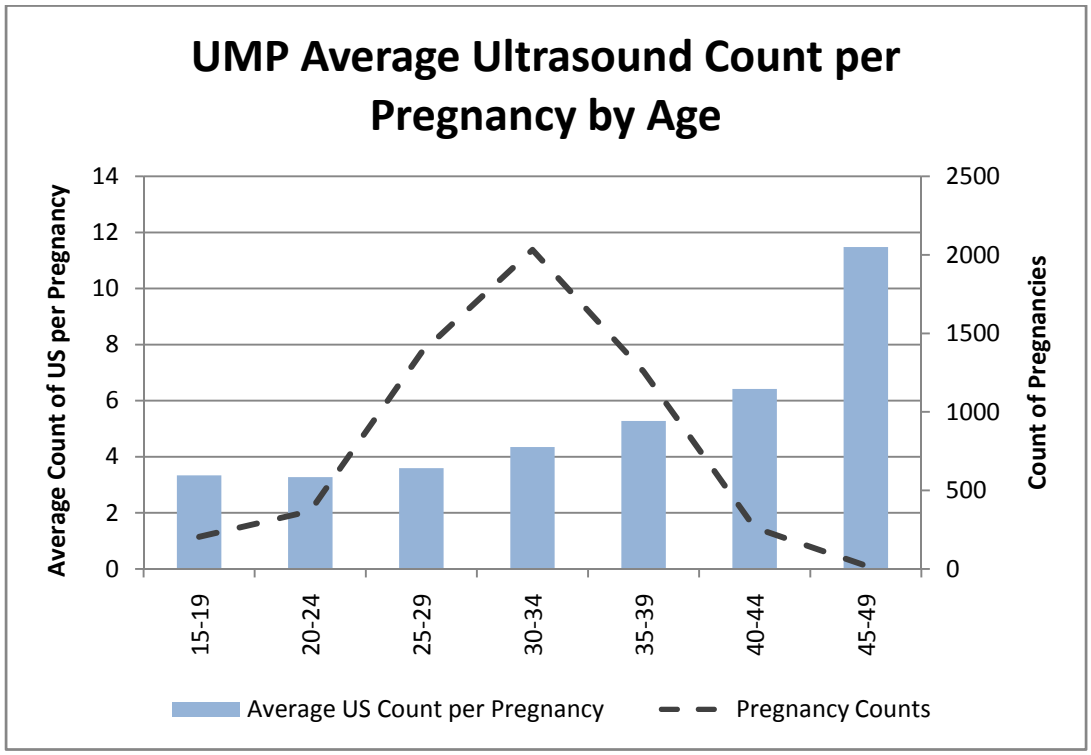


Figure 3b: Average Ultrasound Count per Pregnancy by Age Group – DSHS 2006-2009

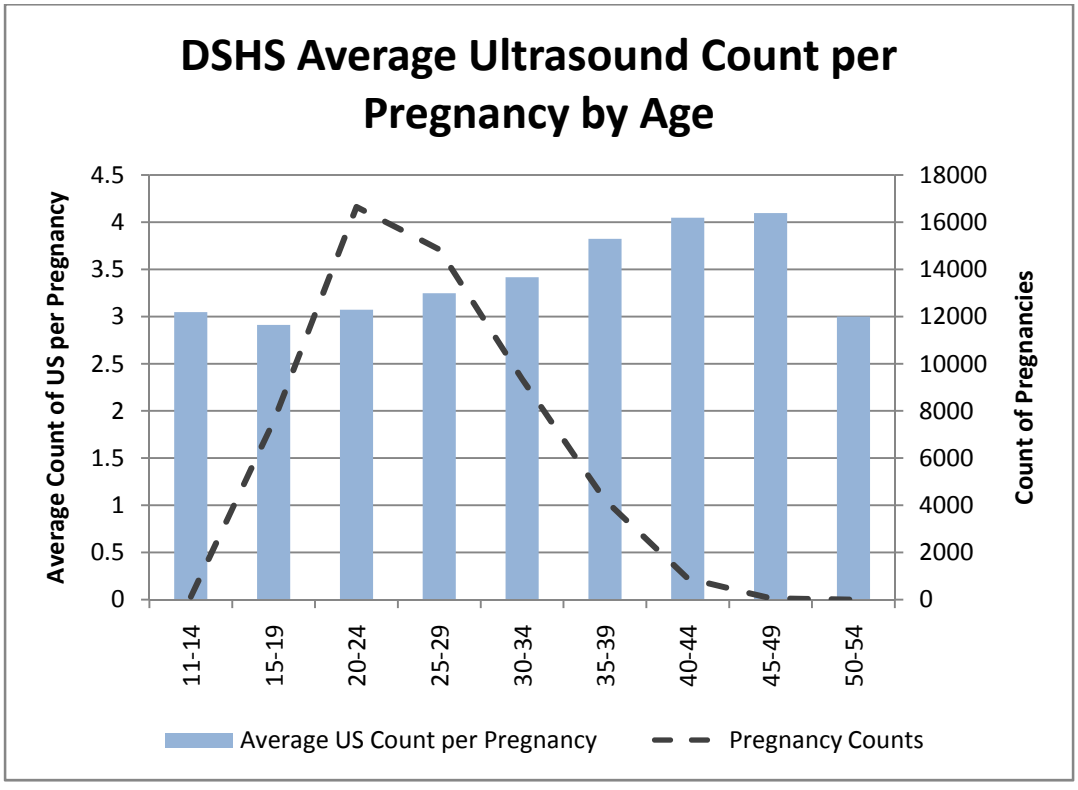
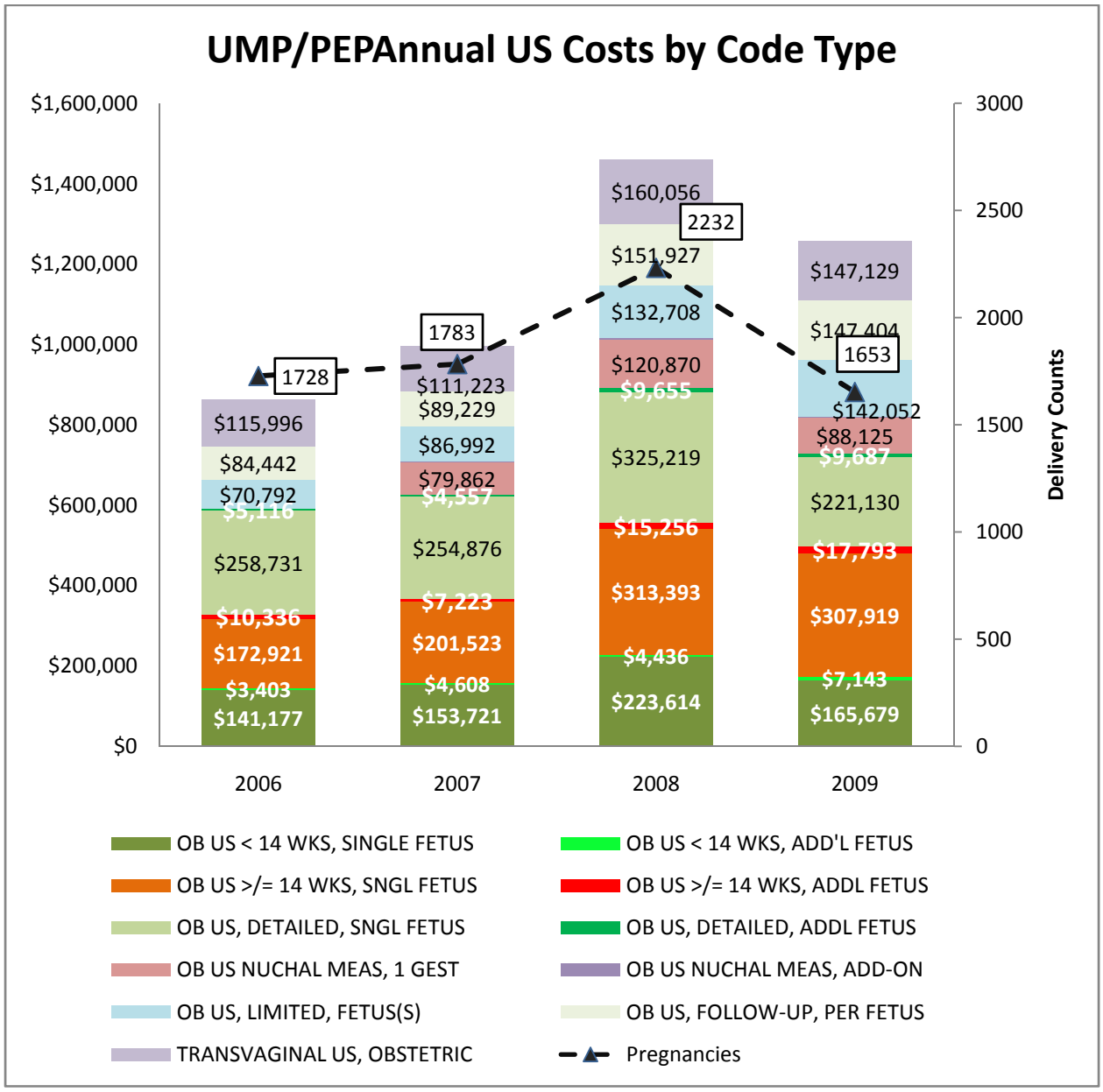


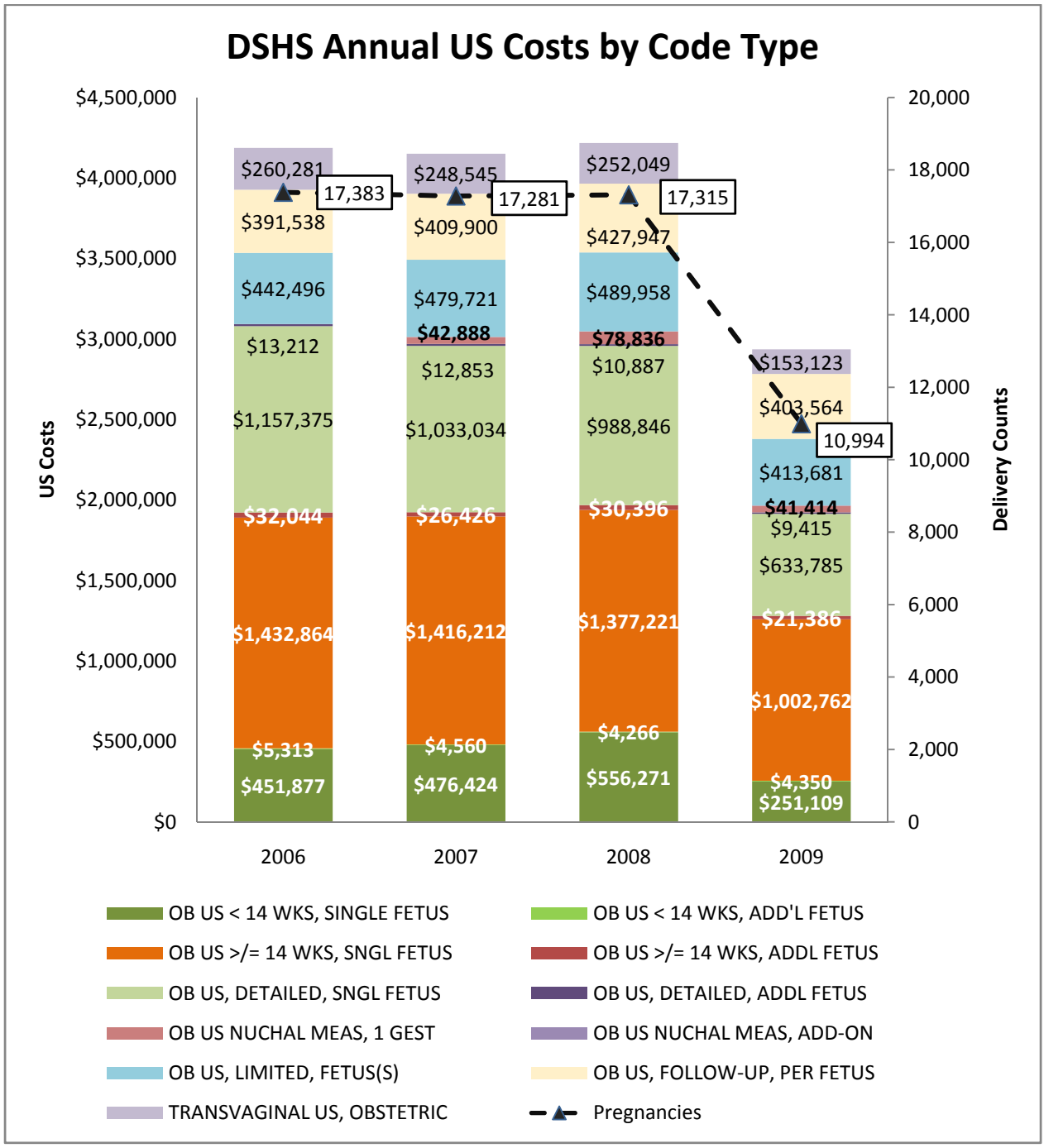
Figure 4a: Annual US Costs by Code/ Type, UMP/PEP 2006-2009



NOTE: The drop in payments between 2008 and 2009 occurred in all agency data, and may be explained by a drop in deliveries (see super-imposed line) due to economic pressures.

The color sections starting at the bottom of each bar correspond to the legend top left, then the top right, second left, etc. Some color sections are too small to be seen on this size chart. For readability, only sections that are visible have $ figures displayed.

Figure 4b: Annual US Costs by Code/ Type, UMP 2006-2009



NOTE: DSHS data included only US in completed pregnancies, and therefore both US and pregnancies are undercounted in 2009 data.

The color sections starting at the bottom of each bar correspond to the legend top left, then the top right, second left, etc. Some color sections are too small to be seen on this size chart. For readability, only sections that are visible have $ figures displayed.

Figure 5a: US Counts by Trimester, UMP 2006-2009

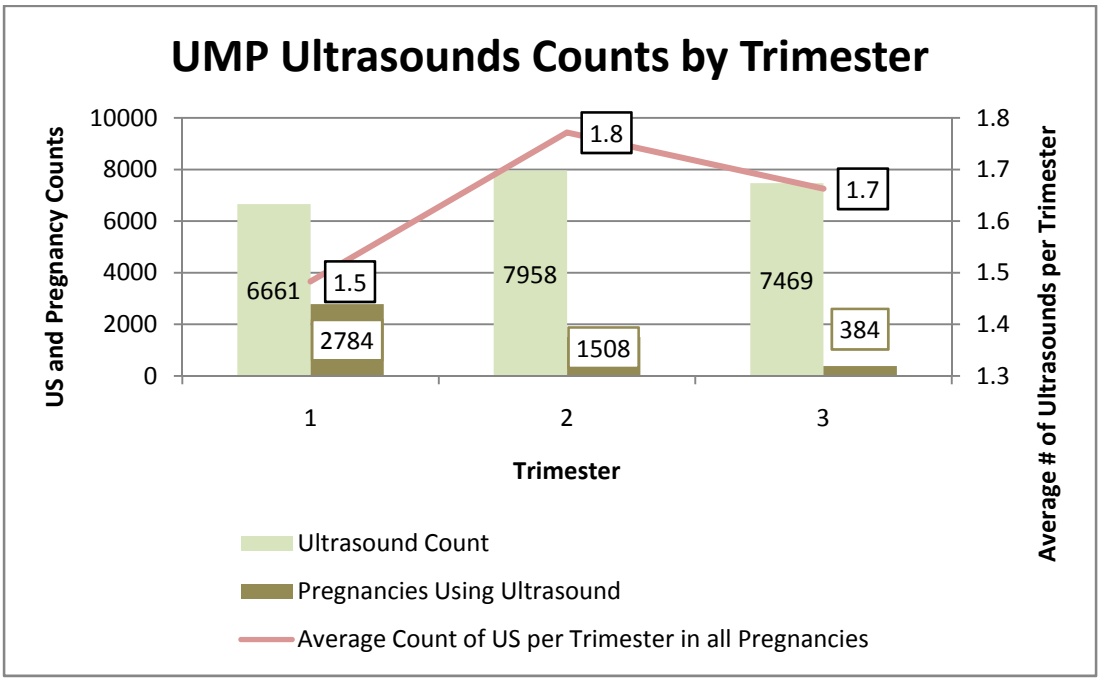


Figure 5b: US Counts by Trimester, DSHS 2006-2009

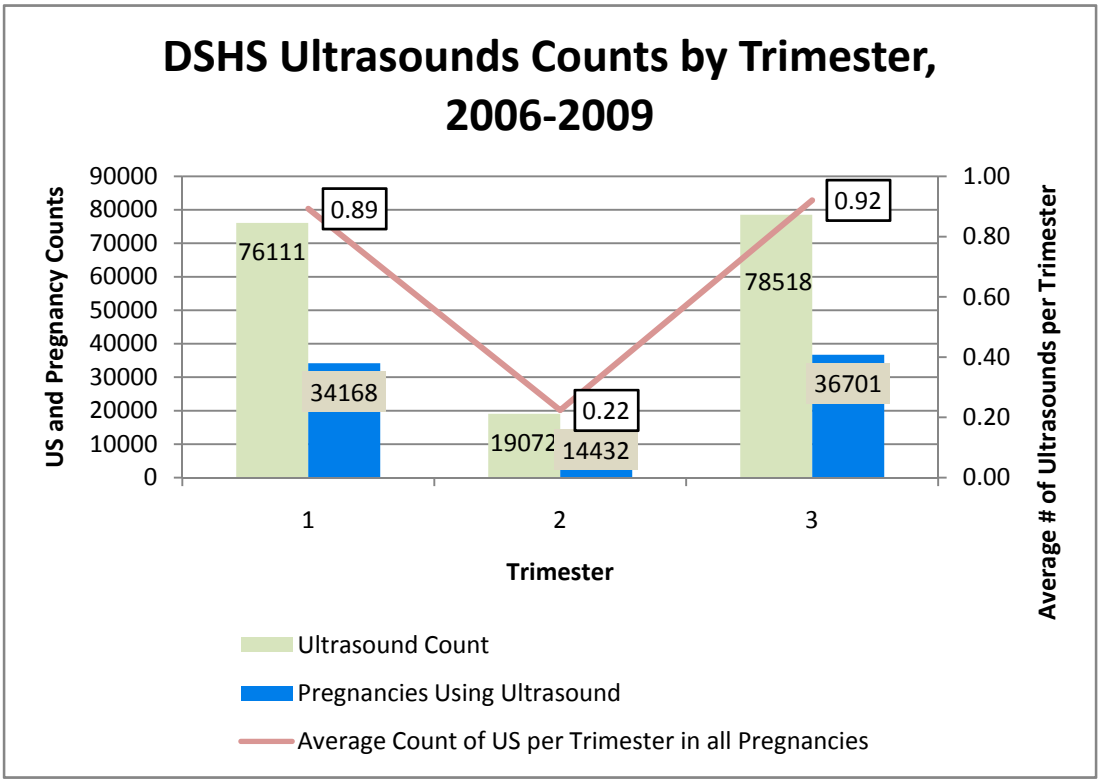
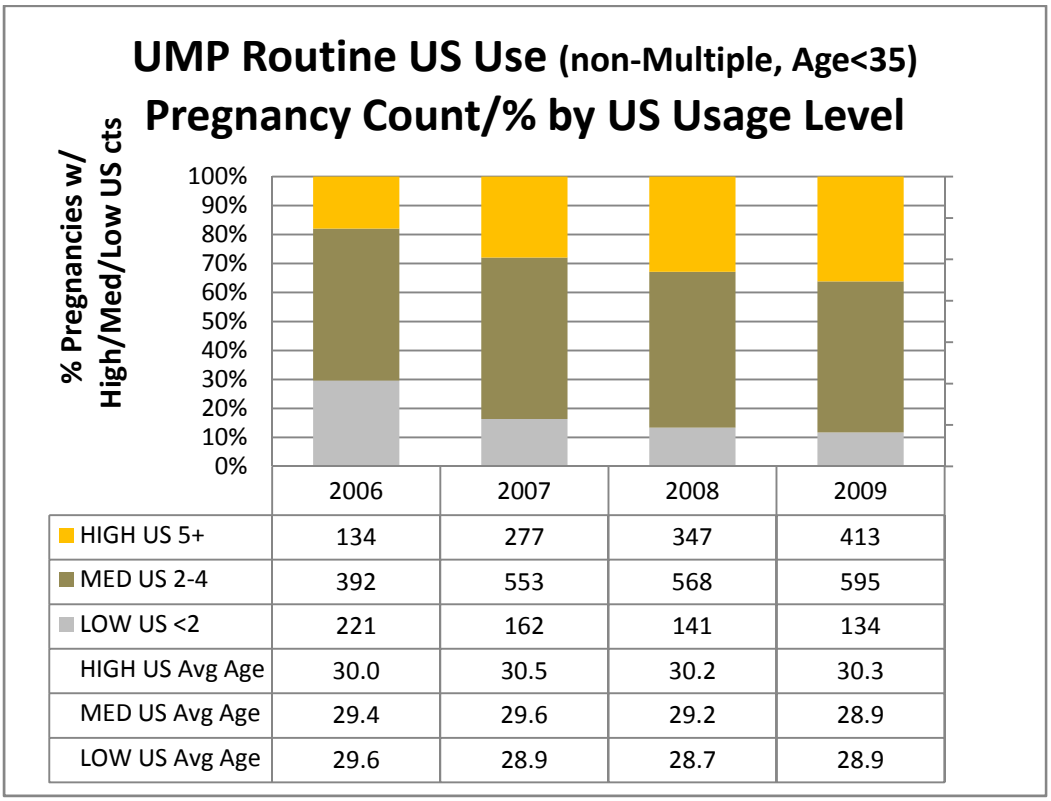
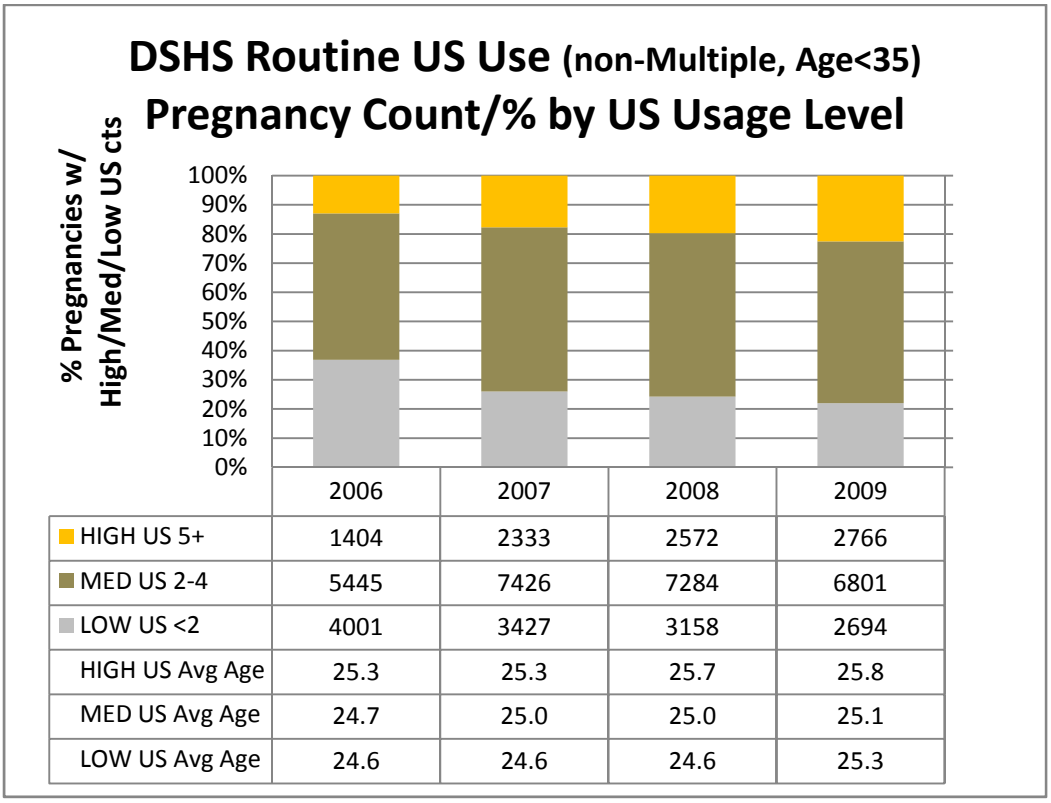


Figure 6a. Routine US Use by Usage Level, UMP 2006-2009



	2006	2007	2008	2009
HIGH US 5+	134	277	347	413
MED US 2-4	392	553	568	595
LOW US <2	221	162	141	134
HIGH US Avg Age	30.0	30.5	30.2	30.3
MED US Avg Age	29.4	29.6	29.2	28.9
LOW US Avg Age	29.6	28.9	28.7	28.9

Figure 6b. DSHS Routine US Use by Usage Level



	2006	2007	2008	2009
HIGH US 5+	1404	2333	2572	2766
MED US 2-4	5445	7426	7284	6801
LOW US <2	4001	3427	3158	2694
HIGH US Avg Age	25.3	25.3	25.7	25.8
MED US Avg Age	24.7	25.0	25.0	25.1
LOW US Avg Age	24.6	24.6	24.6	25.3

Related Medical Codes		
Ultrasound codes	76801	OB US < 14 WKS, SINGLE FETUS
	76802	OB US < 14 WKS, ADD'L FETUS
	76805	OB US >/= 14 WKS, SNGL FETUS
	76810	OB US >/= 14 WKS, ADDL FETUS
	76811	OB US, DETAILED, SNGL FETUS
	76812	OB US, DETAILED, ADDL FETUS
	76813	OB US NUCHAL MEAS, 1 GEST
	76814	OB US NUCHAL MEAS, ADD-ON
	76815	OB US, LIMITED, FETUS(S)
	76816	OB US, FOLLOW-UP, PER FETUS
	76817	TRANSVAGINAL US, OBSTETRIC

Scope of the Report

The scope of this report is defined as:

Patient group: Pregnant women
Intervention: US used for screening purposes or for guiding patient management
Comparators: No screening, screening by other methods, or concealment of US findings
Outcomes: Change in patient management; frequency of Cesarean section and abortion; maternal and fetal health outcomes; frequency of preterm birth

The following key questions will be addressed:

1. What is the evidence of efficacy and effectiveness of ultrasonography? Including consideration of:
    a. Test accuracy
    b. Change in patient management
    c. Reductions in perinatal morbidity and mortality
    d. Rate of labor induction for postterm pregnancy
    e. Rate of Caesarian section
    f. Rate of abortion for fetal anomaly
    g. What is the evidence of the safety of ultrasonography? [Including consideration of adverse events type and frequency (mortality, major morbidity, other)]

2. What is the evidence that ultrasonography has differential efficacy or safety issues in subpopulations? Including consideration of:
    a. Gestational age
    b. Other patient characteristics or evidence based patient selection criteria
    c. Type of scanning machine and software, reader training, and other operational factors

d. Provider type, setting, or other provider characteristics
e. Healthcare system type, including worker's compensation, Medicaid, state employees

3. What is the evidence of cost implications and cost-effectiveness of ultrasonography? Including consideration of:
    a. Short-term costs
    b. Long-term costs

DESCRIPTION

Ultrasonography (US) utilizes sound waves at different frequencies for imaging. Real-time sonography is used to confirm the presence of fetal viability by observation of fetal heart rate and fetal movements. Most US transducers utilize frequencies between 3 MHz and 5 MHz for adequate resolution during pregnancy. Transvaginal US (TVU) is performed using a probe placed in the patient's vagina. The US procedure is usually performed with the patient in a semi-recumbent position. Due to closer anatomic proximity of the US probe to the uterus, gestational sac can be discovered as early as 6 weeks with TVU (AIUM, 2007; ACOG, 2009).

The sonographic sound waves transform into heat energy. The US transducer contains piezoelectric crystals, which permit emission and reception of US waves following stimulation by an electrical current. Different properties of various tissues determine the reflection of echoes and thus the image produced by the machine. The amount of energy absorbed depends upon the type of tissue, duration of exposure, and the US route or mode. Based on clinical application, different modes, including B (brightness), M (motion), or Doppler (D), are employed on the US machine (ACOG, 2009; Houston et al., 2009).

Due to inability to measure fetal temperature elevations, a measure of thermal output was developed for video display on the US machine. The video display continuously displays the acoustic, thermal, and mechanical indices to safeguard against exceeding safety standards. The thermal index is an estimate of increase in temperature from acoustic output. Previous research suggests that a thermal index below 1.0 does not have any potential risks. The mechanical index is used to estimate the potential risk of cavitation from heat generated by real-time imaging. The significance of acoustic, thermal, and mechanical indices is in terms of the teratogenic effects of hyperthermia in the growing fetus (ACOG, 2009; Houston et al., 2009).

PRACTICE GUIDELINES

The following sources were searched for practice guidelines: BlueCross/BlueShield TEC Assessments, Canadian Agency for Drugs and Technology in Health (CADTH), Centre for Reviews and Dissemination (CRD), Cochrane Library, Institute for Clinical Systems Improvement (ICSI), and MEDLINE. Additionally, the websites of the American College of Obstetricians and Gynecologists (ACOG), American College of Radiology (ACR), and American Institute of Ultrasound Medicine (AIUM) were searched. Seven guidelines that were published for physicians in the United States and that defined the indications for US screening in pregnancy were selected (AIUM, 2007; ACR, 2008a, ACR, 2008b; ACOG, 2009; ACR 2009a; ACR, 2009b; ICSI, 2010). The Appraisal of Guidelines Research & Evaluation (AGREE) Instrument provided a framework for the quality assessment of selected practice guidelines; items relating to the methodology of developing recommendations were emphasized in the quality assessment. Appendix I

provides a more detailed discussion of the selected guidelines summarized below. Table 1 provides an overview.

American Institute of Ultrasound Medicine (AIUM) (poor quality): A *Practice Guideline for the Performance of Obstetric Ultrasound Examinations* issued in 2007 advises that fetal US performed in the first, second, and third trimester of pregnancy can be safe and beneficial to diagnose, evaluate, or confirm a number of clinical indications related to fetal and maternal health (AIUM, 2007).

American College of Obstetricians and Gynecologists (ACOG) (fair quality): A 2009 *Practice Bulletin for Ultrasonography in Pregnancy* concluded that US examination is a safe and accurate method of determining precise gestational age, number of fetuses, viability, and location of the placenta, as well as diagnosing a number of major fetal anomalies (level A for good and consistent evidence) (ACOG, 2009). There is some support for the use of US to assist in the detection of fetal growth disturbances and abnormalities in amniotic fluid volume (level B for limited or inconsistent evidence). The Bulletin advises that in general, optimal timing of a single US examination is between 18 weeks and 20 weeks of gestation (level C for expert opinion or consensus). The Bulletin also points out that although US accurately estimates gestational age and detects multiple gestations and major fetal anomalies, whether these benefits translate to either fetal or maternal health outcomes remains unproven. The guideline mentions that overall, there is little evidence that US reduces the rate of perinatal morbidity.

American College of Radiologists (ACR) (fair quality): The ACR has issued several guidelines or "Appropriateness Criteria" that address the use of US for clinical conditions related to pregnancy (ACR, 2008a; ACR, 2008b; ACR, 2009a; ACR, 2009b). Details for all five guidelines are presented in Appendix II, but only the recommendations most relevant to the focus of this report are discussed here. The ACR guidelines were not designed to answer questions about whether US surveillance in high-risk pregnancy or universal screening in low-risk or unselected pregnancies is appropriate. Rather, they emphasize accuracy and the association between certain detectable conditions and outcomes but do not directly address the question of the association between US screening or US-guided management and fetal or maternal outcomes. The recommendation ratings are based on a 9-point scale, with 9 representing the highest confidence that the technology is "usually appropriate."

- *Appropriateness Criteria for Growth Disturbances and the Risk of Intrauterine Growth Restriction (IUGR):* This guideline focuses on the use of US to confirm a suspicion of IUGR. It recommends that US of a pregnant uterus is safe and usually appropriate for determining fetal measurement, growth, amniotic fluid, fetal anatomic survey, and activity patterns (rating 9) (ACR, 2008a).

- *Appropriateness Criteria for Multiple Gestations:* This guideline recommends transabdominal or transvaginal US (TVU) as safe and appropriate for patients with a high or low index of suspicion for multiple gestations, or in patients who have already been diagnosed with twins (rating 9). The guideline advises that the evidence does not support the use of transabdominal or TVU with umbilical artery Doppler as a method of assessment for twins (rating 4) (ACR, 2008b).

- *Appropriateness Criteria for First Trimester Bleeding:* This guideline recommends transabdominal or transvaginal pelvic US as generally safe and appropriate in patients with a positive urine or serum pregnancy test, when correlated with other testing, and in patients who were diagnosed with twins on an initial US (rating 9). Pelvic US with Doppler imaging is

recommended as generally safe and appropriate; however, pulsed Doppler of the embryo is not recommended (rating 4) (ACR, 2009a).

- *Appropriateness Criteria for Second and Third Trimester Bleeding*: The guideline generally recommends transabdominal pelvic US for a variety of conditions (rating 9), with TVU as an alternative or as follow-up (rating 8 or 9) (ACR, 2009b).

The Institute of Clinical Systems Improvement (ICSI) (fair): The guideline *Routine Prenatal Care* makes these comments and recommendations (ICSI, 2010):

- The existing evidence does not support routine US examinations in low-risk pregnancies since there is no evidence of improved perinatal outcomes but US is an option that may be considered (the guideline does not specify the conditions under which US in low-risk populations is appropriate).
- If a single screening US is to be performed, ACOG recommends between 18 weeks and 20 weeks as the optimal time (see previous discussion). There is no evidence to support the use of routine US in low-risk pregnancies beyond 24 weeks gestation (the Cochrane Review of US in late pregnancy by Bricker et al., [2008], discussed in this report, is cited as support for this statement).
- US can be used for gestational dating and anatomy evaluations, and for assessing possible genetic abnormalities. US may be useful to confirm a questionable fetal position/presentation.
- If testing for fetal aneuploidy is to be conducted, ICSI recommends nuchal translucency testing during the first trimester, between 10 weeks and 13 weeks, to enhance the identification of Down syndrome.
- Early sonography may confirm dating when gestational age is uncertain or there are antecedent medical complications, including pregestational diabetes mellitus, or previous complications (high-risk pregnancies).
- ICSI considers a plan for serial US and antepartum fetal testing reasonable for the management of hemoglobinopathies during pregnancy.
- ICSI acknowledges the continued improvement in the identification of congenital anomalies using superior equipment by more experienced examiners.
- Three-dimensional (3D) and four-dimensional (4D) US are not recommended for routine use during pregnancy.

Summary

Fair-quality guidelines from ACOG, ACR, and ICSI are consistent with each other and with the literature in describing US as a reasonably safe procedure that accurately provides a wealth of information about pregnancy status and fetal health. Although the guidelines from ACOG allude to the questionable relationship between routine use of US and maternal and fetal outcomes, recommendations were not formed with this in mind. The ICSI guidelines take into consideration the lack of evidence supporting routine use of US in low-risk pregnancy, especially in late pregnancy, but do not fully address the use of US in high-risk pregnancy. None of the guidelines addresses the use of US to monitor cervical length. None of the guidelines considers evidence pertaining to the long-term effects on child growth and development, differential effectiveness and safety, or cost-effectiveness.

METHODS

Search Strategy and Selection Criteria

This health technology assessment assumes that impact on patient management and maternal and fetal outcomes is the most direct way to measure the clinical utility (effectiveness) of US as a screening and prognostic tool. An attempt was made to locate recent reviews and practice guidelines that could provide estimates of accuracy, but a complete systematic literature search related to accuracy was not conducted, and the selected sources were not critically appraised. Rather, the systematic literature search focused on assessments of the impact of US on patient management and maternal and fetal outcomes and on safety. Systematic reviews and any type of controlled study were to be considered eligible if they addressed any of the Key Questions and met these "PICO" criteria:

> **P**atient group: Pregnant women
> **I**ntervention: US used for screening purposes or for guiding patient management
> **C**omparators: No screening, screening by other methods, or concealment of US findings
> **O**utcomes: Change in patient management; frequency of Cesarean section and abortion; maternal and fetal health outcomes, including frequency of preterm birth

Uncontrolled studies, studies that did not report one of the outcomes identified in the foregoing PICO statement, and studies of the use of US during labor were excluded.

Initially, evidence for this report was obtained by searching for relevant systematic reviews in the following databases: BlueCross/BlueShield TEC Assessments, Canadian Agency for Drugs and Technology in Health (CADTH), Centre for Reviews and Dissemination (York University), and Cochrane Library. Additional systematic reviews were selected from a search of the MEDLINE database for the dates 2005 through July 2010, using various limits to identify systematic reviews. After the most recent systematic reviews were identified, a search of the peer-reviewed medical literature using the MEDLINE and EMBASE databases was conducted in order to identify studies published after the selected systematic reviews. This search covered 2008 to July 2010 for studies of clinical utility (impact on outcomes) and October 2007 to July 2010 for studies of safety. Detection of fetal anomaly was considered an eligible fetal health outcome, but selection was limited to studies looking at general detection rates; studies focusing on genetic anomalies were excluded. The same study design selection criteria observed by the systematic reviews were used in this search: randomized controlled trials (RCTs) or quasi-randomized trials for clinical utility studies and any interventional or observational study with a control group for safety studies. Search terms included *ultrasonography* and variations, combined with *pregnancy* as keywords, subject words, and title words. The search was limited to the English language and to human subjects. See Appendix I for additional details. To make sure any studies of 3D and 4D US other than accuracy assessments were not missed, Hayes Medical Technology Directory reports on 3D/4D US (Hayes, 2005a; Hayes, 2005b; Hayes, 2005c; Hayes, 2005d; Hayes, 2006a; Hayes 2006b), and their annual update searches, were also reviewed. Various targeted searches for the years 2000 to July 2010 were conducted in an attempt to identify studies relevant to key questions but excluded by the systematic reviews. See Appendix I for additional details.

The MEDLINE and EMBASE databases were initially searched for economic evaluations and cost descriptions published between 2000 and July 2010. Studies were selected if they examined the cost, cost consequences, or cost-effectiveness of US. Because of the large number of available studies, studies

that addressed routine screening for single abnormalities were excluded. Several excluded studies evaluated US as a method of screening for Down syndrome (Caughey et al., 2002; Cusick et al., 2003; Shohat et al., 2003; Harris, 2004; Odibo et al., 2005; Gekas et al., 2009). A cost-effectiveness study of targeted versus universal screening for vasa praevia was also excluded (Cipriano et al., 2010). A general update search, using the main search terms, was conducted for the time frame August 2010 through September 10, 2010.

Quality Assessment

The evidence for this report comes primarily from systematic reviews and meta-analyses. No formal quality assessment tools were used to independently evaluate individual studies included in the reviews. With some exceptions, the report relies primarily on information about the studies and their quality as supplied by review authors. Taking this information into account, and following the GRADE system, a quality rating of *very low, low, moderate,* or *high* will be applied to bodies of evidence that answer components of Key Questions 1, 2, and 3.

Economic evaluations and cost descriptions (Key Question 4) were judged according to whether outcomes and costs were collected at the same time, the quality of sources for effectiveness estimates, the appropriateness of the time frame, and whether sensitivity analyses were reported. Generalizability will be described in terms of the populations included, treatment protocols followed or assumed, whether the study was conducted in the United States, and the age of the study.

LITERATURE REVIEW

Efficacy and Effectiveness: Accuracy

The selected literature suggests that US has variable accuracy, depending on the target condition. As a screening tool, it is often combined with other tests. Review articles report sensitivities of 34% to 99%, but information about specificity, positive predictive value, and negative predictive value was not readily available.

Gestational Age: A term pregnancy is defined as live-born infant who has completed 37 weeks of gestation. Different gestational age estimation methods are currently used and none of them is completely valid and reliable. During early first trimester, when no structures are visible within the gestational sac, the gestational age may be estimated from the sac diameter, and later the crown rump length is measured to estimate the gestational age. The crown rump length is the longest demonstrable length of the embryo or fetus, excluding the limbs and the yolk sac. An analysis of the best method of gestational age estimation for research purposes cited studies showing that differences in accuracy between US dating and dating based on last menstrual period are not clinically meaningful (Lynch and Zhang, 2007).

Fetal Abnormalities: US is used during the first and second trimester for assessment of fetal anatomical abnormalities, some of which are caused by chromosomal abnormalities. Aneuploidy, or chromosomal abnormality, is often associated with both major anatomical malformations and with minor markers (or soft signs) that show up on US. The soft signs are useful only for screening purposes and are not sufficient for a prognosis of abnormality in the absence of aneuploidy or a major structural malformation. During the first trimester, measurement of the soft marker fetal nuchal translucency (NT) (a measure of the thickness of the area below the skin in the back of the neck) and maternal serum

markers (β-HCG and PAPP-A) is a highly sensitive screening test for Down syndrome. This combined first-trimester testing has been found to have detection rates between 82% and 87% with a false-positive rate of 5%. The optimal time for performance of NT is 11 to 13 weeks of gestation. NT is also associated with chromosomal aneuploidy other than the abnormality associated with Down syndrome and with structural defects and sometimes appears in fetuses that have normal outcomes. There is an association between increased NT and cardiac defects in euploid (normal number of chromosomes) fetuses. Overall, US has a sensitivity of approximately 40% (range 13-82%) for detecting fetal anomalies. This estimate is based on a review, cited in guidelines published by the American College of Obstetrics and Gynecology (ACOG), of 36 studies (n=900,000 fetuses). Accuracy varied by how anomaly was defined, characteristics of the population studied, expertise of operators, and how anomalies were ascertained (ACOG, 2009). Another review reported that US screening during the first and second trimesters has 81% sensitivity for open neural tube defects, 96% to 100% for anencephaly, 5% to 60% for congenital heart disease, and 60% for genitourinary abnormalities (ACOG, 2007; Flood and Malone, 2008; Shaw et al., 2008; Gagnon et al., 2009; Pathak and Lees, 2009;).

Macrosomia: In a systematic review, Coomarasamy et al. (2005) reported no difference in accuracy between US-determined fetal weight (EFW) and abdominal circumference (AC) for the prediction of a macrosomic baby at birth.

Multiple Gestation: US is used to screen and manage multiple pregnancy. The determination of zygosity[1] and chorionicity[2] can guide antenatal management and screening of chromosomal abnormalities in multiple pregnancies. The sensitivity and specificity of US in detection of chorionicity are 89.8% and 99.5%, respectively, during the first trimester. Sensitivity remains the same but specificity decreases to 94.7% in the second trimester. In the systematic review selected as evidence of the effectiveness of routine US in early pregnancy (Whitworth et al., 2010), US was found to significantly reduce the failure to detect multiple pregnancy by 24 to 26 weeks by 93% in pooled analysis (1% failure versus 39% failure) and to significantly reduce failure to detect multiple pregnancy before birth by 88% (no failures versus 9% failure). Furthermore, the detection of fetal anomaly for multiple gestations in early pregnancy was more than three times more likely with the use of US (Martin et al., 2009; Whitworth et al., 2010). (See ***Efficacy and Effectiveness: Clinical Utility of US in Low-Risk Pregnancies***.)

Measurement of Cervical Length and Prediction of Preterm Birth: A systematic review with meta-analysis concluded that TVU determination of cervical length at 20 weeks to 24 weeks is a good predictor of spontaneous preterm birth in asymptomatic women with *twin* pregnancies (Conde-Agudelo et al., 2010). Specificity, which would be the key measure for ruling out risk of preterm birth, was 91% to 97% for preterm birth at different time points between 28 and 37 weeks and at cutoffs of 20 mm or 25 mm as the definition of short cervix. Specificity was low (76% to 81%) in women with twin pregnancy but increased risk because of a history of preterm birth. Specificity was 74% at best in women with twin pregnancy and symptoms of preterm labor. Another systematic review and meta-analysis demonstrated high rule-out accuracy for TVU determination of cervical length in women with *singleton* pregnancy and intact membranes but symptoms of preterm labor (Sotiriadis et al., 2010). At a cutoff value of 15 cm, specificity was 90.5% and the negative likelihood ratio was 0.51 for birth within 7 days of presentation; specificity was 93.7% and the negative likelihood ratio was 0.63 for preterm birth before 34 weeks. (The authors reported that a negative likelihood ratio < 0.2 is considered good for rule-out situations.) Meta-

[1] Identical or fraternal twins.
[2] Whether or not twin fetuses share a placenta.

analysis results from studies that used higher cutoff values to define short cervix suggested that higher cutoffs improve rule-out accuracy only slightly.

Ectopic Pregnancy: Ectopic pregnancy is typically ruled out if an intrauterine pregnancy can be confirmed (although this can, of course, lead to a false negative in the case of a heterotopic twin pregnancy, which is a rare occurrence). The specificity of emergency department TVU for intrauterine pregnancy exceeds 98% and sensitivity exceeds 90% in most studies (McRae et al., 2009).

Preeclampsia: A systematic review (Cnossen et al., 2008) has shown Doppler US to be more accurate in diagnosing preeclampsia during the second trimester than in the first trimester, with increased pulsatility index with notching serving as the best predictor. The positive likelihood ratio (increased probability of detection) was 21.0 among high-risk patients and 7.5 among low-risk patients. Second-semester sensitivity and specificity for pulsatility index with notching was 23% and 99% in low-risk women and 19% and 99% in the high-risk women. The authors did not document the corresponding first trimester sensitivity and specificity values.

DUS can also determine abnormal umbilical artery (UA) flow, which is associated with a higher risk for IUGR. The detection rate of uterine artery screening for preeclampsia or IUGR at any stage of gestation is better for severe than for mild disease. Increased resistance indices in the first trimester are particularly effective in identifying preterm, rather than term, preeclampsia (Cnossen et al., 2008; Zhong et al., 2010). One systematic review evaluated 37 studies that assessed different combinations of biochemical and DUS markers for preeclampsia. The authors found that in low-risk pregnant women, a combination of serum tests measured in first or early second trimester, combined with DUS in the second trimester, has a sensitivity between 60% and 80% and specificity greater than 80%. In high-risk populations during first trimester, the combination of PP13 with pulsatility index determined by DUS showed both sensitivity and specificity of 90% in a single study limited to severe preeclampsia (Giguère et al., 2010).

Intrauterine Growth Restriction (IUGR): The reported sensitivity of DUS for diagnosis of IUGR (birth weight < 10th percentile) has ranged from 34% to 97%. One study assessed abdominal circumference (AC) as a predictor of IUGR and reported that AC<10th percentile has a sensitivity of 62.25%, specificity of 90.7%, positive predictive value of 67.35%, and a negative predictive value of 89.8%. Previous research shows that weight estimation has higher specificity but a much lower sensitivity compared with DUS (Ott, 2002; Platz and Newman, 2008; Falo, 2009). ACOG guidelines describe use of US for detection of IUGR to be a screening tool; further follow-up evaluation is required (ACOG, 2009).

Efficacy and Effectiveness: Clinical Utility of US in High-Risk Pregnancies

Search Results

Three systematic reviews with meta-analysis were selected for their assessment of US in high-risk pregnancy: (1) a Cochrane Review assessing the effect of fetal and umbilical Doppler US (DUS) on maternal care and fetal outcomes (Alfirevic et al., 2010); (2) a Cochrane Review assessing the use of transvaginal ultrasound (TVU) and measurement of cervical length to prevent preterm birth in certain situations (Berghella et al., 2009); and (3) a systematic review of surveillance of cervical with TVU rather than patient history as the basis for cerclage to prevent preterm birth due to short cervical length (Blikman et al., 2008).

The review by Alfirevic et al. (2010) (see Appendix IIIA) included RCTs or quasi-randomized controlled trials of fetal DUS, compared with no DUS or with concealment of DUS findings, in women considered to be at high risk for fetal compromise. Some studies specified particular risk factors such as suspected IUGR or hypertension as inclusion criteria. Advanced age was not mentioned as a risk factor in the study information provided by the review authors. When reported, study inclusion criteria specified gestational age > 24 weeks (late pregnancy); six studies did not report inclusion criteria. The review excluded TTTS as a risk category because TTTS is less common than growth discordance, or selective intrauterine growth restriction, in multiple pregnancy and because TTTS requires unique monitoring and treatment strategies. The review also excluded utero-placental Doppler US because of an upcoming Cochrane review on this topic. Controls and whether DUS was used in combination with other fetal assessments varied across studies. The authors selected 18 trials (10,156 patients), four of which were published only as conference proceedings or reported in private communication. Only one trial was conducted in the United States; the remaining studies were conducted in Europe, Africa, or Australia.

The review by Berghella et al. (2009) (see Appendix IIIB) investigated the effectiveness of assessing cervical length with TVU in order to prevent preterm birth in women at risk because of factors other than obstetric history. Five RCTs (n=506), two of which were not published in full, were included. Three studies enrolled women with singleton pregnancy and suspected preterm labor; one enrolled patients with singleton pregnancy and premature partial rupture of membrane PPROM) and looked only at the safety of US; and one study enrolled asymptomatic women with twin pregnancies. The study involving PPROM will be discussed under **Safety**.

The systematic review by Blikman et al. (2008) (see Appendix IIIC) also looked at TVU measurement of cervical length, but the comparison was surveillance in patients at risk because of previous preterm birth, followed by cerclage when indicated versus prophylactic, or elective, cerclage based on patient history. The surveillance period was roughly second trimester in most studies. Six published studies (n=653) were selected, two of which were RCTs.

A search of the MEDLINE and EMBASE databases did not reveal any studies of the utility of US in high-risk pregnancy that were published after the systematic reviews, except for a single trial (Simcox et al., 2009) on the same topic addressed by Blikman et al. (2008). Surveillance occurred every 2 weeks up until 24 weeks (just short of the end of second trimester).

Findings

Change in Patient Management, DUS: In the review by Alfirevic et al. (2010), DUS was associated with an absolute reduction of 10.6 percentage points in the frequency of antenatal admissions (relative risk [RR] 0.72, 95% CI 0.60-0.88; 893 patients in two singleton studies, one of which was not published in full). Analysis of the fully published study (n=426) showed that compared with CTG, DUS was associated with an absolute reduction of 13.6 percentage points in antenatal admissions.

Change in Patient Management, TVU: Nonsignificant pooled estimates in the meta-analysis conducted by Berghella et al. (2009) suggested that use of TVU to measure cervical length could result in an increase in maternal hospitalization, a decrease in use of tocolysis, and an increase in administration of steroids for fetal lung maturity in women with singleton pregnancy with suspected preterm labor. Berghella et al. did not comment on how to interpret these findings or provide information on the

appropriateness of the patients' hospitalization or treatment. One of the included studies (Alfirevic et al., 2007) reported a significant and substantial decrease in the frequency of *inappropriate* treatment (RR 0.16, 95% CI 0.05-0.39; n=41). Inappropriate treatment was defined as tocolytics plus steroids given and delivery ≥ 7 days later, or *no* treatment and delivery < 7 days later. Another included study (Ness et al., 2007) reported that failure to give steroids prior to preterm delivery was *not* increased in the TVU knowledge arm, but actual data were not reported.

The review by Blikman et al. (2008) reported that, whereas all patients in the history-predicated control groups underwent cerclage (by design), only 40% to 64% of patients in the TVU groups underwent cerclage (study groups ranged from 73 to 177 patients in six observational studies and RCTs). In an RCT (n=253) published later, TVU surveillance increased the use of cerclage from 20% to 32% (RR 1.6, 95% CI 1.03 to 2.47) (Simcox et al., 2009). The number of hospital admissions was similar in the TVU surveillance and history-guided groups of this trial, but mean hospital stay was significantly longer by about two days in the TVU group. Simcox et al. also found that use of progesterone, tocolysis, and antenatal steroids increased with the use of TVU surveillance, but the difference was significant only for progesterone (RR 1.55, 95% CI 1.06 to 2.25; absolute increase from 25% to 39%).

Reductions in Perinatal Morbidity and Mortality, DUS: Pooled estimates by Alfirevic et al. (2010) showed significant relative reductions in the risk of three composite outcomes: serious neonatal morbidity (RR 0.13, 95% CI 0.02 to 0.99; 500 pregnancies in one study), any perinatal death after randomization (RR 0.71, 95% CI 0.52 to 0.98), and any potentially preventable perinatal death (RR 0.67, 95% CI 0.46 to 0.98) (10,125 pregnancies in 16 studies for the two death estimates). These findings represent very small reductions in absolute risk (0.5, 2.8, and 0.4 percentage points). Number-needed-to-treat (NNT) calculations suggest that 203 high-risk pregnant women would have to be screened with DUS to prevent one perinatal death, 246 to prevent a preventable perinatal death, and 36 to prevent one incident of serious neonatal morbidity. There was a significant but modest reduction in length of infant hospital stay overall (1076 patients in three singleton studies) and in a subset analysis of DUS versus CTG alone (576 patients in two singleton studies). The review authors considered the overall quality of evidence for impact on the two primary outcomes (any perinatal death and serious neonatal morbidity) to be very low because of missing study information and too few events for precision[3]. They point out that the confidence intervals for the effect estimates include values very close to 1, which would mean no effect. There was some evidence of publication bias with respect to perinatal death and significant heterogeneity with respect to neonatal morbidity. A sensitivity analysis restricted to the three high-quality studies resulted in a RR that was somewhat smaller than the overall RR for any perinatal death (0.61 for the high-quality subset, compared with 0.71 for all studies), but the RR for the high-quality studies was nonsignificant. Results across the high-quality studies did not present a consistent pattern of DUS-control differences.

Alfirevic et al. (2010) reported RRs for numerous specific fetal outcomes that favored DUS but were nonsignificant. Examples include stillbirth, Apgar < 7 at 5 minutes, gestational age at birth, and abnormal neurological development at 9 months. The RRs for some specific outcomes favored the control groups, but there were no outcomes for which pooled estimates of effect were significant and favored the control groups.

[3] Lack of precision in this sense means the confidence interval around the pooled RR estimate was wide, suggesting the true RR may be very different from the estimate.

Alfirevic et al. (2010) also separately analyzed the subset of studies that compared DUS alone with CTG alone. In general, the pooled estimate of effect across these studies was less likely to be significant than in the entire set of studies. The pooled RR for the review's first primary outcome, any perinatal death, favored DUS (RR 0.45) compared with CTG but was nonsignificant (2813 patients in four studies, including two good-quality RCTs). The review's other primary outcome, serious neonatal morbidity, could not be assessed for this subset of studies.

Reductions in Perinatal Morbidity and Mortality, Transvaginal US: The review by Berghella et al. (2009) evaluated the impact of TVU assessment of cervical length on the incidence of preterm birth, which might be seen as an intermediate outcome measure for neonatal morbidity and mortality. TVU results were concealed from clinicians in the control groups. Three RCTs (n=290) included women with singleton pregnancies who already had symptoms and/or signs of preterm labor. The pooled estimate calculated by Berghella et al. for the effect on preterm birth at < 37 weeks (256 patients in two RCTs) did favor knowledge of the TVU findings (RR 0.59), but it was statistically nonsignificant with a wide confidence interval that included the possibility of a substantial increase in preterm birth. The pooled estimate for preterm birth at < 34 weeks was similar. One of the two RCTs contributing to this RR had few limitations (Ness et al., 2007); this trial (n=100) showed preterm birth to be substantially less frequent in the TVU arm (13.0%) compared with the arm that underwent fetal fibronectin testing (36.3%, *P*=0.01). Specific results for the other RCT (Palacio et al., 2006) (n=149), which was published as abstract only and was of unclear quality, were not available. The only significant effect in this population was an increase of a little over 4 days in gestational age at birth (mean difference 0.64 weeks, 95% CI 0.3 to 1.25; 290 patients in three RCTs). There was a positive but nonsignificant effect on birth weight, and there were no instances of perinatal death.

Berghella et al. (2009) included an unpublished RCT of 125 patients (Gordon et al., 2006) with no symptoms of preterm labor but at risk because of twin pregnancy. TVU measurement of cervical length was performed at regular intervals. Pooled estimates (1 study, n=125) for preterm birth at < 36 weeks favored the control group, while estimates for < 34, 32, and 30 weeks favored knowledge of TVU findings. The estimates were nonsignificant with wide confidence intervals. Estimates for gestational age at delivery and birth weight were slightly positive but nonsignificant.

The two RCTs (n=170) (Althuisius et al., 2000; Beigi et al., 2005) included in the review by Blikman et al. (2008) enrolled only women with classic risk factors for shortened cervical length, i.e., previous preterm birth or second-trimester pregnancy loss. The RRs of preterm birth calculated for these two studies were 1.05 and 1.25, which indicates a greater frequency of preterm birth in the TVU groups; however, both estimates were nonsignificant. The observational studies also showed nonsignificant differences favoring history-predicated cerclage with one exception, a prospective cohort study (Higgins et al., 2004; n=135) that favored TVU (RR=0.14, nonsignificant). In studies where data were available for pregnancy loss at < 24 weeks, outcomes again favored the history-predicated group in one RCT (RR=1.08, nonsignificant) (Beigi et al., 2005) but were mixed in the observational studies. Estimates from all studies generally had very wide confidence intervals.

A more recent RCT (Simcox et al., 2009) enrolled asymptomatic women (n=253) with singleton pregnancy and considered to be at high risk of preterm birth because of ≥ one previous delivery between 16 and 34 weeks. The study made the same type of comparison considered by Blikman et al. (2009). Women in the TVU arm underwent scanning every two weeks up to 24 weeks and cerclage was

performed if the cervix shortened to ≤ 20 cm. For women in the history-indicated arm, clinicians were free to apply whatever criteria they wished, but a decision on whether to use cerclage was made before randomization for all women. The primary outcome was preterm birth < 34 weeks. The study was powered to detect an absolute risk reduction of 20 percentage points, assuming an incidence of 40% in the history-indicated arm. Preterm birth occurred in 15% of women in each arm. Pregnancy loss before 24 weeks and PPROM were less frequent in the TVU arm but differences were nonsignificant. This was a good-quality study with a clear description of randomization and allocation concealment methods, minor loss to follow-up, and intention-to-treat (ITT) analysis. However, the increased use of progesterone in the TVU group may have created a bias in favor of TVU with respect to preterm birth rates.

Cesarean Section and Induction of Labor, DUS: Pooled estimates reported by Alfirevic et al. (2010) indicated a reduced incidence of Cesarean section (RR 0.90, 95% CI 0.84 to 0.97; 7918 pregnancies, 14 studies) and of emergency Cesarean section in particular (RR 0.81, 95% CI 0.67 to 0.98; 6175 pregnancies, 10 studies) to be reduced with the use of DUS. There was some evidence suggestive of publication bias for the estimate of effect on emergency Cesarean section. The pooled estimate for elective Cesarean section slightly favored no DUS but was nonsignificant. The incidence of induction of labor was also significantly reduced (RR 0.89, 95% CI 0.80 to 0.99), according to data from 10 studies (5633 pregnancies). Absolute risk reduction was small (two to three percentage points). NNT calculations indicated 39 women with high-risk pregnancy would have to be screened to prevent one Cesarean section, 47 to prevent one emergency Cesarean section, and 31 to prevent one induction of labor. Alfirevic et al. considered the evidence for these outcomes to be of low quality because of missing information, heterogeneity, and/or possible publication bias. Estimates were also characterized by imprecision.

In comparisons of DUS versus CTG alone (1473 pregnancies in three studies, including two high-quality RCTs), DUS was associated with an *increase* in the incidence of *elective* Cesarean section (RR 1.53, 95% CI 1.12 to 2.09) and a *decrease* in the incidence of *emergency* Cesarean sections (RR 0.66, 95% CI 0.52 to 0.84) (Alfirevic et al., 2010). The absolute risk reduction for emergency Cesarean section was 45 percentage points and the NNT was 12. The authors of the systematic review reported that although formal meta-analysis was not possible because of too few studies, the lack of heterogeneity between the two sets of studies (elective and emergency) suggested that the difference in the effect of US is real. Among the DUS-versus-CTG subset of studies, there was no significant impact on rate of induction of labor.

Rate of Abortion for Fetal Anomaly: The systematic review of fetal DUS to screen for fetal anomaly did not report data on the impact of US on rate of abortions in high-risk pregnancies.

Summary of Efficacy/Effectiveness Evidence in High-Risk Pregnancies

Although most studies were RCTs, only a few were clearly of high quality. Nevertheless, sensitivity analysis by Alfirevic et al. (2010) suggested that overall estimates for DUS were not affected by study quality. Blinding was absent from all studies, which poses a potential bias in studies' results for rates of Cesarean section and induction of labor. In other words, clinicians' decisions regarding these interventions could be affected by their knowledge of whether or not patients had been allocated to receive US. Lack of blinding is not likely to bias the other outcomes, which reflect natural events that are objectively assessed. Quantity of evidence, precision of estimates or lack thereof, and the review

authors' findings with respect to publication bias were also considered in making a quality assessment for each body of evidence.

Change in Patient Management: Moderate-quality evidence showed a moderate reduction in antenatal admissions with the use of DUS. Low-quality evidence failed to demonstrate a significant effect on maternal hospitalization , use of tocolysis, or use of steroids for fetal lung immaturity when TVU was used to measure cervical length in women with suspected preterm labor (two small trials). However, low-quality evidence (single small trial) suggested that assessment of cervical length with TVU in women with singleton pregnancy and signs or symptoms of preterm labor substantially reduces *unnecessary* use of tocolysis and steroids. Low-quality evidence (n=77 to n=177; two RCTs, one prospective cohort study, and three retrospective studies), also suggests that use of cerclage may be reduced when patients at risk of cervical insufficiency are monitored with TVU, but these results were contradicted by the results of the most recent RCT (n=253).

Reduction in Perinatal Mortality and Morbidity: Very-low-quality evidence suggests that DUS screening in patients at high risk of fetal compromise modestly reduces overall perinatal mortality (10,125 pregnancies in 16 RCTs or quasi-randomized studies) and serious neonatal morbidity (598 pregnancies in three RCTs). Pooled results suggest that more than 200 pregnant women would have to be screened to prevent one perinatal death and 36 women would have to be screened to prevent one incidence of neonatal morbidity. Although the body of evidence is large, it is hampered by imprecise estimates of relative risk that do not rule out the possibility of negligible effect, by missing information from study reports, and by heterogeneity or publication bias for some measures. It has been suggested that the reduction in the rate of perinatal mortality associated with prenatal US is attributable to pregnancy termination (ACOG, 2009); the available evidence does not allow an assessment of whether reduced perinatal mortality is due to better patient management or pregnancy termination. DUS alone appears to have a weaker effect than DUS in combination with other screening, but only a few studies used DUS alone and these measured a limited range of outcomes.

There is promising but low-quality, inconclusive evidence suggesting that management based on TVU assessment of cervical length reduces the incidence of preterm birth in women already thought to be in preterm labor (256 pregnancies, three RCTs), and low-quality evidence suggesting no impact of TVU surveillance in patients with twin pregnancy. Preterm birth is only an intermediate measure but is a leading cause of ultimate perinatal health outcomes (Blikman et al., 2008).

Moderate-quality evidence suggests that surveillance of cervical length with TVU has no impact on preterm birth rates or pregnancy loss at < 24 weeks in women at risk because of previous preterm birth. These findings were reported by three RCTs (n=423) for preterm birth and two RCTs (n= 350) for pregnancy loss; the studies had no limitations other than lack of blinding. However, the RCT results were contradicted by a prospective cohort study (n=135).

Cesarean Section and Induction of Labor: Low-quality evidence suggests that in general, DUS screening in patients at high risk of fetal compromise modestly reduces the incidence of Cesarean section (7918 pregnancies in 14 RCTs or quasi-randomized trials) and induction of labor (5633 pregnancies in 10 studies). The number of women who would have to be screened to prevent one event, based on these data, is 39 (any Cesarean section), 47 (emergency Cesarean section), or 31 (induction of labor). High-quality evidence from three RCTs (total, n=1473) suggested that when specifically compared with CTG alone, DUS *increases* the rate of *elective* Cesarean sections by more than 50% but reduces the rate of

emergency Cesarean sections by 44% in relative terms and 45 percentage points in absolute terms (NNT=12).

Rate of Abortion for Fetal Anomaly: No evidence was available.

Generalizability of the Evidence: Alfirevic et al. (2010) point out that there are no uniform criteria governing intervention when DUS is abnormal; protocols varied among the selected studies. Berghella et al. (2009) also acknowledge that there is uncertainty regarding the most efficacious interventions following determination of cervical length with TVU. Thus, the pooled estimates of differences in outcome between groups managed with and without knowledge of US may be affected by the variation in treatment protocols across studies and may not be generalizable to other settings.

Efficacy and Effectiveness: Clinical Utility of US in Low-Risk Pregnancies

Search Results

Two meta-analyses were selected for their assessment of US in low risk pregnancy: a Cochrane Review assessing the effect of US during *early* pregnancy (< 24 weeks) (Whitworth et al., 2010) and a Cochrane Review assessing the effect of US during *late* pregnancy (> 24 weeks) (Bricker et al., 2008). Authors defined unselected or low risk pregnant patients as those who did not have any clinical indication for US, including a medical condition or previous pregnancy complication.

The review by Whitworth et al. (2010) (see Appendix IVA) included RCTs plus quasi-randomized and cluster-randomized controlled trials (11 trials; 37,505 patients) that compared routine US with selective US for specific indications, in unselected or low risk pregnant women during early pregnancy (< 24 weeks). Studies of patients who underwent more than one US scan were excluded. Most studies included only women with singleton pregnancy although two studies were reported as including some patients with multiple births or other high-risk pregnancy (Bennett et al., 1982; Eik-Nes et al., 1984**)**. The authors included one study that reported long-term childhood outcome data. Most of the included studies took place in Europe; two were conducted in the United States.

The review by Bricker et al. (2008) (see Appendix IVB) was designed to compare routine US with selective US for specific indications; the population studied comprised unselected or low risk women during *late* pregnancy (> 24 weeks). Eight trials (n=27,024) included pregnant patients with singleton pregnancy, except three studies that were reported to include both low- and high-risk pregnant patients (Eik-Nes et al., 1984; Proud et al., 1987; Salvesen et al., 1992). The authors included one study that reported long-term childhood outcome data. One study was conducted in the United States; the others were conducted in Europe. Bricker et al. (2008) explained differences in study protocols:

> "Ultrasound examination options differed between trials, with some offering no routine scans at any time in pregnancy to the control group, some offering routine scans to all participants earlier in pregnancy (before 24 weeks' gestation), and some offering routine scan at all stages of the trial, but only revealing results of late pregnancy ultrasound (after 24 weeks' gestation) for the study groups."

A search of the MEDLINE and EMBASE databases did not reveal any studies of the utility of US in low-risk pregnancy that were published after the systematic reviews.

Routine US in Early Pregnancy

Change in Pregnancy Management: Whitworth et al. (2010) reported a pooled estimate from two RCTs (n=9502) suggesting a reduction in the number of antenatal visits, but the estimate was nonsignificant. A pooled estimate from five RCTs (n=17,685) showed no impact on antenatal hospital admissions. The review also included an RCT (n=602) (Crowther et al., 1999) that examined whether early US scanning reduces the number of inappropriately timed serum screening tests or inappropriately timed repeat fetal anomaly US scanning due to incorrect gestational age at the time of performance of a previous US. Inappropriate testing was less frequent in the US group, but the difference was not significant.

Reductions in Perinatal Morbidity and Mortality: Pooled estimates by Whitworth et al. (2010) showed no statistically significant effect on perinatal mortality (with or without the inclusion of lethal malformations) or on perinatal morbidity-related outcomes. Two of 10 RCTs (n=35,735) that assessed perinatal mortality were quasi-randomized trials and results of sensitivity analysis after excluding the two trials showed no difference between groups in terms of perinatal mortality. The overall rate of perinatal mortality is low in high-resource settings and most RCTs included in this SR were performed in such settings. When children with lethal malformations were excluded, rates of perinatal mortality in the routine US and control groups were similar (0.53 % versus 0.56%). Pooled estimates for a variety of morbidity-related outcomes were all nonsignificant (3906 to 19,337 patients in four to eight studies). Examples of these outcomes included low and very low birth weight, small for gestational age, 5-minute Apgar score ≤ 7, or admission to neonatal intensive care unit.

Cesarean Section and Induction of Labor: Pooled estimates by Whitworth et al. (2010) showed US to have no effect on the rate of Cesarean section during early pregnancy (22,193 patients in five studies). Their analysis did show a statistically significant reduction in rate of induction of labor *for postterm pregnancy* (RR 0.59, CI 0.42 to 0.83; 25,516 pregnancies in eight studies); some heterogeneity was detected. There was also a significant difference in terms of induction of labor *for any reason* (RR 0.78, 95% CI 0.63 to 0.97; 24,790 studies; seven studies). The absolute difference between US and control groups for both measures of induction of labor was 1 percentage point. NNT calculations indicated 100 women with low-risk pregnancy would have to be screened to prevent one labor induction. All multiple pregnancies were detected before labor in the routine US group, whereas 12 of the 133 multiple pregnancies in the control group remained undetected at the onset of labor (RR 0.12, 95% CI 0.03 to 0.54). Whitworth et al. concluded that the reduced incidence of induction of labor in routine US group may be the result of improved estimation of gestational age and earlier detection of multiple pregnancy. Citing the lack of blinding among routine US patients, authors point out that it is possible that the decision to induce labor for postterm pregnancy may be influenced by knowledge of treatment allocation.

Rate of Abortion for Fetal Anomaly: Pooled estimates by Whitworth et al. (2010) showed a statistically significant increase in the rate of abortion for fetal anomaly (RR 2.23, 95% CI 1.10 to 4.54; 28,256 pregnancies; 5 studies), but the rates in each group were very small (0. 17% and 0.07%, respectively). A related finding, based on a smaller set of evidence, was that detection of fetal abnormality in early pregnancy increased from 4% to 16% and detection of major fetal anomaly before birth increased from 9% to 32% (387 patients in 2 studies).

Summary of Efficacy/Effectiveness Evidence for Early Pregnancy

The evidence pertaining to US in early pregnancy is derived from large RCTs that were generally without major limitations except lack of blinding. Sensitivity analyses, in which two quasi-randomized trials were excluded, did not change effect estimates. Heterogeneity across studies was reported for some outcomes, but the authors used random effects models in these situations. Lack of blinding was considered a weakness in the evidence for impact on Cesarean section and induction of labor. Imprecision and whether more than one study was available were also taken into account in the following quality determinations for bodies of evidence.

Change in Patient Management: High-quality evidence showed routine US before 24 weeks to have no effect on hospital utilization. Low-quality evidence (single study with imprecise estimates) showed a reduction in inappropriately timed serum scan and repeat US fetal anomaly scans.

Reductions in Perinatal Mortality and Morbidity: High-quality evidence showed US before 24 weeks to have no effect on perinatal or neonatal mortality or morbidity.

Cesarean Section and Induction of Labor: Moderate-quality evidence (lack of blindness) showed US before 24 weeks to have no effect on the frequency of Cesarean sections but to be associated with a modest reduction in the incidence of induction of labor (NNT=100).

Rate of Abortion for Fetal Anomaly: High-quality evidence showed US < 24 weeks to more than double the rate of abortion for fetal anomaly; the absolute increase was 0.10 percentage points (0.07% to 0.17%).

Generalizability of the Evidence: Whitworth et al. (2010) point out that most studies were conducted in high-resource settings, where overall perinatal mortality is relatively low and fetal abnormality as opposed to quality of care makes a larger contribution to mortality. It is possible that early US could reduce perinatal mortality and/or morbidity in settings with fewer resources. The authors pointed out that there was considerable variation in the timing of US and protocols but did not attempt to analyze which strategies might be more likely to be effective. Furthermore, the reviewed evidence pertains only to pregnancies in which a single US screening scan was performed.

Routine US Late Pregnancy

Change in Patient Management: Pooled estimates by Bricker et al.(2008) showed a small and statistically nonsignificant increase in the number of antenatal admissions (5396 pregnancies; 4 studies) and a small nonsignificant decrease in the rate of subsequent US scans (2536 pregnancies; 2 studies), comparing routine US with controls (moderate-quality evidence). Confidence intervals were wide and did not rule out the possibility of an important effect in either direction. Low-quality evidence showed no difference in the use of cardiograph (2000 pregnancies, 1 study).

Reductions in Perinatal Morbidity and Mortality: Pooled estimates by Bricker et al. (2008) showed no statistically significant effect on perinatal mortality (24,276 patients, seven studies), stillbirth (21,708 patients, five studies), or neonatal death (21,708 patients, five studies). Confidence intervals were wide, e.g., 0.94 (95% CI 0.55 to 1.61) for perinatal mortality, ruling out neither benefit nor harm. Post-term delivery was significantly reduced (RR 0.69, 95% CI 0.59 to 0.81; 17,151 patients in two studies), but the absolute difference was only one percentage point (NNT=100). There was no effect on moderate or

severe neonatal morbidity, according to a single study (n=15,281) (Ewigman et al., 1993). Differences in other outcomes related to perinatal morbidity, such as pre-term delivery, gestational age at delivery, low birth weight, neonatal resuscitation and ventilation, Apgar score, and admission to special care baby unit were either very slight or were nonsignificant (4510 to 20,298 patients in one to four studies). Many of the morbidity-related estimates were imprecise. Bricker et al. (2008) included one RCT (Ewigman et al., 1993) that showed a third trimester detection rate of 22% of fetal anomalies in the screened group and only 6.5% in the control group; none of the other studies addressed this measure. Bricker et al. (2008) commented that an improved detection rate did not appear to translate into improvement in infant survival.

A subgroup analysis suggested an exception to the finding of no effect on stillbirths. A pooled estimate showed a statistically significant relative risk reduction when stillbirths related to congenital abnormalities were excluded (RR 0.05; 95% CI 0.00 to 0.90; 2902 pregnancies; 2 studies). NNT calculations indicated 100 women with low-risk pregnancy would have to be screened to prevent one still birth. Results from one of the two studies suggested that this finding is due to the integration of placental grading into routine third trimester US (Proud et al., 1987). The other study included in this pooled estimate was quasi-randomized (Neilson et al., 1984). Bricker et al. (2008) conducted a sensitivity analysis by excluding this trial and found no significant change in the study findings, possibly due to small sample size of this trial. The same two studies also contributed to the postterm delivery assessment.

Cesarean Section and Induction of Labor: Pooled estimates by Bricker et al. (2008) for Cesarean section were not statistically significant, but the Cesarean section rate was somewhat higher in the screened group (RR 1.06; 21,035 pregnancies; 5 studies), especially for emergency Cesarean section (RR 1.11; 5884 pregnancies; 4 studies). Pooled estimates showed no effect on induction of labor (22,663 patients, 6 studies).

Rate of Abortion for Fetal Abnormality: This outcome was not addressed in the systematic review by Bricker et al. (2008), presumably because the review focused on US in late pregnancy.

Serial US and DUS in Late Pregnancy

Results from a single trial (2834 pregnancies) provided a comparison of serial US and DUS versus selective US during late pregnancy (Newnham et al., 1993). Authors randomized low risk pregnant women into two groups: a treatment group receiving US and continuous-wave DUS at 18, 24, 28, 34, and 38 weeks of gestation (n=1415) and a control group receiving single US imaging at 18 weeks (n=1419). Bricker et al. (2008) reported two statistically significant differences, both of which favored selective imaging without DUS (the control): birth weight less than the third percentile (RR 1.66, 95% CI 1.10 to 2.51) and birth weight less than the 10^{th} percentile (RR 1.36, 95% CI 1.10 to 1.68). The authors of the study argue that this evidence of intrauterine growth restriction may be a chance finding but that it is also plausible that frequent exposure to US may have affected fetal growth. Substantial relative differences in perinatal mortality, stillbirths, and neonatal deaths favored the group with serial US and DUS, but differences were nonsignificant. Results showed no effect on antenatal admissions, use of cardiograph, gestational age at delivery, Apgar score, neonatal interventions, Cesarean section, or induction of labor.

Summary of Efficacy/Effectiveness Evidence for Routine US in Late Pregnancy

The evidence pertaining to US in early pregnancy is derived from large RCTs that were generally without major limitations except lack of blinding. Sensitivity analyses in which the one pseudo-randomized trial was excluded showed no change in effect estimations. Heterogeneity across studies was reported for some outcomes, but the authors used random effects models in these situations. Lack of blinding was considered a weakness in the evidence for impact on Cesarean section and induction of labor. Imprecision and whether more than one study was available were also taken into account in the following quality determinations.

Change in Patient Management: Moderate-quality evidence (imprecision) showed routine US after 24 weeks to have no effect on antenatal admissions or frequency of follow-up US scans. Low-quality evidence (single studies) showed that neither routine US nor routine serial US and DUS had an effect on use of cardiographs, and the study of serial US plus DUS showed no effect on antenatal admissions.

Reductions in Perinatal Mortality and Morbidity: Moderate-quality evidence (imprecision) showed routine US after 24 weeks to have no effect on perinatal mortality or morbidity. Low-quality evidence (single study) suggested that combining routine US with placental grading reduced stillbirths. Low-quality evidence (single study and some imprecision) showed routine serial US plus DUS to have no effect on perinatal mortality, neonatal death, or perinatal morbidity in general; however, there was a clear adverse effect on the risk of very low birth weight.

Cesarean Section and Induction of Labor: Moderate-quality evidence (lack of blinding) showed routine US after 24 weeks to have no effect on the frequency of Cesarean section or induction of labor. Low-quality evidence (single study without blinding) showed routine serial US and DUS to have no effect on the frequency of Cesarean section or induction of labor.

Safety of Routine US

Search Results

One systematic review with meta-analysis was selected to evaluate the safety of US during pregnancy (Torloni et al., 2009) (see Appendix V). This review included RCTs and other prospective and retrospective controlled studies, including case-control studies that assessed any type of short- or long-term effect of at least one in-utero exposure to US during pregnancy in unselected or low-risk pregnant women. A total of 41 studies were selected. All but two (identified in Appendix V by underscoring) of the controlled trials were RCTs; for the sake of simplicity, all controlled trials will be referred to RCTs. Most studies included low-risk women with singleton pregnancy and no previous history of a medical condition. Among the controlled trials, all but two involved US performed in the second trimester. Studies were published in both the United States and Europe. Additionally, the Cochrane Review by Whitworth et al. (2010) provided some pooled estimates of the effect of US on child growth and development. The two reviews used partially overlapping sets of studies. As described previously, only RCTs and quasi-randomized trials were included in the review by Whitworth et al. (2010), and selection was further restricted to studies evaluating a single screening US in early (< 24 weeks) pregnancy. Most trials included only singleton pregnancies. Torloni et al. (2009) concluded that in-utero exposure to US is relatively safe for mother and fetus but cautioned that not all effects, particularly long-term effects, are known. They also were not able to identify the safest use of US in terms of gestational age, US

parameters, or fetal position. Whitworth et al. (2010) did not state a conclusion about the safety of US but did call for more research on long-term neurological effects.

An RCT (Carlan et al., 1997) included in the systematic review by Berghella et al. (2009), an RCT (Newnham et al., 1993) included in the systematic review by Bricket et al. (2008), and a trial selected from the recent primary literature for its evaluation of TVU determination of cervical length (Simcox et al., 2009) also reported safety-related data. Four additional observational studies published after the search time frame observed by Torloni et al. (2009) were selected. These assessed the association between in utero US and non–right-handedness (Rodriguez and Waldenström, 2008), neuroblastoma (McLaughlin et al., 2009), childhood brain tumor (Stålberg et al., 2008), and autism spectrum disorders (Grether et al., 2010).

Dose-Response Relationship Between US Exposure and Fetal Outcomes: Torloni et al. (2009) reported several instances of a dose-response relationship. There was a statistically significant association between a higher number of US exposures (≥ 3 versus 1) and low birth weight (odds ratio [OR] 1.27, 95% CI 1.02 to 1.58). Mean length and head circumference of fetuses exposed to ≥ 3 US scans during pregnancy were also slightly but significantly lower than those exposed to only one scan (weighted mean difference −0.26 cm, 95% CI −0.45 to −0.07 and −0.15 cm, 95% CI −0.29 to −0.01, respectively). Perinatal mortality was reduced in fetuses exposed to one US versus no US (OR 0.56 95% CI 0.40 to 0.78), most likely due to improved diagnosis of any pregnancy-related pathological conditions. Lastly, the risk of non–right-handedness in boys was elevated with two US scans compared with one scan or no scans (OR 1.32; 95% CI 1.02 to 1.71). These calculations were each based on a small number of trials. In addition to the direct dose-response assessments reported by Torloni et al., the findings reported by Bricker et al. (2008) suggest that a more intense regimen of serial US plus DUS may have an adverse effect on the incidence of very low birth weight (see ***Efficacy and Effectiveness: Clinical Utility of US in Low-Risk Pregnancies,*** *Serial US and DUS in Late Pregnancy*).

Maternal Outcomes: Pooled estimates of nine RCTs (total, n=25,200) by Torloni et al. (2009) showed that US during pregnancy does not increase the risks of maternal admission to the hospital. Results of a retrospective cohort study showed that US during pregnancy does not increase postpartum and intrapartum complications, although the authors did not identify specific complications.

One of the RCTs selected for the Cochrane Review by Berghella et al. (2009) randomized 92 women with preterm prelabor rupture of membranes (PPROM) at 24 weeks to 34 weeks to either weekly TVU for monitoring cervical length or no US (Carlan et al., 1997). The incidence of chorioamnionitis (28% versus 20%), endometritis (6% versus 9%), and neonatal infection (17%) was similar in the two groups. The observed differences were not statistically significant. There were three neonatal deaths and all occurred in the TVU group, but they were not apparently related to the TVU probe. The authors of another one of the RCTs (Alfirevic et al., 2007) included by Berghella et al. reported observational data from their institution about the safety of withholding treatment, on the basis of TVU determination of long cervix (> 15 mm), from 216 women with signs and symptoms of preterm labor (< 36 weeks) . Almost all (99%) of the women delivered after 7 days, suggesting a 1% false negative rate. The authors estimated that more than 4000 women with suspected preterm labor would have to be included in a trial designed to show that management based on TVU determination of cervical length was as safe as the conventional approach, i.e. to show statistical noninferiority. A more recent RCT (Simcox et al., 2009) reported that complications (maternal pyrexia, chorioamnionitis in women with preterm birth,

stillbirth or neonatal death, and infant bronchopulmonary dysplasia or intraventricular hemorrhage) were somewhat more frequent in the TVU surveillance group, but differences were nonsignificant.

No systematic search for studies evaluating the safety of TVU was conducted for this report.

Mortality: Pooled data from eight RCTs (total, n=32,962) showed no effect on fetal mortality, and pooled data from 13 RCTs (total, n=46,553) showed no significant effect on perinatal mortality. Pooled estimates of fetal and perinatal mortality from observational studies were also nonsignificant. Pooled data from seven RCTs (total, n=30,942) showed no significant effect on neonatal mortality, but the confidence interval included the possibilities of harm as well as benefit.

Apgar Score and Physical Measures: Pooled estimates from RCTs showed no effect on Apgar score, either at 1 minute (27,299 pregnancies in 10 RCTs) or at 5 minutes (22,150 pregnancies in 12 RCTs), but the confidence interval for Apgar score < 7 at 5 minutes was wide (Torloni et al., 2009). A pooled estimate from two cohort studies (total, n=213,138) suggested a 17% reduction in the odds of 5-minute Apgar score < 7 (Belfrage et al., 1987; Sylvan et al., 2005), but Torloni et al. did not comment on possible confounders. A pooled estimate of mean birth weight (35,894 pregnancies in 9 RCTs) was nonsignificant, but a wide confidence interval included the possibility of both a decrease of 18 g and an increase of 19 g. Nevertheless, there was no effect on the odds of birth weight < 2500 g according to pooled data from nine RCTs (total, n=24,271) or small-for-gestational age (44,745 pregnancies in 13 RCTs), and pooled estimates from smaller sets of RCTs showed no significant effect on very low birth weight, mean length at birth, and mean head circumference at birth. Birth weight effects estimated from observational studies were either nonsignificant or explained by a known confounder.

Perinatal Morbidity: A meta-analysis of RCTs showed no effect on preterm birth (34,049 pregnancies in 10 RCTs), neonatal resuscitation (17,151 pregnancies in 8 RCTs), admission to neonatal special care unit (33,408 pregnancies in 13 RCTs), neonatal seizures (6459 pregnancies in 2 RCTs), or congenital malformations in general (15,281 pregnancies in 2 RCTs) (Torloni et al., 2009). However, pooled data from two cohort studies (total, n=12,451) suggested that pregnant women exposed to US during pregnancy are almost twice as likely to give birth to babies with cardiac defects.

Childhood Cancers: A pooled estimate with wide confidence interval (14,057 pregnancies in 8 case-control studies) found that in utero exposure of US during pregnancy does not increase the risk of childhood cancers such as leukemia, CNS tumors, reticuloendothelial tumors or other childhood malignancies (Torloni et al., 2009). Observational data also showed no association with specific types of tumors. A nested case-control study not included by Torloni et al. reported on the association between US and neuroblastoma (McLaughlin et al., 2009). In a stratified analysis, the ORs for diagnosis at any age, at 1 to 6 months, at 7 to 18 months, and at ≥ 18 months were in the range of 1.0 to 1.3 and were all nonsignificant. The overall study group consisted of 529 cases and 12,010 controls, but 11 cases and 330 controls had missing data pertaining to US exposure. Children without any birth certificate record of US exposure served as the reference. Another case-control study (512 cases and 539 matched controls) reported a nonsignificant OR of 0.94 for the association between US exposure and any brain tumor; ORs by type of tumor were also nonsignificant (Stålberg et al., 2008).

Child Growth and Development: Pooled estimates of two RCTs (total, n=2422) and two cohort studies (total, n=40,841) showed a statistically significant but weak association between US exposure and being non–right-handed in boys (Torloni et al., 2009), whereas neither Torloni et al. nor Whitworth et al.

(2010) found any significant association between US exposure and non–right-handedness in children in general (4715 pregnancies in 2 studies in each review; partial overlap between the two sets of studies, same total n is coincidental). A prospective cohort study (n=1714) included in neither systematic review conducted a stratified analysis by number of US scans (1 to > 5) (Rodriguez and Waldenström, 2008). The study found that after adjustment for other risk factors, there was no significant association between US and non–right-handedness, mixed-handedness, or left-handedness within any stratum, except for a protective effect on left-handedness in children who had been exposed to > 5 scans (OR 0.22, 95% CI 0.05 to 0.97). (NOTE: Rodriguez and Waldenström also found that although non–right-handedness and mixed handedness were associated with language difficulties and attention deficit hyperactivity disorder [ADHD] symptoms, left-handedness was not. Mixed handedness was defined as no preference for either hand or a preference for the right hand with some activities and a preference for the left hand with other activities.)

Overall, pooled data from small sets of RCTs showed no effect on any measure of speech development although one RCT (Salvesen et al., 1994) showed a protective effect in terms the likelihood of being referred to speech therapist. In general, exposure of US in utero did not negatively affect growth and neurological development during infancy. Pooled data from small sets of RCTs demonstrated no effect on childhood visual problems, various measures of neurological development, dyslexia, school performance. Data from two cohort studies (total, n=40,841) suggested a small increase in non–right-handedness for in boys. Some estimates were imprecise, i.e., confidence intervals ruled out neither benefit nor harm. These findings are further confirmed by the pooled estimates of RCTs by Whitworth et al.(2008) that showed that US during pregnancy does not increase the risks of impaired childhood development, school performance in terms of poor oral and comprehension reading, poor spelling and arithmetic at school, dyslexia, auditory and visual impairment, and ambidexterity. This conclusion was based on the findings of a study (Salvesen et al., 1993) that was common to both reviews. A recently published case-control study (362 cases and 393 controls) found that after adjustment for other factors, there was no association between US exposure and autism spectrum disorder (Grether et al., 2010).

Development after Childhood: One large cohort study reported a weak association between US exposure and subnominal intellectual performance among adult men exposed to US in utero (OR 1.19, 95% CI 1.12 to 1.27; 167,059 pregnancies) (Kieler et al., 2005). Evidence from a single cohort study (n=370,945) (Stålberg et al., 2007) showed no association with schizophrenia or with other psychoses.

Summary of Safety of Routine US

Most of the outcomes assessed in the systematic review by Torloni et al. (2009) were addressed by RCTs, the bulk of which were ranked by Torloni et al. as “B” for unclear allocation concealment on an A-B-C scale. Adequacy of allocation concealment was the only factor considered for RCTs. Torloni et al. provided no other details on study quality, such as loss to follow-up. It should be noted that method of allocation concealment was not a required reporting element at the time the available studies were published and the “unclear” ratings may reflect lack of reporting rather than lack of adequate concealment. The following summary assumes that bodies of evidence consisting of ≥ 2 RCTs are of moderate quality unless CIs were imprecise. Bodies of evidence consisting of observational studies only will be considered low quality unless there the quantity of data is very large and/or there is a strong association.

Moderate-quality evidence from RCTs shows a dose-response relationship between the degree of US exposure (≥ 3 versus 1 scan) and reduced birth size and non–right-handedness (boys only), but increased exposure reduces rather than increases perinatal mortality. Moderate-quality evidence from a large volume of RCTs has shown that *overall*, routine US does *not* adversely affect maternal admission to the hospital, fetal mortality, perinatal mortality, or multiple measures of perinatal morbidity. Furthermore, low-quality evidence suggests no impact on postpartum complications, mean birth weight, or Apgar score < 7 at 5 minutes. Although moderate-quality evidence from two large RCTs showed no increase in congenital malformations in general, low-quality evidence from a single large cohort study suggests US in pregnancy nearly doubles the risk of congenital cardiac defect. Moderate quality evidence from a large volume of observational data shows no association between US and childhood cancers. Low-quality evidence has not shown an association between US during pregnancy and childhood growth and development, including cognitive development. Moderate-quality evidence shows no overall effect on non–right-handedness, but an increase in boys. Low quality evidence suggests that US in utero may have a small adverse effect on intellectual performance in men but does not increase the risk of mental illness. Very-low-quality evidence suggests that use of TVU to monitor cervical length in women at imminent risk of preterm birth does not increase the risk of infection or other complications. Furthermore, most studies were conducted prior to 1995; the results of exposure with newer machines with greater acoustic potency may be different.

Differential Efficacy or Safety Issues in Subpopulations

Except for a systematic review of US in the emergency department for assessment of first-trimester bleeding (McRae et al., 2009), evidence of differential efficacy or safety comes from reviews and studies already identified for other key questions.

Gestational Age: The two Cochrane Reviews of US for fetal assessment in low-risk or unselected populations dealt separately with US in early pregnancy (< 24 weeks) and late pregnancy (> 24 weeks) (Bricker et al., 2008; Whitworth et al., 2010). The meta-analyses included in these two reviews suggest that routine US performed in early pregnancy is effective in reducing induction of labor (RR 0.78, CI 0.63 to 0.97) while routine US performed in late pregnancy is not (RR 0.93, not significant [NS]). US had no effect in early or late pregnancy on the other outcomes in common between the two reviews (perinatal death, mean birth weight, Apgar score ≤ 7 at 5 minutes).

Whitworth et al. (2010) analyzed differences among studies in which US was performed before or after 14 weeks (before 14 weeks would be first trimester). The risk of not detecting multiple pregnancy by 24 weeks to 26 weeks was greatly reduced when US was performed after 14 weeks (RR 0.06, CI 0.02 to 0.16), but US had no effect when performed before 14 weeks (RR 0.89, NS). The before-14-weeks results come from a single study of only 9 patients, while the after-14-weeks results were from 6 studies (total, n=286). The risk of induction of labor for postterm pregnancy was also reduced after 14 weeks (RR 0.49, CI 0.31 to 0.77; 5 studies, 23,434 patients) but not earlier (RR 0.99, NS; 2 studies, 1278 patients). No subgroup differences were observed for the effect on perinatal death (34,923 pregnancies in 9 RCTs). The study of US exposure and childhood brain tumors found no differential effect by trimester of exposure (first, second, or third) (Stålberg et al., 2008), and the study of US and autism found no effect by trimester (Grether et al., 2010).

Alfirevic et al. (2010) did not identify any studies of high-risk patients in which US was performed before 24 weeks. Torloni et al. (2009) planned a subgroup analysis of US exposure in the first trimester versus

third trimester but found no controlled trials that used US in the first trimester and only one controlled trial using US in the third trimester.

Other Patient Characteristics or Evidence-Based Patient Selection Criteria: Alfirevic et al. (2010) conducted subgroup analyses according to singleton versus multiple birth and according to five risk factors (small for gestational age or IUGR, hypertension/preeclampsia, diabetes, prolonged pregnancy, and previous pregnancy loss). The effect of US screening in high-risk pregnancies did not significantly differ among subgroups or between subgroups and overall estimates with one exception: although confidence intervals overlapped between the estimated RR of infant intubation or ventilation in singleton pregnancy (2.89, NS) and the RR in multiple pregnancy (0.86, NS), a statistical test for interaction suggested that US did have an effect in singleton pregnancies. The authors offered no explanation for why screening with US would increase the risk of infant intubation. One of the key RCTs (Ewigman et al., 1993) (n=15,151) included in the review of routine US in early pregnancy (Whitworth et al., 2010) and late pregnancy (Bricker et al., 2008) found no differential effects across subgroups defined by postterm pregnancy, multiple births, or small-for-gestational-age infants. The authors of the Cochrane Review on routine US in late pregnancy planned a subgroup analysis by low-risk versus unselected patients and also according to the purpose of the US exam, but the data did not lend themselves to these analyses (Bricker et al., 2008). The study of US and autism found no effect by sex (Grether et al., 2010).

Type of Scanning Machine and Software, Reader Training, and Other Operational Factors: Torloni et al. (2009) conducted a subgroup analysis comparing the in utero exposure of B-mode (routine) US and DUS on perinatal, neonatal, and maternal outcomes. No statistically significant increased risk was reported with the use of DUS compared with routine US, except for one study that reported the mean length of infants exposed to DUS was shorter than in infants exposed to routine US (OR –0.26 cm, 95% CI −0.45 to −0.07 cm) (Newnham et al., 2004). Another RCT, as noted in the discussion of evidence for the effectiveness of US in late pregnancy (Bricker et al., 2008), found that serial US with DUS, compared with selective US, in late pregnancy is associated with low birth weight (Newnham et al., 1993).

No complications were reported in the studies selected for Hayes reports on 3D/4D versus 2D US (Hayes, 2005a; Hayes, 2005b; Hayes, 2005c; Hayes, 2005d; Hayes, 2006a; Hayes, 2006b).

No other evidence was available pertaining to differential effectiveness or safety according to these types of factors.

Provider Type, Setting, or Other Provider Characteristics: A systematic review has shown that emergency department targeted ultrasonogrphy (EDTU) in women presenting to the emergency department with first trimester bleeding may lead to more efficient rule-out of ectopic pregnancy (McRae et al., 2009). (See Appendix VI.) Studies of clinical utility were selected if they measured objective outcomes and compared EDTU with US performed in radiology departments or by gynecology consultants. Eight studies (n > 1778; one study did not report sample size) assessing the effect of EDTU on surgical rupture, time to diagnosis, treatment of ectopic pregnancy, or emergency department length of stay (LOS) were selected. Most were retrospective chart reviews and three studies were published only as abstracts. Two studies (total, n=131) showed that time to surgery was significantly reduced by a mean of 145 minutes to 211 minutes in patients with ectopic pregnancy. Five studies (total, n=1419) showed a significant reduction in emergency department LOS by 59 minutes to 149 minutes, which represents a reduction in the burden on patients. Two of the five studies assessing LOS (total, n=1534), including the

largest study selected by McRae et al., excluded patients with ectopic pregnancy; thus LOS evidence applies largely to the effectiveness of EDTU in confirming IUP, not in accelerating the diagnosis and treatment of ectopic pregnancy. A separate analysis by McRae et al. found EDTU to be highly specific for the detection of intrauterine pregnancy (IUP). Only one study (n=340) assessed actual clinical outcomes; this study showed that the proportion of patients who were found to have ectopic pregnancy were less likely to rupture during surgical exploration; time to surgery was not measured in this study. The authors also pointed out that a pitfall of using EDTU to exclude ectopic pregnancy is that it is inadequate for situations where there are twins and only one is ectopic.

Health Care System Type, Including Worker's Compensation, Medicaid, State Employees: No studies or systematic reviews were identified that looked at the effectiveness of US according to these parameters.

Different US Protocols: A systematic review of different regimens for fetal surveillance for impaired fetal growth concluded that there was insufficient evidence to form a conclusion about the optimal protocol for US (Grivell et al., 2009). The authors looked for RCTs and quasi-randomized trials comparing twice weekly surveillance with testing every 2 weeks. Testing included biophysical profile, nonstress test, umbilical artery and middle cerebral artery Doppler, and uterine artery Doppler. A single pilot RCT (n= 167) was selected. Since only one perinatal death and no cases of serious neonatal morbidity occurred, the effect on these outcomes could not be assessed. There was no difference in emergency Cesarean section, but induction of labor was significantly higher in the twice-weekly group (RR 1.25, 95% CI 1.04 to 1.50); absolute differences were not reported.

Summary: Routine US performed after 14 weeks but before 24 weeks (roughly, second trimester), is effective in reducing the risk of failure to detect multiple pregnancy (low-quality evidence) and the frequency of induction of labor (moderate-quality evidence), whereas routine US performed before 14 weeks (first trimester) or after 24 weeks (roughly, third trimester) does not have these effects. However, the impact on perinatal mortality does not differ between first and second trimesters (high-quality evidence). There are no data specifically pertaining to the safety of US in the first or third trimester, except for two case-control studies showing no association between US exposure in the first, second, or third trimester and either childhood brain tumor or autism. Low-quality evidence has shown no difference in the rate of Cesarean section between twice weekly and every-other-week surveillance for impaired fetal growth, but an increase in the rate of induction of labor with the more frequent regimen. Very-low-quality evidence has suggested that routine serial US plus DUS in late pregnancy does not improve outcomes and may reduce birth size. Very-low-quality evidence suggests that US performed in the emergency department rather than by radiological or gynecological specialists may lead to more efficient rule-out of ectopic pregnancy and improved outcomes. Very-low- to low-quality evidence (has failed to show differential impact on outcomes of DUS screening in high-risk patients, comparing either singleton with multiple-birth pregnancies or comparing patients with different risk factors. Low-quality evidence suggests that routine US in low-risk or unselected patients does not differ in its effects according to maternal or fetal risk factors. No other evidence pertaining to differential effectiveness was available.

NOTE: Economic modeling studies have estimated differential impact on certain outcomes and differential cost-effectiveness according to clinical protocol and high risk subpopulations. Cost descriptions have provided data specific to certain settings. See following discussion.

Cost Implications and Cost-Effectiveness

See Table 5 for a summary of economic evaluations.

Short-Term Costs

Estimated costs from a payer perspective for an US exam in the United States have included $200 in 1998 dollars (Vintzileos et al., 2000) and $52 (range $43 to $74, year unreported) for Medicaid reimbursement (Cahill et al., 2010). Consumer-oriented websites give current prices of $200 to $440 for the cost of a fetal US. The RADIUS trial (Ewigman et al., 1993), the largest in the systematic reviews by Whitworth et al. (2010) and Bricker et al. (2008), found that in the control group, which underwent US only for medical indications, an average of 0.6 scans per pregnancy were performed, whereas an average of 2.2 scans were performed when women underwent a screening exam at 15 to 22 weeks and again at 31 to 35 weeks. The RADIUS study involved 109 practices in the United States.

Durston et al. (2000) (United States): A retrospective chart review of all evaluable women (n=120) with ectopic pregnancy treated during a 6-year period in a single emergency department showed no difference in the proportion of patients with missed ectopic pregnancies or in time to treatment of ectopic pregnancy, comparing three consecutive time periods between 1992 and 1998. These three periods were characterized by (1) availability of an on-call US technician but discouragement of emergency physicians from ordering formal US without obstetrics and gynecology consultation; (2) availability of the technician and no restriction on the emergency physicians for ordering US; and (3) EDTU. The authors estimated that the third approach, by eliminating the need for an on-call technician, could save $299 to $1244 per case of ectopic pregnancy, depending on the cost of implementing EDTU.

Vanara et al. (2004) (Italy): Researchers in Italy concluded on the basis of decision analytic modeling that an organized US screening program for detecting structural fetal abnormalities would save overall costs (Vanara et al., 2004). An organized program was assumed to include an active population-wide offer (e.g., with mailings) of a single US screening scan in the second trimester, follow-up diagnostic tests when indicated, defined treatments, follow-up for diagnosed cases, outcome evaluation, and quality control; additional private exams were assumed. The comparator was the current practice situation of a mean of three public and/or private US examinations per pregnancy. The analysis was from the perspective of a public payer, but the costs of private US exams were included. The following direct costs were included: US screening; follow-up diagnostic US; additional tests when needed; termination of pregnancy; fetal death; lifetime care of malformed infants (medical, development, and special education costs), discounted by 5%; and implementation of the organized program. The authors assumed that every eligible woman would receive one screening US and that 58% of women with a diagnosis of fetal anomaly would elect abortion. Results showed that with the organized program the number of US exams would decrease, diagnosis of malformations would increase, and the number of malformed infants would decrease. Short-term costs associated with US examinations would decrease 44%.

NOTE: The base case assumptions of US sensitivity for fetal structural abnormality were based on data published between 1998 and 2000 (61.4% for usual care and 68% for the organized program); a more recent estimate specifically for structural abnormalities was not identified for this report. The study had no commercial sponsorship; there was no disclosure of author conflict of interest.

Ritchie et al. (2005) (Scotland): A modeling study conducted from the perspective of the Scotland national health service compared direct medical costs across six competing strategies for US screening for any fetal abnormality (Ritchie et al., 2005). The six strategies represented various combinations of first trimester US nuchal translucency assessment or US assessment of gestational age, first trimester and/or second trimester serum testing, and inclusion or exclusion of second trimester US; all six strategies involved at least one US exam. The authors used a single pivotal trial to estimate the accuracy of US nuchal translucency Down syndrome. They took estimates of the sensitivity of US for various structural abnormalities from published studies. They surveyed clinicians for estimates of the proportion of women who accept various tests, including invasive follow-up tests. The cost of performing the US exams, serum tests, and follow-up tests, as well as the cost of offering those tests in a manner that assured informed consent, were included. The cost of test consequences (amniocentesis, chorionic villus sampling, iatrogenic miscarriage related to these procedures, other spontaneous miscarriage, and elective termination) were also included.

The least expensive strategy entailed first trimester US for estimating gestational age plus second trimester double serum test for Down and Edward syndromes and for neural tube defects. The model predicted that such a strategy would detect 65% of abnormalities, miss 96 abnormalities per 50,000 live births, and result in 7.9 iatrogenic losses per 50,000 pregnancies. Using this strategy as the reference strategy, the authors calculated incremental cost-effectiveness ratios (ICERs). The two most cost-effective alternatives entailed a first trimester US for nuchal translucency plus serum screening for Down and Edward syndrome and either no further screening or a second trimester anomaly US scan for each additional abnormality detected. The least cost-effective strategy included the same elements plus a second trimester serum test for neural tube defect. The numerical ICERs are not reported here because only some of them could be reproduced in hand calculations using the formula and data specified in the study report; there appear to be errors in the article.

NOTE: This analysis did not include a strategy in which *no* US screening exam was performed and included only conditions considered serious enough to warrant an offer of pregnancy termination. The authors also point out that their analysis is limited by the fact that published sources do not typically provide false positive rate for second trimester US but rather the proportion of false positives that reach birth or pregnancy termination. No sensitivity analysis was conducted. The literature on accuracy was not reviewed in sufficient detail for this report to allow comment on the reasonableness of the sensitivity assumptions made by Ritchie et al.

Pierce et al. (2001) (setting assumed to be United States): In this unpublished RCT (n=69), per-patient hospital charges were less in patients evaluated with US in the emergency department for first trimester bleeding ($535) than in patients who were referred for US evaluation in a radiology department or by a gynecologist ($926), but the difference was statistically nonsignificant.

Vintzileos et al. (2000) (United States): A cost-effectiveness study from a societal perspective in the United States found that routine second trimester US screening for fetal anomaly could generate cost savings (Vintzileos et al., 2000). The authors estimated that 1 million women per year would be candidates for a second trimester screening exam, i.e., an exam not predicated by any particular indication. Sensitivities for various anomalies were derived from the Routine Antenatal Diagnostic Imaging with Ultrasound (RADIUS) trial (Ewigman et al., 1993); there were separate estimates for tertiary and nontertiary centers. A single published study was the source of an estimate that 66% of women with a diagnosis of serious fetal anomaly would choose pregnancy termination. Short-term costs

(1998 dollars) included the cost of screening US exams ($200 per patient, according to a published fee schedule), follow-up amniocentesis, and follow-up US. RADIUS estimates of the rate of other averted services, combined with data from additional sources, were used to compute additional cost savings from averted use of tocolysis, postterm US fetal testing, and induction of labor. Lastly, savings were calculated for averted lost wages due to hospitalization for tocolysis and postterm fetal testing. Total estimated short-term cost savings were $13,030 per screened patient when initial screening was performed in tertiary centers and $2230 when initial screening was in nontertiary centers.

NOTE: There was no disclosure of author conflict of interest; funding was from the Robert Wood Johnson Foundation. The RADIUS trial was included in the Cochrane Review by Bricker et al. (2008), who considered the studies in general to be a good quality. The RADIUS trial was reviewed for this report, and no serious weaknesses were identified. Results of the cost-effectiveness analysis may not be applicable to situations in which screening identifies a mild or moderate abnormality.

Long-Term Costs

Cahill et al. (2010) (United States): A cost-effectiveness modeling study in the United States explored different options for identifying women who might benefit from the recently discovered potential of progestins in prevention of preterm birth (Cahill et al., 2010). Universal transvaginal ultrasound (TVU) assessment of cervical length followed by progesterone to treat short cervix was found both to prevent preterm births and to save costs, except among women considered high risk because of a history of preterm birth**.** The following four strategies were defined, with the assumption that all strategies would be in conjunction with routine second trimester US anatomic survey and that patients had singleton pregnancies without anomalies:

(1) Universal TVU screening of cervical length at the time of routine second trimester US anatomic survey (18 to 23 weeks) followed by daily vaginal progesterone for women with a short cervix (≤ 15 mm).
(2) High-risk TVU screening, timed as in strategy 1, for women at risk of preterm birth because of previous preterm birth, followed by vaginal progesterone for women with a cervical length ≤ 15 mm.
(3) High-risk, prophylactic treatment with 17-∝-hydroxyprogesterone caproate (17-OHP-C) of *all* patients at with previous preterm birth; no screening before treatment.
(4) No screening; no treatment.

Pooled estimates and ranges with extreme values for the probabilities of various outcomes and for utilities (preferences) associated with neonatal death and abnormality were derived from the literature. Costs included TVU exam, vaginal progesterone or 17-OHP-C for weeks 18 to 34, and lifetime care due to neonatal severe morbidity. Cost estimates were based on local Medicaid reimbursement rates where possible and otherwise were derived from the literature. The assumed cost-effectiveness threshold was $100,000 per quality-adjusted life-year (QALY). The model for a general population showed strategy 1 (universal TVU screening) to be dominant over the other strategies: it was the least costly, and all other strategies produced QALY *losses* compared with strategy 1. Compared with strategy 4 (no screening or treatment), strategy 1 was predicted to prevent 95,920 preterm births out of 4 million deliveries per year, at a total cost savings of $129,400,000. The other strategies (2 and 3) of screening and/or treatment only high-risk patients also prevented preterm births and saved costs but to a lesser degree.

Subpopulation analyses were also conducted. The model for women at high historical risk of preterm birth (280,000 pregnancies per year), either a strategy of screening followed by progesterone if indicated or a strategy of prophylactic treatment of with 17-OHP-C was found to reduce the number of preterm births but to also be more costly, compared with no screening or treatment as the reference. However, compared with prophylactic treatment, screening in high-risk women prevented fewer preterm births and was more costly. For the other subpopulation, women without known risk of preterm birth, screening was found to net a cost savings of $9982 per QALY gained, compared with no screening or treatment as the reference. The authors concluded that TVU screening for short cervical length and treatment with vaginal progesterone may be a cost-effective approach to reducing preterm birth but that further study is necessary to determine whether this approach is practical.

NOTE: There was no commercial funding; no disclosure of author conflict of interest was provided. The quality of studies used for estimates of the sensitivity of US assessment of cervical length and for outcomes of treatment were fair to good on the scale previously used by the U.S. Preventive Services Task Force. One-way and multi-way sensitivity analyses using Monte Carlo multivariate simulations showed the results from the main model to be robust, that is, not sensitive to wide variations in all estimates. The same was true of the model for women without risk of preterm birth. For high-risk women, screening rather than prophylactic treatment became the preferred strategy if the estimate for reduction is preterm birth risk associated with 17-OHP-C were substantially lower. Results may not apply to multiple gestation, pregnancy with fetal abnormality, treatment of short cervix defined at thresholds > 15 mm, prevention of preterm births at other than < 34 weeks, prevention of preterm birth with treatments other than progesterone, assessment of cervical length in women with symptoms of preterm labor, or where second trimester US screening was not already considered routine. Annual discounting[4] of 3% was applied to cost and QALY estimates. Results might not apply when costs are based on private payer reimbursement rates.

Vanara et al. (2004) (Italy): In addition to the short-term cost savings reported by Vanara et al. (2004) in an Italian study, total costs would decrease by 21% with the institution of an organized screening program, as opposed to usual practice. Sensitivity analyses based on higher and lower estimates of diagnostic sensitivity, cost, and the number of private scans showed cost savings ranging from 15% to 30%.

NOTE: The base case sensitivity assumptions (61.4% for usual care and 68% for the organized program) were derived from data published in 1998 and 2000 lower; a more recent estimate was specifically for structural abnormalities was not identified for this report.

Vintzileos et al. (2000) (United States): In addition to the short-term costs and cost savings discussed previously, this study in the United States considered the lifetime incremental direct costs (medical, developmental, special education), and indirect costs (lost productivity) for individuals born with a serious anomaly, using estimates from the RADIUS trial. Final calculations showed a net cost *savings* of $97 to $189 (1998 dollars) per patient if initial screening was performed in tertiary centers and net *loss* of $69 to $161 per patient if screening was performed in nontertiary centers.

[4] Discounting in this sense refers to an adjustment for the lesser cost a postponed expenditure and the diminished value of a delayed benefit.

NOTE: No discounting of lifetime costs was reported.

Summary

No definitive statements may be made about the cost or cost-effectiveness of US in pregnancy, primarily because assumptions about the impact on outcomes have not been based on a comprehensive review of current literature, costs were not trial-based, and three of four cost-effectiveness studies made no comparison with a strategy in which there was no routine US screening. Two economic evaluations have suggested that the use of second trimester US to screen for fetal anomaly may save costs. The first, a cost-effectiveness modeling study (Vintzileos et al., 2000) a United States societal perspective, suggested that universal second trimester US screening for fetal anomaly may generate short-term, direct medical cost savings of $2312 to $13,376 per patient screened, depending on whether the screening were conducted in a nontertiary care center or a tertiary center. The same study also showed that long-term costs, including care for and loss of productivity in individuals born with abnormality, would be reduced with the use of US screening but only if the screening were conducted in a tertiary center. The other economic evaluation, also a modeling study, (Vanara et al., 2000) showed that in Italy, a structured program of universal US screening for fetal abnormality, combined with well-defined protocols, has the potential of reducing short- and long-term costs, as well as reducing the incidence of birth with structural abnormality. Results are not likely to apply to the United States healthcare system, which is considered to be a fragmented system.

Other studies have suggested that the cost implications of universal routine second trimester anomaly screening depend on the patient population and other care being offered. Recent modeling evidence from the United States (Cahill et al., 2010) suggests that compared with strategies that do not include screening for short cervix, universal TVU screening of women with no history of preterm birth, followed by treatment with vaginal progesterone for short cervix, may prevent preterm birth and save direct costs, taking into account the long-term costs associated with caring for individuals born with serious abnormality. The same study suggests that use of US screening and progesterone treatment in *high risk* women reduces the risk of preterm birth but does not save costs and is less cost-effective than a strategy of treatment with 17-OHC-P without screening. Results apply only to situations where the cervical length screen is an add-on to routine second-trimester US fetal assessment. A modeling study in Scotland (Ritchie et al., 2005) demonstrated that if some use of US in pregnancy is assumed, different combinations of US and serum testing have different cost implications, but the study included no comparison with a strategy in which no routine US exams were used.

Performance of EDTU for evaluation of suspected ectopic pregnancy could generate short-term cost savings, according to a retrospective chart review (n=120) and an RCT (n=69) that was published only as an abstract. These findings, published in 2000 and 2001, are consistent with those reported by McRae et al. (2009) (see **Differential Efficacy or Safety Issues in Subpopulations**).

CONCLUSION

Evidence-Based Conclusions

Tables 2 through 5 provide an overview of findings.

Effectiveness in High-Risk Pregnancy: The evidence provides some support for the use of Doppler US to monitor high-risk patients and the use of TVU to identify patients in need of prophylactic treatment because of imminent risk of preterm birth. The evidence does not support the use of TVU surveillance in women with a history of preterm birth. In general, evidence pertaining to these indications is of very low to low quality.

The use of Doppler ultrasound (DUS) for monitoring high-risk patients has been shown to reduce antenatal admissions (moderate-quality evidence). DUS may reduce overall perinatal mortality as well as perinatal morbidity, but pooled estimates and measures of statistical variability have suggested the possibility of very small effects (very low-quality evidence). More than 200 pregnant women would have to be screened to prevent one perinatal death and 36 women would have to be screened to prevent one incidence of neonatal morbidity. DUS may reduce the incidence of Cesarean section and induction of labor (low-quality evidence). The number of women who would have to be screened to prevent one event is 39 (any Cesarean section), 47 (emergency Cesarean section), or 31 (induction of labor). High-quality evidence has shown that compared with cardiograph (CTG) alone, DUS alone increases the rate of elective Cesarean section but reduces the rate of emergency Cesarean section (12 patients would have to undergo DUS assessment rather than to prevent one emergency Cesarean section).

Low-quality evidence suggests that assessment of cervical length with transvaginal ultrasound (TVU) in women who already have signs or symptoms of preterm labor does not increase utilization rates and may reduce unnecessary prophylactic treatment and reduce preterm births. However, moderate-quality evidence has shown that surveillance with TVU on the basis of obstetric history does *not* prevent preterm birth or early pregnancy loss (which is a possible outcome from preterm birth). Low-quality evidence suggests that surveillance reduces the use of cerclage as prophylactic treatment for cervical insufficiency, but there was some conflicting evidence from a trial that demonstrated an increase in utilization rates for numerous services.

Lack of standard protocols for intervention when DUS is abnormal or TVU shows short cervical length hampers the generalizability of pooled results to particular settings.

Effectiveness in Low-Risk Pregnancy, Early Screening: According to moderate- to high-quality evidence, routine ultrasonography (US)in early pregnancy (< 24 weeks) does not change patient management or improve health outcomes or substantially affect delivery mode, at least not in high-resource settings. High-quality evidence indicates that a single routine US in high-resource settings has no effect on hospital utilization although, according to low-quality evidence, it may reduce inappropriately-timed serum scans and repeat fetal anomaly scans. High-quality evidence from large randomized controlled trials (RCTs) supports a conclusion that in high-resource settings, routine US does *not* reduce perinatal or neonatal mortality or morbidity even though it doubles the rate of abortion for fetal anomaly (mortality and morbidity can be due to conditions unrelated to a congenital abnormality). Moderate-quality evidence shows no effect on the frequency of Cesarean section in high-resource settings but a modest reduction in the frequency of induction of labor. US in this population more than doubles the

rate of abortion for fetal anomaly, but the absolute increase is only 0.10 percentage points (high-quality evidence).These various findings might not apply to low-resource settings where perinatal mortality is high and not as likely to be attributable to fetal abnormality and might not apply to all strategies for US timing and follow-up intervention.

Effectiveness in Low-Risk Pregnancy, Late Screening: Low- to moderate-quality evidence has not shown routine US in late pregnancy (> 24 weeks) to change patient management, affect delivery mode, or improve health outcomes. Low- to moderate-quality evidence has shown no effect on antenatal admission, follow-up US scans, or use of in cardiographs. To date, the evidence shows no effect from either routine US (moderate-quality evidence) or routine serial US plus DUS (low-quality evidence) in late pregnancy on perinatal mortality or morbidity. However, a single study suggests that routine US combined with placental grading could reduce stillbirths (low-quality evidence). Moderate-quality evidence indicates that routine US in late pregnancy has no effect on the frequency of Cesarean section or induction of labor.

Safety: Moderate-quality evidence for major outcomes has shown US to be a reasonably safe procedure with no serious short-term adverse effects overall although delivery of multiple scans (≥ 3), compared with a single scan, has a modest effect on birth size (moderate-quality evidence), and DUS compared with no US may be associated with a reduction in birth size (low-quality evidence). A large volume of moderate-quality evidence from RCTs has shown that routine US during pregnancy does not adversely affect maternal hospitalization, fetal or perinatal death, or perinatal morbidity. Furthermore, low-quality evidence shows no impact on postpartum complications, Apgar score, or birth weight. Moderate-quality evidence shows no impact on the overall rate of congenital abnormality, but there is low-quality evidence to suggest that the risk of cardiac congenital abnormality is increased. Moderate-quality evidence shows no association with childhood cancer. Evidence of mixed quality suggests no general impact on developmental outcomes after birth but further research, particularly with respect to neurological development, is needed to allow definite conclusions. According to low-quality evidence, use of TVU to measure cervical length does not increase the risk of infection. The applicability of most of the safety evidence is diminished by the fact that most studies were using older, weaker machines. There is also very little evidence on the safety of US performed in the first and third trimesters.

Differential Effectiveness and Safety: The evidence pertaining to differential effectiveness and safety does not address all potentially useful comparisons and is of variable quality. Routine US performed between 14 and 24 weeks (second trimester) is most likely to detect multiple births (low-quality evidence) and to reduce the frequency of induction of labor (moderate quality), compared with US at other gestational ages. However, high-quality evidence shows no differential effect by gestational age on perinatal mortality, and very-low-quality evidence has shown no differential effect on childhood brain cancer or autism. US screening is associated with non–right-handedness in boys although no general effect has been demonstrated (moderate quality evidence). Attempts to assess differential effects according to multiple versus singleton pregnancy and material risk factors have failed to show significant differences (low-quality evidence). Very-low-quality evidence has suggested that serial US plus DUS, compared with routine US, does not improve outcomes and may reduce birth size. Performance of US in the emergency department for evaluation of first-trimester bleeding rather than by radiological or gynecological specialists may be less burdensome to the patient and improve surgical outcomes in ectopic pregnancy (very-low-quality evidence). There was no evidence pertaining to the effect on outcomes of different types of US scanning software, reader training, operational factors (other than the emergency department and ectopic pregnancy issue), or type of healthcare financing

system. More frequent surveillance may increase the rate of Cesarean section and induction of labor (low-quality evidence).

Cost Implications and Cost-Effectiveness: No definitive statements can be made about the cost or cost-effectiveness of US in pregnancy because assumptions about the impact on outcomes have not been based on a comprehensive review of current literature, costs were not trial-based, and three of four cost-effectiveness studies made no comparison with a strategy in which there was no routine US screening. There is preliminary evidence that routine use of second-trimester US to screen for fetal anomaly may reduce short- and long-term costs. There is also preliminary evidence that universal TVU screening for short cervix may prevent preterm birth and save direct short- and long-term costs but only in low-risk pregnancies (no previous preterm birth) and only as an add-on to routine second trimester US fetal assessment. There have been no economic evaluations of US in other types of high-risk pregnancy. Performance of US to rule out ectopic pregnancy may be less costly if performed in the emergency department.

Gaps in the Evidence

Given the low quality of evidence for US in high-risk patients, large RCTs powered to detect clinically meaningful differences in perinatal mortality, perinatal morbidity, and other outcomes of interest are needed. However, unless the most effective follow-up protocols can be defined for abnormal US results, future evidence will continue to be weakened by heterogeneity. Alfirevic et al. (2010) have pointed out the lack of clear guidelines on the optimal frequency of DUS studies and timing of delivery when DUS is abnormal. This has led to protocol heterogeneity across studies and makes precise conclusions about US in high-risk pregnancy difficult. Alfirevic et al. (2010) also stated that no conclusions are possible regarding more sophisticated DUS tests such as assessment of blood flow in the middle cerebral artery and ductus venosus because of lack of evidence. Berghella et al. (2009) also acknowledge that there is uncertainty regarding the most effective interventions following determination of cervical length with TVU.

Similarly, evidence is lacking as to whether routine US in low-risk pregnancies might be effective with specific strategies. The safety and effectiveness of US performed in the first and third trimesters has not been adequately studied. Most included studies of both high- and low-risk patients were performed before 2000. Significant improvements in the acoustic potency of US machines have been made since then. Assessment of outcomes and adverse events associated with newer machines, including 3D and 4D, is needed. Furthermore, randomized controlled trials conducted in the United States are needed to confirm that findings in European studies are applicable in the United States population and healthcare system.

Not all subpopulations have been studied. For example, there were no assessments of the utility of US in obese women, in whom signals are weaker, or in women with low socioeconomic status, who are more likely to have poor pregnancy outcomes.

More direct evidence is needed regarding impact on clinical decision making, especially studies designed to directly demonstrate whether clinicians change management plans after receiving information from US scans; no such studies were identified for this report. The comparative effectiveness of US by type of setting, e.g., tertiary, secondary, or primary care, has not been addressed except in a single cost-effectiveness modeling study, and evidence applicable to low-resource settings is limited.

Cost-effectiveness studies have not addressed the full range of US applications and have relied largely on modeling rather than trial-based cost data.

LIMITATIONS OF THIS REPORT

The following limitations apply to the methodology used for this report:

- No critical appraisal of the evidence pertaining to the accuracy of US for detecting anomalies or predicting pregnancy outcomes.
- No review of the evidence of differential accuracy by subpopulation, setting, type of US, or any other factor.
- No assessment of the clinical utility (e.g., reduction in amniocentesis) and potential patient harms (undue anxiety) of ultrasound (US) testing for Down syndrome and no assessment of the cost-effectiveness of US testing specifically for Down syndrome. However, the cost-effectiveness of testing for fetal abnormality in general (chromosomal and structural) has been addressed.
- Exclusion of twin-to-twin transfer syndrome (TTTS). This indication was omitted from the systematic review selected for evidence pertaining to US in high-risk pregnancies (Alfirevic et al., 2010) and no search for related studies was conducted for this report.
- Exclusion of utero-placental Doppler. This indication was omitted from the systematic review selected for evidence pertaining to US in high-risk pregnancies (Alfirevic et al., 2010) because of a forthcoming Cochrane Review on this topic.
- No assessment of the impact on abortion rate for fetal anomaly in high-risk pregnancies. This outcome was not addressed by Alfirevic et al. (2010), and no targeted search for evidence was conducted for this report.
- No assessment of the safety of US in high-risk pregnancy.

Table 1. Summary of Practice Guidelines

Key: TVU, transvaginal ultrasound; US, ultrasound

Guideline	Recommendations	Quality
American Institute of Ultrasound Medicine (AIUM) A Practice Guideline for the Performance of Obstetric Ultrasound Examinations (2007)	Fetal US performed in the first, second, and third trimester of pregnancy can be safe and beneficial to diagnose, evaluate, or confirm a number of clinical indications related to fetal and maternal health.	Poor
American College of Obstetricians and Gynecologists (ACOG) Practice Bulletin 101 for Ultrasonography in Pregnancy (2009)	• US examination is a safe and accurate method of determining precise gestational age, number of fetuses, viability, and location of the placenta, as well as diagnosing a number of major fetal anomalies (level A for good and consistent evidence). • There is some support for the use of US to assist in the detection of fetal growth disturbances and abnormalities in amniotic fluid volume (level B for limited or inconsistent evidence). • Optimal timing of a single US examination in the first trimester of pregnancy is between 18 weeks and 20 weeks of gestation (level C for expert opinion or consensus).	Fair
American College of Radiology (ACR) Appropriateness Criteria for Growth Disturbances and the Risk of Intrauterine Growth Restriction (IUGR) (2008)	US of a pregnant uterus is safe and usually appropriate for determining fetal measurement, growth, amniotic fluid, fetal anatomic survey, and activity patterns (rating 9). (Rating scale is 1 to 9, in order of increasing confidence in appropriateness.)	Fair
Appropriateness Criteria for Multiple Gestations (2008)	• Transabdominal or transvaginal US is safe and appropriate for patients with a high or low index of suspicion for multiple gestations, or in patients who have already been diagnosed with twins (rating 9). • Evidence does not support the use of transabdominal or TVU with umbilical artery Doppler as a method of assessment for twins (rating 4).	Fair
Appropriateness Criteria for First Trimester Bleeding (2009)	• Transabdominal or transvaginal pelvic US is generally safe and appropriate in patients with a positive urine or serum pregnancy test, when correlated with other testing, and in patients who were diagnosed with twins on an initial US (rating 9). • Pelvic US with Doppler imaging is generally safe and appropriate; however, pulsed Doppler of the embryo is not recommended (rating 4)	Fair
Appropriateness Criteria for Second and Third Trimester Bleeding (2009)	• Transabdominal pelvic US is safe and appropriate for a variety of conditions (rating 9). • TVU is an alternative or can be used as follow-up (rating 8 or 9).	Fair
Institute for Clinical Systems Improvement (ICSI) Routine Prenatal Care (2010)	• The existing evidence does not support routine US examinations in low-risk pregnancies. • US can be used for gestational dating and anatomy evaluations, and for assessing possible genetic abnormalities. US may be useful to confirm a questionable fetal position/presentation. • Early sonography may confirm dating when gestational age is uncertain or there are antecedent medical complications,	Fair

Guideline	Recommendations	Quality
	including pregestational diabetes mellitus, or previous complications (high-risk pregnancies). • Serial US and antepartum fetal testing is reasonable for the management of hemoglobinopathies during pregnancy. • Three-dimensional (3D) and four-dimensional (4D) US is not recommended for routine use during pregnancy.	

Table 2. Summary of Key Findings: Effectiveness of US in Pregnancy

Key: C-section, Cesarean section; CTG, cardiotocograph; DUS, Doppler ultrasound; hx, history; IOL, induction of labor; LOS, length of stay; NS, (statistically) nonsignificant; NNT, number needed to treat; PTB, preterm birth; PTL, preterm labor; RCT, randomized controlled trial; TVU, transvaginal ultrasound; US, ultrasound

Outcome Measure	Comparison	Type/Quantity of Evidence	Findings	Quality[1]
High-Risk Pregnancies (see Appendices IIIA, IIIB, and IIIC for details from meta-analyses by Alfirevic et al. (2010) and Berghella et al. (2009), and systematic review by Blikman et al. (2008); additional evidence provided by an RCT of TVU surveillance [Simcox et al., 2009])				
Antenatal admissions	DUS vs no-DUS or DUS vs CTG alone in high risk pregnancy	2 RCTS (n=893); 1 RCT (n=426)	Moderate reduction	Moderate
Antenatal admissions	TVU vs no TVU; women w/ suspected PTL	1 RCT (n=93)	NS difference	Low
Antenatal admissions/LOS	TVU surveillance vs hx-based management in women w/ previous PTB	1 RCT (n=253)	No impact on antenatal admissions but significant increase in mean LOS by 2 days	Low
Tocolysis; steroids	TVU vs no TVU; women w/ suspected PTL	2 RCTs (n=102; n=114)	NS difference	Low
Inappropriate use of tocolysis and steroids	TVU vs no TVU; women w/ suspected PTL	1 RCT (n=41)	Significant, substantial reduction (from 90% to 33%)	Low
Use of tocolysis, steroids, progesterone	TVU surveillance vs hx-based management in women w/ previous PTB	1 RCT (n=253)	Substantial increases; significant only for progesterone, from 20% to 32%	Low
Cerclage	TVU surveillance vs hx-based management in women w/ previous PTB	2 RCTs, 4 observational studies (n=77 to n=177) 1 good RCT (n=253)	Reduced Significantly increased from 20% to 32%	Low
Perinatal mortality	DUS vs no-DUS in high-risk pregnancy	16 RCTs/quasi-randomized trials (n=10,125)	Significant, modest absolute reduction; NNT=200	Very low
Serious neonatal morbidity	DUS vs no-DUS in high-risk pregnancy	3 RCTs (n=598)	Significant, modest absolute reduction; NNT=36	Very low
PTB	TVU vs no TVU; women w/ suspected PTL	3 RCTs (n=256)	NS difference according to pooled estimate (But a substantial reduction was demonstrated by the best of the 3 studies.)	Low
PTB	TVU surveillance vs no TVU in women w/ twin pregnancy	1 RCT (n=125)	NS	Low
PTB	TVU surveillance vs hx-based management in women w/ previous PTB	3 RCTs (n=423)	No impact (results contradicted by one prospective cohort study)	Moderate

Pregnancy loss <24 wks	TVU surveillance vs hx-based management in women w/ previous PTB	2 RCTs (n=350)	No impact results contradicted by one prospective cohort study)	Moderate
C-section	DUS vs no-DUS in high risk pregnancy	14 RCTs/quasi-randomized trials (n=7918)	Significant, modest absolute reduction; NNT=39 (NNT=47 for emergency C-section)	Low
C-section, elective	DUS vs CTG alone	3 RCTs (n=1473)	>50% relative increase	High
C-section, emergency	DUS vs CTG alone	3 RCTs (n=1473)	44% relative, 45% absolute reduction; NNT=12	High
IOL	DUS vs no-DUS in high risk pregnancy	1 RCTs/quasi-randomized trials (n=5633)	Significant, modest absolute reduction; NNT=31	Low
IOL	DUS vs CTG alone	2 RCTs (n=576)	NS difference	Low
Rate of abortion	No data			
Early (< 24 wks) Low-Risk Pregnancy (see Appendix IVA for detail from meta-analysis by Whitworth et al. [2010])				
Antenatal visits/hospitalization	Routine US vs concealed/selective US	2 RCTS(n=9502)/ 5 RCTs (n=17,685)	No impact	High
Inappropriately-time serum scan/repeat fetal anomaly US scan	Routine US vs concealed/selective US	1 RCT (n=602)	NS difference	Low
Perinatal mortality/morbidity	Routine US vs concealed/selective US	4 to 8 RCTs (n=3906 to n=19,337	No impact	High
C-section	Routine US vs concealed/selective US	5 RCTs (n=22,193)	No impact	Moderate
IOL for postterm pregnancy	Routine US vs concealed/selective US	8 RCTs (n=25,516)	Significant, modest absolute reduction; NNT=100	Moderate
IOL for any reason	Routine US vs concealed/selective US	7 RCTs (n=24,790)	Significant, modest absolute reduction; NNT=100	Moderate
Rate of abortion for fetal anomaly	Routine US vs concealed/selective US	5 RCTs (n=28,256)	More than doubled; absolute increase of 0.10 percentage points	High
Late (>24 wks) Low-Risk Pregnancy (see Appendix IVB for detail from meta-analysis by Bricker et al. [2008])				
Antenatal admissions	Routine US vs no/concealed/selective US	4 to RCTs (n=5396)	NS difference	Moderate
Frequency of follow-up US	Routine US vs no/concealed/selective US	2 RCTs (n=2536)	NS difference	Moderate
Use of cardiographs	Routine US vs no/concealed/selective US; serial US+DUS vs routine US	1 RCT (n=2000); 1 RCT (n=2834)	No impact	Low

Perinatal mortality/stillbirth/perinatal morbidity	Routine US vs no/concealed/selective US	5 to 7 RCTs (n=21,708 to n=24,276)	No impact	Moderate
Stillbirth, excluding congenital abnormality	Routine US vs no/concealed/selective US	2 RCT (n=2902) Results may be due to placental grading in more influential study.	Significant, modest absolute reduction; NNT=100	Low
Perinatal mortality/morbidity	Serial US+DUS vs routine US	1 RCT (n=2834)	No impact, except possible adverse effect on birth weight	Low
C-section	Routine US vs no/concealed/selective US	5 RCTs (n=21,035)	No impact	Moderate
IOL	Routine US vs no/concealed/selective US	6 RCTs (n=22,663)	No impact	Moderate
Rate of abortion	No data			

[1] Refers to quality of the overall body of evidence for the corresponding outcome. Quality was judged according to a system adapted from the GRADE system. Individual study quality is one of several factors considered. See the summary sections in the **Literature Review** for specific reasons for quality ratings.

Table 3. Summary of Key Findings: Safety of US in Pregnancy

Sources: Meta-analysis by Torloni et al. (2009) (see Appendix V); meta-analysis by Whitworth et al. (2010) (see Appendix IVA). Additional data from subsequently published RCTs: Rodriguez and Waldenström (2008), McLaughlin et al. (2009), Simcox et al. (2009), Stålberg et al. (2008) , and Grether et al. (2010).

Key: RCT, randomized controlled trial; US, ultrasound

Outcome Measure	Comparison	Type/Quantity of Evidence	Findings	Quality[1]
Antenatal admissions	Routine US vs no US or fewer scans	9 RCTs (n=25,200)	No impact	Moderate
Postpartum complications	Routine US vs no US or fewer scans	1 cohort study (n=806)	No impact	Low
Fetal mortality	Routine US vs no US or fewer scans	8 RCTs (n=32,962)	No impact	Moderate
Perinatal mortality	Routine US vs no US or fewer scans	13 RCTs (n=46,553)	No impact	Moderate
Neonatal mortality	Routine US vs no US or fewer scans	7 RCTs (n=30,942)	No impact	Low
Mean birth weight	Routine US vs no US or fewer scans	9 RCTs (n=35,894)	NS difference[2]	Low
Apgar <7 at 5 min	Routine US vs no US or fewer scans	12 RCTs (n=22,150)	NS difference	Low
Perinatal /general morbidity	Routine US vs no US or fewer scans	9 to 13 RCTs (n=22,150 to n=44,745)	No impact	Moderate
Congenital malformations	Routine US vs no US or fewer scans	2 RCTs (n=15,281)	No impact	Moderate
Congenital cardiac malformations	Routine US vs no US or fewer scans	2 cohort studies (n=12,451)	Almost doubles the risk	Low
Childhood cancer	Routine US vs no US or fewer scans	10 case-control studies (n=27,306)	No impact	Moderate
Non–right-handedness	Routine US vs no US or fewer scans; routine US vs no/concealed/selective US	3 RCTs+2 prospective cohort studies	No overall impact; increase in boys	Moderate
Other measures of child growth and development, including cognitive development	Routine US vs no US or fewer scans; routine US vs no/concealed/selective US	Most data from 1 to 2 RCTs (n=1993 to n=4715)	No impact	Low
Intellectual performance in men	Routine US vs no US or fewer scans	1 cohort study (n=16,059)	Significant, small adverse effect	Very low
Mental illness	Routine US vs no US or fewer	1 cohort study (n=370,945)	No impact	Very low

Outcome Measure	Comparison	Type/Quantity of Evidence	Findings	Quality[1]
	scans			
Infection	TVU assessment of cervical length	2 RCTs (n=345)	No impact	Low

[1] Refers to quality of the overall body of evidence for the corresponding outcome. Quality was judged according to a system adapted from the GRADE system. Individual study quality is one of several factors considered. See the summary sections in the **Literature Review** for specific reasons for quality ratings.
[2] Although moderate-quality evidence shows a modest reduction in birth size associated with ≥3 vs 1 scan.

Table 4. Summary of Key Findings: Differential Effectiveness and Safety of Ultrasound in Pregnancy

Sources: Meta-analyses (see Appendices III-VI) and individual trials (Ewigman et al., 1993; Stålberg et al., 2008; Grether et al., 2010)

Key: C-section, Cesarean section; ED, emergency department; IOL, induction of labor; RCT, randomized controlled trial; US, ultrasound

Comparison or Subgroup	Outcome Measure	Type/Quantity of Evidence	Findings	Quality[1]
2nd vs 1st trimester	Detection of multiple pregnancy by 24 wks	7 RCTs (n=295)	Effective only in 2nd trimester	Low
2nd vs 1st trimester	Perinatal mortality	9 RCTs (n=34,923)	No differential effect	High
2nd vs 1st trimester	IOL	8 RCTs (n=25,516)	Effective only in 2nd trimester	Moderate
Late pregnancy (>24 wks) vs early	IOL	13 RCTs (n=47,453)	Effective only in early pregnancy	Moderate-High
3rd vs 2nd vs 1st trimester	Childhood brain cancer	1 case-control study (n=1051)	No difference in effect	Very low
3rd vs 2nd vs 1st trimester	Autism	1 case-control study (n=755)	No difference in effect	Very low
Boys vs girls	Non–right-handedness	2 RCTs (n=2422), 2 cohort studies (n=40,841)	Increased in boys only	Moderate
Singleton vs multiple birth; DUS vs no-DUS in late pregnancy	Perinatal mortality, serious perinatal morbidity	18 RCTs; n=10,156	No difference in effect	Low
SGA fetus vs HTN/preeclampsia vs diabetes vs prolonged pregnancy vs previous pregnancy loss	Perinatal mortality, serious perinatal morbidity	18 RCTs; n=10,156	No difference in effect	Low
Routine serial US+DUS vs routine US in late pregnancy	Perinatal outcomes	2 small RCTs	No improvement in any outcomes; possible small adverse effect on birth size	Very low
US in ED vs by radiological or gynecological specialists	Outcome of evaluation and tx of suspected ectopic pregnancy	8 studies of mixed design; n>1778	More efficient; improved surgical outcomes	Very low
Twice weekly vs every-other-wk surveillance for impaired fetal	C-section growth	1 RCT (n=167)	No difference	Low
Twice weekly vs every-other-wk surveillance for impaired fetal growth	IOL	1 RCT (n=167)	25% relative increase	Low

[1] Refers to quality of the overall body of evidence for the corresponding outcome. Quality was judged according to a system adapted from the GRADE system. Individual study quality is one of several factors considered. See the summary sections in the **Literature Review** for specific reasons for quality ratings.

Table 5. Summary of Key Findings: Cost Implications and Cost-effectiveness of US in Pregnancy

Key: EDTU, emergency department targeted ultrasound; GA, gestational age; HMO, health maintenance organization; hx, history; ICER, incremental cost-effectiveness ratio; NS, (statistically) nonsignificant; NT, nuchal translucency; pt, patient; PTB, preterm birth; QALY, quality-adjusted life-year; RADIUS, Routine Antenatal Diagnostic Imaging with Ultrasound; TVU, transvaginal ultrasound; US, ultrasound

Outcome Measure	Comparison	Type/Quantity of Evidence	Findings	Comments
Short- and long-term direct costs associated w/ US, including cost of implementing organized public program and lifetime care for individuals born w/ severe neonatal morbidity	Organized program of universal 2nd trimester US screening for detecting structural abnormality vs usual practice	1 cost-effectiveness study (decision analytic modeling); perspective of public payer (Italy) (Vanara et al., 2004)	Short-term costs would decrease by 44% Long-term costs would decrease by 21%	Evidence reviewed for this report did not provide a basis of comparison for the study's assumed detection rate of structural abnormality. Results may not be applicable to United States healthcare system.
All direct short-term costs associated w/ providing US, companion tests, and follow-up tests/procedures	6 strategies for US screening for fetal anomaly (all strategies included some type of routine US screening)	1 cost-effectiveness study (modeling); perspective of public payer (Scotland) (Ritchie et al., 2005)	Least expensive strategy: 1st trimester US for estimating GA, 2nd trimester double serum test Compared w/ least expensive strategy, the most cost-effective alternatives entailed 1st trimester US for NT and with no further screening or 2nd trimester anomaly US.	There appear to be errors in some of the reported values for ICERs. A strategy of *no* US was not considered. No sensitivity analysis.
Short-term and long-term direct and indirect (productivity) costs	Routine 2nd trimester US anomaly screen vs no/selective US	1 cost-effectiveness study (modeling); societal perspective (United States) (Vintzileos et al., 2000)	Short-term savings of $13,030/pt if initial screen in tertiary center; $2230/pt if in nontertiary center. Long-term savings $97 to $189/pt if initial screening in tertiary center. Long-term loss of $69 to $161/pt if initial screening in nontertiary center.	Accuracy based on RADIUS trial. Estimate of termination rate may not be applicable to mild/moderate anomalies. Costs were in 1998 dollars.
Long-term direct costs, including lifetime care for individuals born w/ severe neonatal morbidity	Universal TVU to assess cervical length as an *add-on* to routine 2nd trimester US anomaly scan.	1 cost-effectiveness study (modeling); Medicaid perspective (Cahill et al., 2010)	Saved costs, prevented PTB, and produced QALY gains, but only in women w/out hx of PTB.	Assumes availability of TVU in all facilities. Results not applicable if 2nd trimester US screening is not already considered routine. Results may not apply to multiple gestation, pregnancy with fetal abnormality, treatment of short cervix defined at thresholds >15

Outcome Measure	Comparison	Type/Quantity of Evidence	Findings	Comments
				mm, prevention of preterm births at other than < 34 weeks, prevention of preterm birth with treatments other than progesterone, assessment of cervical length in women with symptoms of preterm labor.
Short-term cost of US in emergency evaluation of possible ectopic pregnancy	EDTU vs US w/ involvement of obstetrics or radiology department	1 cost analysis (retrospective chart review); perspective of staff model HMO (United States) (Durston et al., 2000)	EDTU could save $299 to $1244 per case of ectopic pregnancy	Costs from 1992-1998
Hospital charges for US evaluation of 1st trimester bleeding	EDTU vs US w/ involvement of obstetrics or radiology department	1 cost analysis (unpublished RCT); perspective of payer (Pierce et al., 2001)	$535/pt for EDTU; $926/pt for radiology/gynecologist; NS difference	

REFERENCES

Alfirevic Z, Allen-Coward H, Molina F, Vinuesa CP, Nicolaides K. Targeted therapy for threatened preterm labor based on sonographic measurement of the cervical length: a randomized controlled trial. *Ultrasound Obstet Gynecol*. 2007;29(1):47-50.

Alfirevic Z, Stampalija T, Gyte GM. Fetal and umbilical Doppler ultrasound in normal pregnancy. *Cochrane Database Syst Rev*. 2010;8:CD001450.

Althuisius SM, Dekker GA, van Geijn HP, Bekedam DJ, Hummel P. Cervical incompetence prevention randomized cerclage trial (CIPRACT): study design and preliminary results. *Am J Obstet Gynecol*. 2000;183(4):823-829.

American College of Obstetricians and Gynecologists (ACOG). ACOG Practice Bulletin No. 58. Ultrasonography in pregnancy. *Obstet Gynecol*. 2004;104(6):1449-1458.

American College of Obstetricians and Gynecologists (ACOG). ACOG Practice Bulletin No. 101: Ultrasonography in pregnancy. *Obstet Gynecol*. 2009;113(2 Pt 1):451-461.

American College of Radiology (ACR). ACR Appropriateness Criteria®. Growth Disturbances-Risk of Intrauterine Growth Restriction. 2008a. Available at: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonWomensImaging/growthDisturbancesRiskofintrauterinegrowthRestrictionDoc4.aspx. Accessed August 12, 2010.

American College of Radiology (ACR). ACR Appropriateness Criteria®. Multiple Gestations. 2008b. Available at: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonWomensImaging/MultiplegestationsDoc6.aspx. Accessed August 12, 2010

American College of Radiology (ACR). ACR Appropriateness Criteria®. Second and Third Trimester Bleeding. 2009a. Available at: http://www.acr.org/secondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonWomensimaging/secondandThirdtrimesterbleedingDoc9.aspx. Accessed August 12, 2010.

American College of Radiology (ACR). ACR Appropriateness Criteria®. First Trimester Bleeding. 2009b. Available at http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonWomensImaging/firsttrimesterBleedingDoc3.aspx. Accessed August 12, 2010

American College of Radiology (ACR). ACR Appropriateness Criteria® Overview. 2010. Available at: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/overview.Aspx. Accessed August 12, 2010.

American Institute for Medical Ultrasound (AIUM). AIUM Practice Guideline for the Performance of Obstetric Ultrasound Examinations. 2007. Available at: http://www.aium.org/publications/guidelines/obstetric.pdf. Accessed August 12, 2010.

Bahado-Singh RO, Cheng CS. First trimester prenatal diagnosis. *Curr Opin Obstet Gynecol.* 2004;16(2):177-181.

Beigi A, Zarrinkoub F. Elective versus ultrasound-indicated cervical cerclage in women at risk for cervical incompetence. *Med J Islamic Rep Iran.* 2005;19(2):103-107.

Belfrage P, Fernström I, Hallenberg G. Routine or selective ultrasound examinations in early pregnancy. *Obstet Gynecol*. 1987;69(5):747-750.

Bennett MJ, Little G, Dewhurst J, Chamberlain G. Predictive value of ultrasound measurement in early pregnancy: a randomized controlled trial. *Br J Obstet Gynaecol*. 1982;89(5):338-341.

Berghella V, Baxter JK, Hendrix NW. Cervical assessment by ultrasound for preventing preterm delivery. *Cochrane Database Syst Rev*. 2009;(3):CD007235.

Blikman MJC, Le T, Bruinse HW, van der Heijden GJMG. Ultrasound-predicated versus history-predicated cerclage in women at risk of cervical insufficiency: a systematic review. *Obstet Gynecol Surv*. 2008;63(12):803-812.

Bricker L, Neilson JP, Dowswell T. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database Syst Rev.* 2008;(4):CD001451.

Cahill AG, Odibo AO, Caughey AB, et al. Universal cervical length screening and treatment with vaginal progesterone to prevent preterm birth: a decision and economic analysis. *Am J Obstet Gynecol*. 2010;202(6):548.e1-8.

Cargill Y, Morin L, Bly S, et al. Content of a complete routine second trimester obstetrical ultrasound examination and report. *J Obstet Gynaecol Can.* 2009;31(3):272-275, 276-280.

Carlan SJ, Richmond LB, O'Brien WF. Randomized trial of endovaginal ultrasound in preterm premature rupture of membranes. *Obstet Gynecol*. 1997;89(3):458-461.

Caughey AB, Kuppermann M, Norton ME, Washington AE. Nuchal translucency and first trimester biochemical markers for down syndrome screening: a cost-effectiveness analysis. *Am J Obstet Gynecol*. 2002;187(5):1239-1245.

Center for Devices and Radiological Health (CDRH). A Close Look at Ultrasound. Updated August 18, 2010. Food and Drug Administration [website]. Available at: http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm181011.htm. Accessed August 18, 2010.

Centers for Medicare & Medicaid (CMS). Medicare Coverage Database. National Coverage Documents [search: *ultrasound*]. Updated January 14, 2010. Available at: http://www.cms.gov/mcd/search.asp. Accessed August 12, 2010.

Cipriano L, Barth W, Zaric G. The cost-effectiveness of targeted or universal screening for vasa praevia at 18-20 weeks of gestation in Ontario. *BJOG*. 2010;117(9):1108-1118.

Cnossen JS, Morris RK, ter Riet G, et al. Use of uterine artery Doppler ultrasonography to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariable meta-analysis. *CMAJ*. 2008;178(6):701-711.

Conde-Agudelo A, Romero R, Hassan SS, Yeo L. Transvaginal sonographic cervical length for the prediction of spontaneous preterm birth in twin pregnancies: a systematic review and metaanalysis. *Am J Obstet Gynecol*. 2010;203(2):128.e1-12.

Coomarasamy A, Connock M, Thornton J, Khan KS. Accuracy of ultrasound biometry in the prediction of macrosomia: a systematic quantitative review. *BJOG*. 2005;112(11):1461-1466.

Crowther CA, Kornman L, O'Callaghan S, George K, Furness M, Willson K. Is an ultrasound assessment of gestational age at the first antenatal visit of value? A randomised clinical trial. *Br J Obstet Gynaecol*. 1999;106(12):1273-1279.

Cusick W, Buchanan P, Hallahan TW, et al. Combined first-trimester versus second-trimester serum screening for Down syndrome: a cost analysis. *Am J Obstet Gynecol*. 2003;188(3):745-751.

Durston WE, Carl ML, Guerra W, Eaton A, Ackerson LM. Ultrasound availability in the evaluation of ectopic pregnancy in the ED: comparison of quality and cost-effectiveness with different approaches. *Am J Emerg Med*. 2000;18(4):408-417.

Eik-Nes SH, Okland O, Aure JC, Ulstein M. Ultrasound screening in pregnancy: a randomised controlled trial. *Lancet*. 1984;1(8390):1347.

Falo AP. Intrauterine growth retardation (IUGR): prenatal diagnosis by imaging. *Pediatr Endocrinol Rev*. 2009;6(Suppl 3):326-331.

Flood K, Malone FD. Screening for fetal abnormalities with ultrasound. *Curr Opin Obstet Gynecol*. 2008;20(2):139-145.

Fu S, Bao Y, Zhang J, Hu Y, Wang X, Zhang F. Effects of B-ultrasonic radiation during pregnancy on reproductive outcome [Chinese]. *Chung-Hua Yu Fang i Hsueh Tsa Chih (Chin J Prevent Med)*. 2000;34(2):86-88.

Gagnon R, Morin L, Bly S, et al. Guidelines for the management of vasa previa. *J Obstet Gynaecol Can*. 2009;31(8):748-760.

Gekas J, Gagne G, Bujold E, et al. Comparison of different strategies in prenatal screening for Down's syndrome: cost effectiveness analysis of computer simulation. *BMJ*. 2009;338:b138.

Giguère Y, Charland M, Bujold E, et al. Combining biochemical and ultrasonographic markers in predicting preeclampsia: a systematic review. *Clin Chem*. 2010;56(3):361-375.

Gordon M, Robbins A, McKenna D, Howard B, Barth W. Cervical length assessment as a resource to identify twins at risk for preterm delivery (clarity study). *Am J Obstet Gynecol.* 2006;195(6 Suppl 1):S55.

Grether JK, Li SX, Yoshida CK, Croen LA. Antenatal ultrasound and risk of autism spectrum disorders. *J Autism Dev Disord*. 2010;40(2):238-245.

Grivell R, Wong L, Bhatia V. Regimens of fetal surveillance for impaired fetal growth. *Cochrane Database Syst Rev*. 2009;(1):CD007113.

Harris AH. The cost effectiveness of prenatal ultrasound screening for trisomy 21. *Int J Technol Assess Health Care*. 2004;20(4):464-468.

Hayes, Inc. Hayes Medical Technology Directory. Three-Dimensional and Four-Dimensional Ultrasound for Diagnosis of Fetal Head Abnormalities. Lansdale, PA: Hayes, Inc.; November 24, 2005a.

Hayes, Inc. Hayes Medical Technology Directory. Three-Dimensional and Four-Dimensional Ultrasound for Diagnosis of Fetal Limbs and Skeletal Structures. Lansdale, PA: Hayes, Inc.; December 5, 2005b.

Hayes, Inc. Hayes Medical Technology Directory. Three-Dimensional and Four-Dimensional Ultrasound for Fetal Growth and Volume Measurements. Lansdale, PA: Hayes, Inc.; December 20, 2005c.

Hayes, Inc. Hayes Medical Technology Directory. Three-Dimensional and Four-Dimensional Ultrasound for High-Risk Pregnancies and Routine Screening. Lansdale, PA: Hayes, Inc.; November 9, 2005d.

Hayes, Inc. Hayes Medical Technology Directory. Three-Dimensional and Four-Dimensional Ultrasound for Extrafetal and Maternal Structures in Pregnancy. Lansdale, PA: Hayes, Inc.; July 18, 2006a.

Hayes, Inc. Hayes Medical Technology Directory. Three-Dimensional and Four-Dimensional Ultrasound for Fetal Cardiovascular Diagnosis. Lansdale, PA: Hayes, Inc.; February 9, 2006b.

Hellman LM, Duffus GM, Donald I, Sunden B. Safety of diagnostic ultrasound in obstetrics. *Lancet*. 1970;1(7657):1133-1134.

Higgins SP, Kornman LH, Bell RJ, Brennecke SP. Cervical surveillance as an alternative to elective cervical cerclage for pregnancy management of suspected cervical incompetence. *Aust N Z J Obstet Gynaecol*. 2004;44(3):228-232.

Houston LE, Odibo AO, Macones GA. The safety of obstetrical ultrasound: a review. *Prenat Diagn*. 2009;29(13):1204-1212.

Institute for Clinical Systems Improvement (ICSI). Routine Prenatal Care. 2010. Available at: http://www.icsi.org/prenatal_care_4/prenatal_care__routine__full_version__2.html. Accessed August 12, 2010

Kieler H, Haglund B, Cnattingius S, Palmgren J, Axelsson O. Does prenatal sonography affect intellectual performance*? Epidemiology.* 2005;16(3):304-310.

Kurjak A, Vecek N, Hafner T, et al. Prenatal diagnosis: what does four-dimensional ultrasound add? *J Perinat Med.* 2002;30(1):57-62.

Le Ray C, Morin L. Routine versus indicated third trimester ultrasound: is a randomized trial feasible? *J Obstet Gynaecol Can*. 2009;31(2):113-119.

Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol*. 2007;21(Suppl 2):86-96.

Martin JA, Hamilton BE, Sutton PD, et al. *Births: Final Data for 2006*. Hyattsville, MD: National Center for Health Statistics; 2009. National Vital Statistics Reports, Vol. 57, No. 7. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_07.pdf. Accessed August 12, 2010.

McLaughlin CC, Baptiste MS, Schymura MJ, Zdeb MS, Nasca PC. Perinatal risk factors for neuroblastoma. *Cancer Causes Control*. 2009;20(3):289-301.

McRae A, Murray H, Edmonds M. Diagnostic accuracy and clinical utility of emergency department targeted ultrasonography in the evaluation of first-trimester pelvic pain and bleeding: a systematic review. *CJEM*. 2009;11(4):355-364.

Neilson JP, Munjanja SP, Whitfield CR. Screening for small for dates fetuses: a controlled trial. *BMJ.* 1984;289(6453):1179-1182.

Ness A, Visintine J, Ricci E, Berghella V. Does knowledge of cervical length and fetal fibronectin affect management of women with threatened preterm labor? A randomized trial. *Am J Obstet Gynecol*. 2007;197(4):426.e1-7.

Newnham JP, Doherty DA, Kendall GE, et al. Effects of repeated prenatal ultrasound examinations on childhood outcome up to 8 years of age: follow-up of a randomised controlled trial. *Lancet.* 2004;364(9450):2038-2044.

Newnham JP, Evans SF, Michael CA, Stanley FJ, Landau LI. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. *Lancet.* 1993;342(8876):887-891.

Odibo AO, Stamilio DM, Nelson DB, Sehdev HM, Macones GA. A cost-effectiveness analysis of prenatal screening strategies for Down syndrome. *Obstet Gynecol*. 2005;106(3):562-568.

Oto A, Ernst R, Jesse MK, et al. Magnetic resonance imaging of the chest, abdomen, and pelvis in the evaluation of pregnant patients with neoplasms. *Am J Perinatol*. 2007;24(4):243-250.

Platz E, Newman R. Diagnosis of IUGR: traditional biometry. *Semin Perinatol.* 2008;32(3):140-147.

Ott WJ. Diagnosis of intrauterine growth restriction: comparison of ultrasound parameters. *Am J Perinatol.* 2002;19(3):133-137.

Palacio M, Sanchez M, Cobo T, Figueras F, et al. Cervical length measurement to reduce length of stay in patients admitted because of preterm labor. Prospective and randomized trial. Final results [poster session]. *Ultrasound Obstet Gynecol.* 2006;28(4):485.

Pathak S, Lees C. Ultrasound structural fetal anomaly screening: an update. *Arch Dis Child Fetal Neonatal Ed*. 2009;94(5):F384-F390.

Pedrosa I, Lafornara M, Pandharipande PV, et al. Pregnant patients suspected of having acute appendicitis: effect of MR imaging on negative laparotomy rate and appendiceal perforation rate. *Radiology*. 2009;250(3):749-757.

Pedrosa I, Zeikus EA, Levine D, Rofsky NM. MR imaging of acute right lower quadrant pain in pregnant and nonpregnant patients. *Radiographics*. 2007;27(3):721-743.

Pierce DL, Friedman KD, Killian A, et al. Emergency department ultrasonography (EUS) in symptomatic first trimester pregnancy [abstract]. *Acad Emerg Med.* 2001;8:546.

Proud J, Grant AM. Third trimester placental grading by ultrasonography as a test of fetal wellbeing. *BMJ.* 1987;294(6588):1641-1644.

Pugash D, Brugger PC, Bettelheim D, Prayer D. Prenatal ultrasound and fetal MRI: The comparative value of each modality in prenatal diagnosis. *Eur J Radiol.* 2008;68(2):214-226.

Ritchie K, Bradbury I, Slattery J, et al. Economic modelling of antenatal screening and ultrasound scanning programmes for identification of fetal abnormalities. *BJOG*. 2005;112(7):866-874.

Rodriguez A, Waldenström U. Fetal origins of child non-right-handedness and mental health. *J Child Psychol Psychiatry*. 2008;49(9):967-976.

Royal College of Obstetricians and Gynaecologists (RCOG). Obstetric Cholestatis. 2006a. Available at: http://www.rcog.org.uk/files/rcog-corp/uploaded-files/GT43ObstetricCholestasis2006.pdf. Accessed August 12, 2010.

Royal College of Obstetricians and Gynaecologists (RCOG). The Management of Severe Pre-Eclampsia/Eclampsia. 2006b. Available at: http://www.rcog.org.uk/files/rcog-corp/uploaded-files/GT10aManagementPreeclampsia2006.pdf. Accessed August 12, 2010

Salvesen KA, Bakketeig LS, Eik-nes SH, Undheim JO, Okland O. Routine ultrasonography in utero and school performance at age 8-9 years. *Lancet.* 1992;339(8785):85-89.

Salvesen KA, Jacobsen G, Vatten LJ, Eik-Nes SH, Bakketeig LS. Routine ultrasonography in utero and subsequent growth during childhood. *Ultrasound Obstet Gynecol.* 1993;3(1):6-10.

Salvesen KA, Vatten LJ, Bakketeig LS, Eik-Nes SH. Routine ultrasonography in utero and speech development. *Ultrasound Obstet Gynecol.* 1994;4(2):101-103.

Shohat M, Frimer H, Shohat-Levy V, et al. Prenatal diagnosis of Down syndrome: ten year experience in the Israeli population. *Am J Med Genet A*. 2003;122A(3):215-222.

Siddique J, Lauderdale DS, VanderWeele TJ, Lantos JD. Trends in prenatal ultrasound use in the United States: 1995 to 2006. *Med Care*. 2009;47(11):1129-1135.

Simcox R, Seed PT, Bennett P, et al. A randomized controlled trial of cervical scanning vs history to determine cerclage in women at high risk of preterm birth (CIRCLE trial). *Am J Obstet Gynecol*. 2009;200(6):623.e1-e6.

Sotiriadis A, Papatheodorou S, Kavvadias A, Makrydimas G. Transvaginal cervical length measurement for prediction of preterm birth in women with threatened preterm labor: a meta-analysis. *Ultrasound Obstet Gynecol*. 2010;35(1):54-64.

Stålberg K, Haglund B, Axelsson O, et al. Prenatal ultrasound and the risk of childhood brain tumour and its subtypes. *Br J Cancer*. 2008;98(7):1285-1287.

Stampalija T, Gyte GM, Alfirevic Z. Utero-placental Doppler ultrasound for improving pregnancy outcome. *Cochrane Database Syst Rev*. 2010;9:CD008363.

Sylvan K, Ryding EL, Rydhstroem H. Routine ultrasound screening in the third trimester: a population-based study. *Acta Obstet Gynecol Scand*. 2005;84(12):1154-1158.

Torloni M, Vedmedovska N, Merialdi M, et al. Safety of ultrasonography in pregnancy: WHO systematic review of the literature and meta-analysis. *Ultrasound Obstet Gynecol*. 2009;33(5):599-608.

Vanara F, Bergeretti F, Gaglioti P, Todros T. Economic evaluation of ultrasound screening options for structural fetal malformations. *Ultrasound Obstet Gynecol*. 2004;24(6):633-639.

Vintzileos AM, Ananth CV, Smulian JC, Beazoglou T, Knuppel RA. Routine second-trimester ultrasonography in the United States: a cost-benefit analysis. *Am J Obstet Gynecol*. 2000;182(3):655-660.

Whitworth M, Bricker L, Neilson JP, Dowswell T. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database Syst Rev.* 2010;4:CD007058.

Zhong Y, Tuuli M, Odibo AO. First-trimester assessment of placenta function and the prediction of preeclampsia and intrauterine growth restriction. *Prenat Diagn*. 2010;30(4):293-308.

APPENDIX I. GUIDELINES

Seven American guidelines considered to be most pertinent to the topic of this report were selected. Quality was assessed using the Appraisal of Guidelines Research & Evaluation (AGREE) Instrument.

American Institute of Ultrasound Medicine (AIUM) (pp. 2-5): In 2007, the Practice Guideline for the Performance of Obstetric Ultrasound Examinations was issued by AIUM in collaboration with the American College of Radiology (ACR) and the American College of Obstetricians and Gynecologists (ACOG).

Quality: The quality was determined to be poor due to lack of a systematic literature search, lack of reporting of editorial independence, lack of reporting of the evidence base from which the recommendations were derived, no strength of evidence levels, and finally, no link between the stated recommendations and the evidence base. It is unclear if evidence of clinical utility (impact on clinical decision-making or health outcomes) was considered, in addition to the evidence of accuracy, during the development of this clinical guideline.

Recommendations:

Fetal ultrasound performed in the first trimester of pregnancy can be beneficial for a variety of indications, including, but not limited, to the following:

- Confirm presence of intrauterine pregnancy
- Evaluate suspected ectopic pregnancy
- Define cause of vaginal bleeding
- Evaluate pelvic pain
- Estimate gestational age
- Diagnose or evaluate multiple gestations
- Confirm cardiac activity
- As an adjunct to chorionic villus sampling, embryo transfer, and localization and removal of intrauterine device
- Assess certain fetal anomalies, including anencephaly, in high-risk individuals
- Evaluate maternal pelvic masses and/or uterine abnormalities
- Measure nuchal translucency when part of a screening program for fetal aneuploidy
- Evaluate suspected hydatidiform mole

Fetal ultrasound performed in the second and third trimester of pregnancy can be beneficial for a variety of circumstances, including, but not limited to the following:

- Estimation of gestational age
- Evaluation of fetal growth
- Evaluation of vaginal bleeding
- Evaluation of abdominal or pelvic pain
- Evaluation of cervical insufficiency

- Determination of fetal presentation
- Suspected multiple gestation
- Adjunct to amniocentesis or other procedure
- Significant discrepancy between uterine size and clinical dates
- Pelvic mass
- Suspicion of hydatidiform mole
- Adjunct to cervical cerclage placement
- Suspicion of ectopic pregnancy
- Suspicion of fetal death
- Suspicion of uterine abnormality
- Evaluation of fetal well-being
- Suspicion of amniotic fluid abnormalities
- Suspicion of placental abruption
- Adjunct to external cephalic version
- Premature rupture of membranes and/or premature labor
- Abnormal biochemical markers
- Follow-up of fetal anomaly
- Follow-up of placental location for suspicion of placental previa
- History of previous congenital anomaly
- Evaluation of fetal condition in late prenatal care
- Assess findings that may increase risk of aneuploidy
- Screening for fetal anomalies

Additional recommendations include the following:

- Fetal ultrasound for diagnostic purposes is generally safe for use during pregnancy. However, fetal ultrasound should be performed only when there is a valid medical reason, using the "as low as reasonably achievable" (ALARA) principle, to determine the lowest possible ultrasound exposure settings in order to obtain necessary diagnostic data (p. 8).

NOTE: The AIUM also provided comments and recommendations pertaining to performance of an obstetric ultrasound, including the qualification and responsibilities of personnel; imaging parameters, and measurement specifics; documentation of exam; equipment specificity; quality control and improvement, and infection control. However, these particular recommendations are not relevant to the focus of this technology assessment, and were not reviewed in detail.

The American College of Obstetricians and Gynecologists (ACOG): In 2009, the Practice Bulletin for Ultrasonography in Pregnancy was issued by ACOG.

Quality: The quality was determined to be fair. Evidence was obtained from a systematic literature search, and guidelines from other relevant organization and expert opinions from obstetrician-gynecologists were also reviewed as evidence. However, no clear methodology for development of

recommendations was described. The strength of the guideline was further weakened by the lack of editorial independence, the lack of definitions regarding "good" or "limited" evidence, and the absence of a link between individual studies and each recommendation.

Recommendations:

Recommendations based on "good and consistent evidence" (Level A) (p. 459):

- When used appropriately, ultrasound examination is a safe and accurate method of determining precise gestational age (most accurately determined in the first half of pregnancy), number of fetuses, viability, and location of the placenta.
- Ultrasonography can be used for the diagnosis of a number of major fetal anomalies.

Recommendations based on "limited or inconsistent evidence" (Level B):

- Ultrasonography can assist in the detection of fetal growth disturbances.
- Ultrasonography can detect abnormalities in amniotic fluid volume.

Recommendations based on "consensus and expert opinion" (Level C):

- In the absence of specific indications, optimal timing of a single ultrasound examination in the first trimester of pregnancy is at 18 weeks to 20 weeks of gestation.
- The benefits and limitations of ultrasonography should be discussed in detail with all patients.

In addition to the above summary recommendations, the ACOG also issued the following clinical considerations:

- Ultrasonography may reduce the rate of perinatal mortality, mainly through termination of pregnancy for congenital malformation. In addition, ultrasonography accurately estimates gestational age and detects multiple gestations, and major fetal anomalies; however, how these benefits translate to health outcomes remains unproven. Overall, there is also little evidence that US reduces the rate of perinatal morbidity (p.456).

- The ACOG reported on the various clinical indications appropriate for ultrasonographic examination during the first, second, and third trimester of pregnancy (pp. 452, 454). These indications are the same as reported by the AIUM's Clinical Guideline [see above, American Institute of Ultrasound Medicine (AIUM)].

NOTE: The ACOG also provided comments and recommendations pertaining to imaging parameters for standard fetal examination, ultrasound facility accreditation, and documentation and quality assurance; however, these recommendations are not relevant to the focus of this technology assessment, and hence, were not reviewed in detail.

The American College of Radiologists (ACR): The ACR has issued several guidelines or "Appropriateness Criteria" that address the use of ultrasonography for a variety of clinical conditions related to pregnancy (ACR, 2008a; ACR, 2008c; ACR, 2009a; ACR, 2009b).

Quality of ACR Guidelines: The quality was determined to be fair. A general overview regarding the development of Appropriateness Criteria ("Overview of Appropriateness Criteria") states that a literature search, development of an evidence table, and the quality assessment of the evidence are standard practice (ACR, 2010). However, the search criteria specific to each guideline are not reported an explicit link between the final recommendations and supportive evidence is lacking. Although an evidence table with a strength of evidence rating for each study can be accessed through a link embedded in each guideline, the studies that support each recommendation are not identified and there is not a clear correspondence between the literature review and the recommendations or between the recommendation categories and the strength of evidence ratings. There is no disclosure of potential conflict of interest among panel members.

Recommendation Ratings:

1,2,3: Usually not appropriate
4,5,6: May be appropriate
7,8,9: Usually appropriate

Appropriateness Criteria for Growth Disturbances and the Risk of Intrauterine Growth Restriction (IUGR) (ACR, 2008a): A summary of the recommendations for four clinical scenarios are described as follows (pp. 1-3):

- <u>For risk assessment of IUGR:</u>
  - Ultrasound of a pregnant uterus is safe and "usually appropriate" for determining fetal measurement, growth, amniotic fluid, fetal anatomic survey, and activity patterns (*rating 9*).
  - Ultrasound of a pregnant uterus for a biophysical profile and assessment of umbilical arteries, evidence is indeterminate (*rating 4*).
  - Ultrasound assessment of cerebral to umbilical artery ratio, cerebral arteries, and uterine arteries is not appropriate (*rating 4*).

- <u>For small fetus, low/low-normal fluid, and follow-up studies:</u>
  - Ultrasound of pregnant uterus is indicated at follow-up intervals of 2 weeks to 4 weeks, and more frequently in third trimester or as delivery approaches (*rating 9*).
  - Ultrasound of pregnant uterus for biophysical profile is appropriate (*rating 8*); Doppler ultrasound of pregnant uterus may provide supplemental data but is not a stand-alone test (*rating 8*).

- <u>For very small fetus, normal fluid, and follow-up studies:</u>
  - Ultrasound of pregnant uterus is appropriate for follow-up intervals between 2 weeks and 4 weeks (*rating 9*).

- ✓ Interval of growth assessment is reduced both as estimates of fetal size drop below 10% and as pregnancy advances into third trimester and towards possible urgent delivery.
- Ultrasound of pregnant uterus for biophysical profile is appropriate (*rating 9*).
- Doppler ultrasound of pregnant uterus may provide supplemental data, but should not be used as a stand-alone test (*rating 8*).

- For normal-sized fetus, low/absent fluid, and follow-up studies:
  - Ultrasound of pregnant uterus should have an optimal follow-up interval of 2 weeks (*rating 9*).
  - Ultrasound of pregnant uterus for biophysical profile is appropriate (*rating 9*).
  - Doppler ultrasound of pregnant uterus may provide supplemental data, but should not be used as a stand-alone test (*rating 8*).

- For the four clinical scenarios described above, ACR considers ultrasound to be generally safe for use during pregnancy.

- The Relative Radiation Levels (RRL) of the above indications have been rated as "O" which is indicative of the lowest level of radiation exposure (0 millisieverts [mSv] for both adult and pediatric patients).

Appropriateness Criteria for Multiple Gestations (ACR, 2008b): A summary of the recommendations for four clinical scenarios are described as follows (pp. 1-2):

- For patients with a high index of suspicion for multiple gestations:
  - Evidence supports transabdominal or transvaginal ultrasound for patients with a high and low index of suspicion for multiple gestations (*rating 9*).
  - For pts who were diagnosed w/ twins on initial ultrasound, evidence supports transabdominal or transvaginal ultrasound (same scan) (*rating 9*) to assess the following:
    - ✓ chorionicity and amnionicity
    - ✓ anatomic survey
    - ✓ amniotic fluid
    - ✓ twin sizes/discordance
  - Evidence does not support transabdominal or transvaginal ultrasound with umbilical artery Doppler assessment for each twin (*rating 3*).
  - The RRL of the above examinations have been rated as "O" which is indicative of the lowest level of radiation exposure (0 mSv for both adult and pediatric patients).

Appropriateness Criteria for First Trimester Bleeding (ACR, 2009a): A summary of the recommendations is described as follows (p. 1):

- For patients with a positive urine or serum pregnancy test:

- Evidence supports the use of transabdominal pelvic ultrasound, when correlated with quantitative β-hCG levels (*rating 9*).
- Evidence supports the use of transvaginal pelvic ultrasound, when correlated with quantitative β-hCG levels, and M-mode for fetal heart rate (*rating 9*).
- Evidence supports the use of pelvic ultrasound with Doppler imaging, however, pulsed Doppler of the embryo is not recommended (*rating 7*).
- There is indiscriminate evidence regarding the use of pelvic MRI, with or without contrast (*rating 4*).
- The RRL for the above examinations have been rated as "O" which is indicative of the lowest level of radiation exposure (0 mSv for both adult and pediatric patients).
- Pelvic CT, with or without contrast, is not recommended (*rating 1*).
  - ✓ The RRL for pelvic CT has been rated to be between 1 mSv and 10 mSv for adults and 0.3 mSv to 3 mSv for children.

Appropriateness Criteria for Second and Third Trimester Bleeding (ACR, 2009b): A summary of the recommendations for six clinical conditions is described as follows (pp. 1-2):

- <u>For patients with no other signs or symptoms other than second or third trimester bleeding:</u>
  - Evidence supports the use of transabdominal ultrasound (*rating 9*).
  - Evidence supports the use of transvaginal ultrasound, *if* transabdominal ultrasound is inconclusive (*rating 8*).
    - ✓ Transvaginal ultrasound is contraindicated if there is evidence of ruptured membranes, open cervix with bulging amniotic sac at or below the external os.
  - Evidence supports the use of transperineal ultrasound, if transabdominal ultrasound is inconclusive *and* there is clinical concern about performing transvaginal ultrasound (*rating 8*).

- <u>For patients with internal cervical os not visible by transabdominal ultrasound:</u>
  - Evidence supports the use of transvaginal ultrasound (*rating 9*).
    - ✓ Transvaginal ultrasound should not be used if cervix appears completely open.
  - Evidence supports the use of transperineal ultrasound (*rating 9*).
  - Evidence is indiscriminate with regard to the use of repeat transabdominal ultrasound (*rating 4*).

- <u>For patients with placenta previa diagnosed before 32 weeks:</u>
  - Evidence supports the use of transabdominal ultrasound (*rating 9*).
  - If the cervix and placenta are not visualized with a transabdominal ultrasound, then transvaginal or transperineal ultrasound should be attempted.
  - Evidence supports the use of transvaginal ultrasound if transabdominal ultrasound is inconclusive (rating 9).
  - Evidence supports the use of transperineal ultrasound, if transabdominal ultrasound is inconclusive and if there is concern regarding transvaginal ultrasound (*rating 7*).

- ✓ Transperineal ultrasound is contraindicated if there is evidence of ruptured membranes or open cervix with bulging amniotic sac at or below the external os.

- For patients with uterine contractions, pain, > 20 weeks:
  - Evidence supports the use of transabdominal ultrasound (*rating 9*).
  - Evidence supports the use of transvaginal ultrasound, if transabdominal ultrasound is inconclusive (*rating 8*).
    - ✓ Transvaginal ultrasound is contraindicated if there is evidence of ruptured membranes, or open cervix with bulging amniotic sac at or below the external os.
  - Evidence supports the use of uterus transperineal, if transabdominal ultrasound is inconclusive, and there is concern regarding transvaginal ultrasound (*rating 7*).

- For patients with persistent low lying placenta:
  - Evidence supports the use of transabdominal ultrasound (*rating 9*).
  - Evidence supports the use of transvaginal ultrasound, with Doppler (color and spectral Doppler ultrasound can exclude vasa previa if transabdominal ultrasound is inconclusive (*rating 8*).
  - Evidence supports the use of transperineal ultrasound, if there is concern regarding transvaginal ultrasound (*rating 7*).

- For patients with placenta previa and history of caesarean delivery:
  - Evidence supports the use of transabdominal ultrasound with Doppler (*rating 9*).
  - Evidence supports the use of transvaginal ultrasound with Doppler (color and spectral) (*rating 8*).
  - Evidence supports the use of transperineal ultrasound, if there is concern regarding transvaginal ultrasound (*rating 7*).
  - Evidence supports the use of pelvic MRI without contrast, as an adjunct to ultrasound and for preoperative planning (*rating 7*).

- The RRL for the above examinations have been rated as "O" which is indicative of the lowest level of radiation exposure (0 mSv for both adult and pediatric patients).

The Institute of Clinical Systems Improvement (ICSI): ICSI issued the 2010 guideline, Routine Prenatal Care (ICSI, 2010). Conclusion grades were assigned only to key conclusions and/or recommendations as determined by the ICSI working group members. Study class, based on an evidence rating system, was listed after comments provided by the working group throughout the guideline document (p. 7).

Quality: Fair. A guideline development protocol available on the organizational website defines a process for systematic literature search, but the search for this topic was not described in the guidelines document. Only one recommendation (US for detecting fetal aneuploidy) was accompanied by detailed critical appraisal of the evidence. Given the broad topic for this guideline, these omissions are understandable.

Conclusion Grades:

Grade I: The evidence originates from well-designed studies, and the results are free of significant limitations.
Grade II: The evidence originates from well-designed studies; however, there is some uncertainty regarding the conclusion because of inconsistencies in the results or because of other study limitations.
Grade III: The evidence originates from well-designed studies; however there is substantial uncertainty regarding the conclusion due to inconsistencies in the results or because of serious study limitations.
Grade Not Assignable: There is no evidence available that directly supports or refutes the conclusion.

Evidence Grading System:

A. Primary Reports of New Data Collection:
Class A: Randomized, controlled trial
Class B: Cohort study
Class C: Nonrandomized trial with concurrent or historical controls; case-control study; study of sensitivity and specificity of a diagnostic test; population-based descriptive study
Class D: Cross-sectional study; case series; case report

B. Reports that Synthesize or Reflect upon Collections of Primary Reports:
Class M: Meta-analysis; systematic review; decision analysis; cost-effectiveness analysis
Class R: Consensus statement; consensus report; narrative review
Class X: Medical opinion

Recommendations:

A summary of comments and recommendations pertaining to the use of ultrasound in routine prenatal care was reported as follows:

- For screening purposes, obstetric ultrasound performed between 18 weeks and 20 weeks gestation is considered to be optimal timing for dating and anatomy evaluations, and for assessing possible genetic abnormalities (R), particularly since there is evidence of continued improvement in the identification of congenital anomalies using superior equipment by more experienced examiners (A) (p. 45).

- Early sonography may confirm dating in situations of uncertain age or antecedent medical complications, including pregestational diabetes mellitus or previous complications (M) (p. 45).

- ICSI considers a plan for serial ultrasounds and antepartum fetal testing reasonable for the management of hemoglobinopathies during pregnancy, since these medical conditions put women at risk for spontaneous abortion, preterm labor, intrauterine growth retardation (IUGR), and stillbirth (R) (p. 18).

- When screening for fetal neural tube defect, an ultrasound performed at 18 weeks to 20 weeks gestation may be more sensitive and superior to maternal serum alpha-fetoprotein testing (B) (p. 35).

- Ultrasonography for nuchal translucency testing during the first trimester of pregnancy, between 10 weeks and 13 weeks enhances the detection of Down's syndrome (C) (Conclusion Grade I, p. 35).

- Ultrasound may be useful to confirm a questionable fetal position/presentation (p. 48; *no report or overall conclusion grade provided*).

- The overall existing evidence does not provide support for universal, routine ultrasound in low-risk pregnancies since there is no evidence of improvement in perinatal outcomes (A). More specifically, there is no evidence to support the use of routine ultrasound in low-risk pregnancies beyond 24 weeks gestation (M) (p. 44).

- Three-dimensional (3D) and four-dimensional (4D) ultrasound are not recommended for routine use during pregnancy (R) (p. 45).

EXCLUDED GUIDELINES

These guidelines were identified in searches but were excluded as not pertinent to the report topic or not written for American practitioners.

Sponsor, Year Published	Guideline Topic/Title
American College of Emergency Physicians (ACEP), 2006	Emergency Ultrasound Imaging Criteria Compendium
American College of Obstetricians and Gynecologists (ACOG), 2003	Management of Preterm Labor
ACOG, 2007	Screening for Fetal Chromosomal Abnormalities
ACOG, 2008	Asthma During Pregnancy
ACR, 2008b	Assessment of Gravid Cervix
AIUM, 2008	AIUM Practice Guideline for Ultrasonography in Reproductive Medicine
Finnish Medical Society, 2008	Ultrasound Scanning During pregnancy
ICSI	Management of Labor
Emergency Medicine Quality Council, 2006	Undifferentiated Vaginal Bleeding/Abdominal Pain Suggestive of Ectopic Pregnancy Clinical Pathway
National Heart, Lung, and Blood Institute, 2005	Managing Asthma During Pregnancy: Recommendations for Pharmacologic Treatment
National Collaboration Centre for Women's and Children's Health (NCC-WCH) on behalf of NICE, 2004	Caesarean Section
NICE, 2008a	Diabetes in Pregnancy-Management of Diabetes and Its Complications from Preconception to Postnatal Period
National Collaboration Centre for Women's and Children's Health (NCC-WCH) on behalf of NICE, 2008b	Antenatal Care. Routine Care for the Healthy Pregnant Woman.
Royal College of Obstetricians and Gynaecologists (RCOG), 2005	Placenta Previa and Placenta Previa Accreta: Diagnosis and Management
RCOG, 2006a	Preterm Prelabor Rupture of Membranes
RCOG, 2006b	Obstetric Cholestasis

RCOG, 2006c	The Management of Early Pregnancy Loss
RCOG, 2006d	The Management of Breech Presentation
RCOG, 2006e	The Management of Severe Pre-eclampsia/Eclampsia.
RCOG, 2007	Thromboembolic Disease in Pregnancy and the Puerperium: Acute Management
Society of American Gastrointestinal and Endoscopic Surgeons (SAGES)	Guidelines for the Diagnosis, Treatment and Use of Laparoscopy for Surgical Problems During Pregnancy
SOGC, 2001a	Guidelines for US in Labour and Delivery
SOGC, 2001b	US Cervical Assessment in Predicting Preterm Birth
SOGC, 2003a	Use of 1st Trimester US
SOGC, 2003b	Use of Fetal Doppler in Obstetrics
SOGC, 2005a	Fetal Soft Markers on Obstetric US
SOGC, 2005b	US Evaluation of 1st Trimester Pregnancy Complications
SOGC, 2005c	Obstetric US Biological Effects and Safety
Society of Obstetricians and Gynaecologists of Canada (SOGC), 2007a	Prenatal Screening for Fetal Aneuploidy
SOGC 2007b (Journal of Obstetrics and Gynaecology Canada)	Fetal Health Surveillance: Antepartum and Intrapartum Consensus Guideline
SOGC 2007c	Diagnosis & Management of Placenta Previa
SOGC, 2008	Guidelines for the Management of Pregnancy at 41+0 to 42+0 weeks
SOGC, 2009a	Content of a Complete Routine Second Trimester Obstetrical US Exam and Report
SOGC, 2009b	Guidelines for the Management of Vasa Previa
SOGC, 2009c	Evaluation of Prenatally Diagnosed Structural Congenital Anomalies

APPENDIX II. SEARCH STRATEGY

Basic Search Terms

A variety of search strategies in the MEDLINE and EMBASE databases were used to identify different types of evidence.

The following basic search string was used for MEDLINE searches:

> *("ultrasonography"[Subheading] OR "ultrasonography"[All Fields] OR "ultrasound"[All Fields] OR "ultrasonography"[MeSH Terms] OR "ultrasound"[All Fields] OR "ultrasonics"[MeSH Terms] OR "ultrasonics"[All Fields]) AND ("pregnancy"[MeSH Terms] OR "pregnancy"[All Fields])*

The following keywords were used for EMBASE searches:

> *(ultrasonography or ultrasonogram or ultrasonics or ultrasound) and pregnancy*

All searches were limited to Humans or Human Subjects and English language.

Search for Systematic Reviews

Initially, evidence for this report was obtained by searching for relevant systematic reviews in the following databases: Blue Cross Blue Shield TEC Assessments, Canadian Agency for Drugs and Technology in Health (CADTH), Centre for Reviews and Dissemination (York University), and Cochrane Library. Additional systematic reviews were selected from a search of the MEDLINE database for the time span 2005 through July 2010, using various limits:

- Limited to Practice Types: meta-analysis, practice guideline, consensus development conference, NIH
- Limited to Journal Groups: systematic review
- Limited to “systematic review” in Title/Abstract

(three separate searches, connected by “OR”)

Searches for Primary Studies Published After the Systematic Reviews

For studies of clinical utility (impact on outcomes) of ultrasound (US), MEDLINE and EMBASE were searched for the years 2008 to July 2010. For studies of the safety of US, the time frame was October 2007 to July 2010. The time spans were chosen based on the search time frames in the selected systematic reviews.

Additional Searches for Evidence Excluded by Systematic Reviews

Since the selected review of US in high-risk patients focused on Doppler US, the years 2000 to July 2010 (MEDLINE only) were searched for studies of the clinical utility of non-Doppler US in high-risk patients. The basic search was combined with the PubMed clinical trial filter (broad/sensitive) and with *high risk, hypertension, asthma, diabetes, renal physiopathology, renal physiopathology, renal dysfunction, pyelonephritis,* and *cardiovascular disease* as keywords and subject headings; the search was then

combined using "NOT" with a keyword search using the term *Doppler*. (There were no results other than an investigation of the use of US for assessment of cervical insufficiency.)

Since the review of US in high-risk patients did not mention maternal age as a risk factor, the years 2000 to July 2010 (MEDLINE only) were searched for studies of the clinical utility of US advanced age patients. The basic search was combined with the PubMed clinical trial filter (broad/sensitive) and with the phrase *maternal age and risk* as keyword. (There were no eligible results.)

As a check to make sure studies measuring impact of US on abortion rate had not been missed, the basic search was repeated for the years 2000 to July 2010 (MEDLINE only) and combined with *abortion* or *termination* in the Title/Abstract. (There were no eligible results.)

For evidence of the comparative clinical utility of 3D and 4D US versus 2D US, the Hayes Medical Technology Directory Reports on *Three-Dimensional and Four-Dimensional Ultrasound for Fetal Cardiovascular Diagnosis, Three-Dimensional and Four-Dimensional Ultrasound for Fetal Growth and Volume Measurements, Three-Dimensional and Four-Dimensional Ultrasound for Diagnosis of Fetal Head Abnormalities, Three-Dimensional and Four-Dimensional Ultrasound for Fetal Limbs and Skeletal Structures, Three-Dimensional and Four-Dimensional Ultrasound for High-Risk Pregnancies and Routine Screening,* and *Three-Dimensional and Four-Dimensional Ultrasound for Extrafetal and Maternal Structures in Pregnancy* were reviewed. Results from annual update searches conducted by Hayes on these topics were also reviewed. No studies other than studies of accuracy were identified.

For studies of the clinical utility of US by setting, the years 2000 to July 2010 were searched (MEDLINE only). The basic search was combined with *managed care, health maintenance organization, HMO, government, worker's compensation, workmen's compensation,* and *Medicaid* as keywords and with *Workmen's Compensation, Managed Care Programs, Delivery of Health Care, Delivery of Health Care Integrated,* and *State Medicaid* as subject headings.

For studies of the cost implications and cost-effectiveness of US, the years 2000 to July 2010 (MEDLINE only) were searched by combining the basic search with *((((economic analysis) OR (economic evaluation)))) OR (((((cost AND (analysis OR benefit OR effective* OR consequence OR minimization)))) OR (("Costs and Cost Analysis"[MeSH] OR "Cost-Benefit Analysis"[MeSH])))).*

APPENDIX III. SYSTEMATIC REVIEWS EVALUATING THE EFFECTIVENESS AND SAFETY OF ULTRASOUND IN HIGH-RISK PREGNANCY

A. Doppler Ultrasound of Fetal Circulation (Alfirevic et al., 2010)

Key: BPD, bronchopulmonary dysplasia; BPP, biophysical profile; C-section, Cesarean section; CI, confidence interval; CTG, cardiotocograph; DUS, Doppler ultrasound; GA, gestational age; grp(s), group(s); GRADE, Grading of Recommendations Assessment, Development and Evaluation; HTN, hypertension; IOL, induction of labor; ITT, intention-to-**treat**; IUGR, intrauterine growth restriction; IVH, intraventricular hemorrhage; LOS, length of stay; MA, meta-analysis; MD, mean difference; NEC, necrotizing enterocolitis; NICU, neonatal intensive care unit; NNT, number needed to treat; pt(s), patient(s); NS, not significant; PTL, preterm labor; RR relative risk; SCBU, special care baby unit; SE, standard error; SGA, small for gestational age; SMD, standardized mean difference; SR, systematic review; TTTS, twin-twin transfer syndrome; US, ultrasound

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
Alfirevic et al. (2010) Cochrane Review MA to assess whether use of fetal and umbilical DUS improves subsequent obstetrical care and fetal outcomes. *Singleton pregnancies:* Almstrom, 1992; Biljan, 1992; de Rochambeau, 1992; **Haley, 1997**; Neales, 1994; **Nienhuis, 1997;** Ott, 1998*; Trudinger, 1987; Tyrrell, 1990; Williams, 2003 *Multiple pregnancies:* Giles, 2003 *Singleton+multiple (mixed) pregnancies:* Johnstone, 1993 *Unspecified:* Burke, 1992; **Hofmeyr, 1991;** Nimrod, 1992; Norman, 1992;	n=10,156 (18 studies) Search Cochrane Trials Register through September 2009 *Study inclusion criteria:* RCTs and quasi-randomized trials of fetal DUS (DUS) vs no DUS or vs concealment of DUS findings (no-DUS); published or unpublished; all languages *Study exclusion criteria:* Utero-placental DUS (unless combined w/ fetal DUS); comparisons of different forms of DUS *Pt inclusion criteria:* Pregnancies considered	*Analysis:* MA w/ fixed effects and random effects models. ITT analysis wherever available data permitted. Publication bias assessed w/ funnel plots of pooled estimate vs SE wherever >10 studies. Random-effects MA where statistical or clinical heterogeneity was found; otherwise, fixed-effects. Subgrp analyses planned *a priori* for primary outcomes only: (1) singleton vs multiple and monochorionic vs dichorionic twins for all outcomes; (2) SGA fetus, HTN/pre-eclampsia, diabetes, prolonged pregnancy, and previous pregnancy loss for primary outcomes only. Significance of subgrp differences	*Significant pooled estimates favoring DUS:* DUS vs no-DUS: Any perinatal death, serious neonatal mortality (1 singleton study), any potentially preventable perinatal death, C-section (any), emergency C-section, IOL, infant hospital LOS, antenatal admission DUS alone vs CTG: Elective C-section, emergency C-section, infant hospital LOS, antenatal admission, phototherapy *Summary of subgrp differences* DUS vs no-DUS: Effects did not differ. DUS alone vs CTG: Effects did not differ. *Nonsignificant pooled estimates favoring DUS:* DUS vs no-DUS: Stillbirth, neonatal death, Apgar <7 at 5 min, elective C-section, operative vaginal birth, GA at birth, neonatal fitting seizures, neonatal admission to SCBU and/or NICU, hypoxic ischemic encephalopathy, birth wt, phototherapy (1 singleton study), abnormal neurological development at 9 mos, fetal distress in labor, periventricular leucomalacia, antenatal hospital stay DUS along vs CTG: Any perinatal death, stillbirth, neonatal death, any potentially preventable death,	Pooled estimates of published and unpublished studies suggest that DUS reduced the overall risk of adverse fetal outcome; reduced the need for C-section and IOL; and reduced hospital utilization. Compared w/ CTG, DUS may offer the same benefits, but statistically significant benefits were demonstrated only for C-section and some measures of utilization. *Limitations of SR:* *Limitations of studies according to Alfirevic et al.:* 3 good-quality studies (Haley, 1997; Hofmeyr, 1991; Nienhuis, 1997) and the others of unclear quality, possibly because of publication before

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
Pattinson, 1994 *Conducted in the United States.	to be at high risk for fetal compromise *Pt exclusion criteria:* TTTS *Study characteristics:* Most used DUS in both grps, w/ results revealed to clinicians only in the test grp. Comparators varied (e.g., CTG, US biometry). *Clinical hx/baseline pt characteristics:* See # studies by birth category in column 1. GA as an inclusion criterion varied from 24 to >40 wks; NR in 6 studies	assessed by whether CIs for pooled estimates overlapped. *Quality assessment:* Cochrane Handbook methods (adequate sequence generation, allocation concealment, incomplete outcome data addressed, selective reporting*, other biases) for individual studies; GRADE system for bodies of evidence. *Refers to failure to report all prespecified outcomes. *Outcomes:* Primary: Any perinatal death after randomization; serious neonatal morbidity (hypoxic ischemic encephalopathy, IVH, BPD, or NEC) 2ndary: 29 additional prespecified outcomes identified (see following information) Non-prespecified: 8 outcomes (see following information) *Excluding chromosomal abnormalities, termination of pregnancies, birth before	Apgar <7 at 5 min, C-section (any), spontaneous vaginal birth, operative vaginal birth, IOL (singleton), neonatal fitting seizures (singleton), neonatal admission to SCBU, GA at birth, birth wt, antenatal admission (singleton) *Significant pooled estimates favoring no-DUS:* DUS vs no-DUS: None DUS along vs CTG: None *Nonsignificant pooled estimates favoring no-DUS:* DUS vs no-DUS: Spontaneous vaginal birth, infant intubation or ventilation, preterm labor, infant respiratory distress syndrome (1 singleton study), intraventricular hemorrhage, birth at <34 wks, hospitalization for IUGR, birth wt <5th percentile DUS along vs CTG: Infant requiring intubation or ventilation (singleton) *No eligible studies of these outcomes:* Fetal acidosis, oxytocin augmentation, neonatal resuscitation required, meconium aspiration, bronchopulmonary dysplasia, necrotizing enterocolitis, long-term infant neurodevelopmental outcomes, women's view of their care Studies did not measure outcomes at different time points. Insufficient information for sensitivity analyses for missing data. *Authors' conclusions:* DUS of the umbilical artery should be incorporated into protocols for fetal monitoring in pregnancies thought to be at risk of placental insufficiency because of hypertensive disorders and small for date fetuses; it is not clear whether women w/ risk factors such as postterm, diabetes, and uncomplicated dichorionic twin pregnancy should	CONSORT reporting guidelines were issued; some evidence of possible publication bias; no consensus on frequency of DUS or interventions that should follow abnormal DUS; most studies did not report recruitment rates. *Other limitations noted:* Bias in favor of DUS in some studies (e.g., Pattinson, 1994), DUS combined w/ biophysical assessment but no biophysical assessment in control arm and only some pts in intervention arm received DUS (Tyrrell, 1990); only 2 studies (Neales, 1994, Nienhuis, 1997) reported blinded assessment of outcomes; most studies conducted in Europe, Australia, and South Africa Assessment by Alfirevic et al. of overall evidence by key outcomes seems reasonable.

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
		fetal viability (as defined by trialists), and fetal death before use of the intervention.	undergo DUS. Alfirevic et al. (2010) concluded that DUS of the umbilical artery should be incorporated into protocols for fetal monitoring in pregnancies thought to be at risk of placental insufficiency. Their assessment of the evidence was that women w/ hypertensive disorders and small for date fetuses are good candidates for DUS but that it is not clear whether women w/ risk factors such as postterm, diabetes, and uncomplicated dichorionic twin pregnancy should undergo DUS.	
DUS vs no-DUS, change in pt management				
Singleton pregnancies: Almstrom, 1992; Neales, 1994	n=893 (2 singleton studies)	Antenatal admission (not prespecified)	*Singleton:* RR 0.72 (CI, 0.60-0.88) Events (DUS, no-DUS): 125/450, 170/443 Absolute risk reduction: 10.6 % NNT: 9	
	n=142 (1 singleton study)	Hospitalization for IUGR	RR: Singleton (1.03); NS differences	Not a prespecified outcome. No data for multiple births.
	n=426 (1 multiple birth study)	Antenatal hospital stay	MD: Multiple birth (–0.60); NS differences.	Not a prespecified outcome. No data for singleton births.
DUS vs no-DUS, perinatal mortality and morbidity				
Singleton pregnancies: Almstrom, 1992; Biljan, 1992; **Haley, 1997**; Neales, 1994; **Nienhuis, 1997;** Ott, 1998; Trudinger, 1987; Tyrrell, 1990; Williams, 2003 *Multiple pregnancies:* Giles, 2003 *Mixed pregnancies:* Johnstone, 1993; Newnham, 1991 *Unspecified:*	n=10,125 pts (16 studies)	Any perinatal death after randomization	*All pts:* RR 0.71 (CI, 0.52-0.98) Event rate (DUS, no-DUS): 1.2%, 1.7% (reported by Alfirevic et al.) Absolute risk reduction: 0.5 % NNT: 203 (CI, 103-4352) (reported by Alfirevic et al.) RRs for singleton (0.59) and for multiple, mixed, and unspecified births (CI, 0.68-0.88) were all NS; differences NS. RRs for SGA/IUGR (0.72), hypertension/pre-eclampsia (3.57), and previous pregnancy loss (0.26) were NS;	No data for diabetes or prolonged pregnancy subgrps. Alfirevic et al. considered this body of evidence to be *very-low* quality (missing information, too few events for precision, funnel plot asymmetry suggesting publication bias).

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
Burke, 1992; **Hofmeyr, 1991;** Norman, 1992; Pattinson, 1994			differences were NS; w/in SGA/IUGR risk subgrp (12 92 pts, 4 studies), RRs for singleton an unspecified subgrps followed same pattern. Sensitivity analysis did not demonstrate any effect of study quality but was underpowered; for the 3 high-quality studies, RR 0.61 (CI, 0.24-1.53). 1 vs 0 events (Haley, 1997), 0.9% vs 1.7% (Hofmeyr, 1991), 1.4% vs 0 (Nienhuis, 1997).	
Singleton pregnancies: Tyrrell, 1990 *Mixed pregnancies:* Newnham, 1991 *Unspecified:* Norman,, 1992	n=598 (3 studies)	Serious neonatal morbidity	*Singleton (n=500, 1 study):* RR 0.13 (CI, 0.02-0.99) Events (DUS, no-DUS): 1/250, 8/250 Absolute risk reduction: 2.8% NNT: 36 RR for mixed births (2.95) was NS. Data available only for previous pregnancy loss risk subgrp; numbers too small to allow estimate.	Data for all 3 studies not pooled because of so few babies w/ serious neonatal morbidity; significant heterogeneity across 3 studies. No studies of multiple pregnancies or for most risk subgrps. Alfirevic et al. considered this body of evidence to be *very low* quality (missing information, seriously too few events for precision).
Singleton pregnancies: Almstrom, 1992; Biljan, 1992; **Haley, 1997**; Neales, 1994; **Nienhuis, 1997;** Ott, 1998; Trudinger, 1987; Tyrrell, 1990; Williams, 2003 *Multiple pregnancies:* Giles, 2003 *Mixed pregnancies:* Johnstone, 1993; Newnham, 1991	n=10,225 (16 studies)	Any potentially preventable perinatal death	*All pts:* RR 0.67 (CI, 0.46-0.98) Event (DUS, no-DUS): 42/5053, 64/5172 Absolute risk reduction: 0.4 % NNT: 246 NS RRs for singleton, multiple, mixed, and unspecified births (CI, 0.58-0.88); differences NS.	

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
Unspecified: Burke, 1992; **Hofmeyr, 1991;** Nimrod, 1992; Norman, 1992; Pattinson, 1994				
	n=9560 (15 studies)	Stillbirth	NS RRs: All pts (0.65), singleton and multiple births (CI, 0.61-0.67), mixed birth (1.20), unspecified birth (1.00).	
	n=8167 (13 studies)	Neonatal death	NS RRs: All pts (0.81); singleton, multiple, mixed, and unspecified births (CI, 0.69-1.01); NS differences.	
Singleton pregnancies: Almstrom, 1992; **Haley, 1997**; Tyrrell, 1990	n=1076 (3 singleton studies)	Length of infant hospital stay	*Singleton:* SMD –0.28 (CI, –0.40 to –0.16)	
	n=6321 pts (7 studies)	Apgar <7 at 5 min	NS RRs: All pts (0.92), singleton (0.70), mixed (1.22), and unspecified births (0.93); NS differences.	No data specific to multiple births. Alfirevic et al. considered this body of evidence to be *low* quality (missing information, too few events for precision). Funnel plot asymmetry was nonsignificant.
	n=3136 (6 studies)	Infant intubation or ventilation	NS RRs: All pts (1.42); singleton (2.89), multiple (0.86), mixed (1.25); NS differences. Test for interaction suggested possible effect in singleton but not multiple birth pregnancies.	
	n=4066 (8 studies)	GA at birth	NS MDs: All pts; singleton, multiple, mixed, and unspecified births (–0.20 to 0.54); NS differences.	
	n=150 (1 study of singleton births)	Neonatal fitting seizures	NS RR: Singleton (0.35)	No data for multiple, births.
	n=626 (2 studies)	Preterm labor	NS RRs: All pts, singleton, not specified (1.03-1.18); NS differences.	No data for multiple births.
	n=107 (1 study of singleton births)	Infant respiratory distress syndrome	NS RR: Singleton (1.06)	No data for multiple births.

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
	n=9334 (12 studies)	Neonatal admission to SCBU and/or NICU	NS RRs: All pts (0.95); singleton, multiple, mixed, unspecified births (CI, 0.92-0.98); NS differences.	
	n=1045 (2 studies)	Hypoxic ischemic encephalopathy	NS RRs: All pts (0.65); singleton (0.09), mixed (4.91); NS differences.	No data specific to multiple births.
	n=2008 (4 studies)	Intraventricular hemorrhage	NS RRs: All pts (1.42); singleton (1.26), mixed (2.95); NS differences.	No data specific to multiple births.
	n=3887 (7 studies)	Birth wt	NS MDs: All pts (31.33); singleton, mixed (–48.0 to 11.89); NS differences.	No evidence specific to multiple births.
	n=976 (2 studies)	Birth at <34 wks	NS RRs: All pts (2.04); singleton, unspecified births (1.17-3.90); NS differences.	Not a prespecified outcome. No evidence specific to multiple births.
	n=150 (1 singleton study)	Phototherapy for neonatal jaundice	NS RR: Singleton (0.15); NS differences.	Not a prespecified outcome. No data for multiple births.
	n=289 (1 singleton study)	Birth wt <5th percentile	NS RR: Singleton (1.16); NS differences.	Not a prespecified outcome. No data for multiple births.
	n=545 (1 multiple birth study)	Periventricular leucomalacia	NS RR: Multiple birth (0.33); NS differences.	Not a prespecified outcome. No data for singleton births.
DUS vs No-DUS, C-section and IOL				
Singleton pregnancies: Almstrom, 1992; de Rochambeau, 1992; **Haley, 1997**; Neales, 1994; **Nienhuis, 1997;** Trudinger, 1987; Williams, 2003 *Multiple pregnancies:* Giles, 2003 *Mixed) pregnancies:* Johnstone, 1993; Newnham, 1991 *Unspecified:* Burke, 1992; **Hofmeyr, 1991;** Nimrod, 1992; Norman, 1992	n=7918 (14 studies)	C-section (elective and emergency)	*All pts:* RR 0.90 (CI, 0.84-0.97) Events (DUS, no-DUS): 921/3876, 1063/4042) Absolute risk reduction: 2.5% NNT: 39 *Singleton:* RR 0.84 (CI, 0.75-0.95) Events (DUS, no-DUS): 364/1425, 456/1504 Absolute risk reduction: 4.8 % NNT: 21 NS RRs for multiple, mixed, and unspecified (CI, 0.77-0.82); NS differences.	Alfirevic et al. considered this body of evidence to be *low* quality (missing information; visual evidence of asymmetry in funnel plot although not quite statistically significant).
Singleton pregnancies: Almstrom, 1992; **Haley, 1997;**	n=6175 (10 studies)	C-section (emergency)	*All pts:* RR 0.81 (CI, 0.67-0.98)	Significant funnel plot asymmetry (*P*=0.09),

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
Neales, 1994; **Nienhuis, 1997;** Trudinger 1987 *Multiple pregnancies:* Giles, 2003 *Mixed pregnancies:* Johnstone, 1993; Newnham, 1991 *Unspecified:* Burke, 1992; **Hofmeyr, 1991**			Events (DUS, no-DUS): 367/3033, 449/3142 Absolute risk reduction: 15.3 % NNT: 47 *Singleton (n=1482, 5 studies):* RR 0.58 (CI, 0.43-0.78) Events (DUS, no-DUS): 61/724, 113/758 Absolute risk reduction: 6.5 % NNT: 15 NS RRs for multiple, mixed, and unspecified births (CI, 0.78-1.06); NS differences.	suggesting publication bias.
	n=6627 (11 studies)	C-section (elective)	NS RRs: All pts (1.07); singleton, multiple, and mixed births (CI, 0.87-1.11); for unspecified birth (1.30); NS differences.	
Singleton pregnancies: Almstrom, 1992; **Haley, 1997**; Neales, 1994; Trudinger, 1987; Tyrrell, 1990 *Multiple pregnancies:* Giles, 2003 *Mixed pregnancies:* Johnstone, 1993, Newnham, 1991 *Unspecified:* Burke, 1992; Norman, 1992	n=5633 (10 studies)	IOL	*All pts:* RR 0.89 (CI, 0.80-0.99) Events (DUS, no-DUS): 838/2777, 955/2856 Absolute risk reduction: 2.6 % NNT: 31 NS RRs for multiple, mixed, and unspecified births (CI, 0.79-1.10); NS differences.	Alfirevic et al. considered this body of evidence to be *low* quality (missing information, evidence of heterogeneity). Funnel plot asymmetry as NS. Additional analysis ("prediction interval") suggests future study w/ RR >1 cannot be ruled out.
	n=2504 pts (5 studies)	Spontaneous vaginal birth	NS RRs: All pts (1.04); singleton, multiple, mixed, and unspecified births (CI, 0.98-1.07); NS differences.	
	n=2813 (4 studies)	Operative vaginal birth	NS RRs: All pts, singleton, unspecified births (CI, 0.92-0.97); NS differences.	No evidence, multiple or mixed pregnancies.
	n=289 (1 singleton study)	Fetal distress in labor	NS RR: Singleton (0.35); NS differences.	Not a prespecified outcome. No data for multiple births.

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
DUS vs No-DUS, other				
	n=137 (1 singleton study)	Abnormal neurological development at 9 mos	RR: Singleton (0.61); NS differences	Not a prespecified outcome. No data for multiple births.
DUS Alone vs CTG Alone, change in pt management				
Singleton: Almstrom, 1992	n=426 (1 singleton study)	Antenatal admission (risk)	*Singleton:* RR 0.70 (CI, 0.55-0.90) Events (DUS, CTG): 69/214, 97/212 Absolute risk reduction: 13.6 % NNT: 40	No data on multiple births.
	n=426 (1 singleton study)	Antenatal admission (mean #)	NS MD: Singleton (–0.60)	No data on multiple births.
DUS Alone vs CTG Alone, perinatal mortality and morbidity				
Singleton: Almstrom, 1992; **Haley, 1997**; Williams, 2003 *Not specified:* **Hofmeyr, 1991**	n=2813 (4 studies)	Any perinatal death after randomization	RRs: All pts (0.45); singleton (0.34), not specified (0.52); NS differences between subgrps. SGA/IUGR (0.33), hypertension/pre-eclampsia (3.57); NS differences. Sensitivity analysis did not demonstrate any effect of study quality but was underpowered; RR 0.58 (CI, 0.20-1.73) for 2 high-quality studies.	No data specific to multiple births and no data for pts w/ diabetes, prolonged pregnancy, or previous pregnancy loss.
	n=2813 (4 studies)	Stillbirth	NS RRs: All pts (0.48); singleton (0.28), unspecified (1.05); NS differences.	No data on multiple births.
	n=1473 (3 studies)	Neonatal death	NS RRs: All pts (0.52); singleton (1.02), unspecified (1.35); NS differences.	No data on multiple births.
	n=2813 (4 studies)	Any potentially preventable death	NS RRs: All pts (0.38); singleton (0.41), unspecified (0.35); NS differences.	No data on multiple births.
	n=2663 (3 studies)	Apgar <7 at 5 min	NS RRs: All pts (0.86); singleton (0.83), unspecified (0.93); NS differences.	No data on multiple births.
Singleton: Almstrom, 1992; **Haley, 1997** *Not specified:* **Hofmeyr, 1991**	n=576 (2 singleton studies)	Infant LOS	*Singleton:* SMD –0.25 (CI, –0.41 to –0.08)	No data on multiple births. Authors believe the number of babies involved was too few to allow conclusions (p. 19).

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
Singleton: **Haley, 1997**	n=150 (1 singleton study)	Phototherapy	*Singleton:* RR 0.15 (CI, 0.01-2.87) Events (DUS, CTG): 0/73, 3/77	No data on multiple births.
	n=576 (2 singleton studies)	Infant requiring intubation or ventilation	NS RR: Singleton (1.54)	No data on multiple births.
	n=150 (1 singleton study)	Neonatal fitting seizures	NS RR: Singleton (0.35)	No data on multiple births.
	n=2813 (4 studies)	Neonatal admission to SCBU	NS RRs: All pts (0.87); singleton (0.80), unspecified (1.00); NS differences.	No data on multiple births.
	n=1473 (3 studies)	GA at birth	NS MDs: All pts (0.23); singleton (0.26), unspecified (0.20); NS differences.	No data on multiple births.
	n=2813 (4 studies)	Birth wt	NS MDs: All pts (38.41); singleton (49.32), unspecified (–4); NS differences.	No data on multiple births.
DUS Alone vs CTG Alone, C-section and IOL				
	n=2813 (4 studies)	C-section, elective and emergency	NS RRs: All pts (0.89*); singleton (0.89), unspecified (0.91); NS differences. *Trend toward significance (CI, 0.79-1.01)	No data on multiple births.
Singleton: Almstrom, 1992; **Haley, 1997** *Not specified:* **Hofmeyr, 1991**	n=1473 (3 studies)	C-section, elective	*All pts:* RR 1.53 (CI, 1.12-2.09) Events (DUS, CTG): 88/725, 59/748 *Singleton (576 pts, 2 studies):* RR 1.69 (CI, 1.07-2.67) Events (DUS, CTG): 44/287, 26/289	No data for multiple births.
Singleton: Almstrom, 1992; **Haley, 1997** *Not specified:* **Hofmeyr, 1991**	n=1473 (3 studies)	C-section, emergency	*All pts:* RR 0.66 (CI, 0.52-0.84) Events (DUS, CTG): 30/287, 55/289 Absolute risk reduction: 45 % NNT: 12 *Singleton (576 pts, 2 studies):* RR 0.55 (CI, 0.36-0.83) Events (DUS, CTG): 63/438, 90/459 Absolute risk reduction: 5.5 % NNT: 19	No data for multiple births.

Authors/Study Design (good-quality studies are bolded; studies not published in full or unpublished are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors DUS; MD <1 or SMD <1 signifies decrease, i.e., favors DUS for adverse outcomes and favors no-DUS for GA and wt at birth)	Conclusions/Comments/ Limitations
	n=576 (2 singleton studies)	IOL	NS RR: Singleton (0.67)	No data on multiple births.
	n=1323 (2 studies)	Spontaneous vaginal birth	NS RRs: All pts, singleton , and unspecified (CI, 1.04-1.06); NS differences.	No data on multiple births.
	n=2663 (3 studies)	Operative vaginal birth	NS RRs: All pts, singleton , and unspecified (0.92-1.01); NS differences.	No data on multiple births.

[1]Unless otherwise noted, significance levels for *P* values is 0.05 and CIs are 95%.
[2] Unless otherwise noted, calculations of NNT and absolute risk reduction were performed with data supplied in the article.

(Appendix III continued)

B. Systematic Review of Transvaginal Ultrasound for Assessment of Cervical Length (Berghella et al., 2009)

Key: CI, confidence interval; CL, cervical length; GA, gestational age; grp(s), group(s); FFN, fetal fibronection; FN, false negative; IOL, induction of labor; MA, meta-analysis; MD, mean difference; NICU, neonatal intensive care unit; NR, not reported; NNT, number needed to treat; pt(s), patient(s); PPROM, preterm prelabor rupture of membranes; PTB, preterm birth; PTL, preterm labor; RCT, randomized controlled trial; RR relative risk; SMD, standardized mean difference; SR, systematic review; TVU, transvaginal ultrasound; TVU-CL, cervical length measured by transvaginal ultrasound; US, ultrasound

Authors/Study Design (best-quality study is bolded; studies *not* published in full are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR<1 favors TVU-CL; MD <1 or SMD <1 signifies decrease, i.e., favors TVU-CL for adverse outcomes)	Conclusions/Comments/ Limitations
Berghella et al. (2009) Cochrane Review MA to assess effectiveness of antenatal management based on TVU-CL screening for prevention of PTB. *Symptomatic, singleton, signs and/or symptoms of PTL:* Alfirevic, 2007; **Ness, 2007**; Palacio, 2006 *Symptomatic, singleton, signs and/or symptoms of PPROM:* Carlan, 1997 (primary outcome was safety [infection], not effect on management) *Asymptomatic, twins, at risk for PTL:* Gordon 2006	n=507 (5 RCTs) Searched Cochrane Trials Register and MEDLINE to September 2008. *Study inclusion criteria:* RCTs, cluster-randomized trials, and quasi-randomized trials; both published and unpublished trials; pregnant women being screened w/ TVU assessment of CL; GA 14-34 wks; knowledge of vs concealment of TVU-CL results *Primary population of interest* was singleton gestations w/ signs and/or symptoms of PTL. *Study exclusion criteria:* NR *Study characteristics:* In some studies, both arms underwent TVU-CL and pts were randomized to grps where results were or were not available the managing	*Analysis:* MA w/ fixed effects and random effects models. Subgrp analyses: PTL and PPROM; singleton and twin Significance of subgrp differences assessed by which CIs for pooled estimates overlapped. *Quality assessment:* (adequate sequence generation, allocation concealment, incomplete outcome data addressed, selective reporting*, other biases) for individual studies. *Refers to failure to report all prespecified outcomes. *Outcomes:* Primary: PTB (<37 wks for singleton, <34 wks for Twin) 2ndary: 21 additional prespecified outcomes, including a composite measure of perinatal outcome (perinatal death,	*Significant outcomes favoring knowledge of TVU-CL:* Symptomatic, singleton, signs and/or symptoms of PTL: GA at delivery *NS outcomes favoring knowledge of TVU-CL:* Symptomatic, singleton, signs and/or symptoms of PTL: PTB <37 wks and <34 wks, birth wt <2500 g, tocolysis* Symptomatic, singleton, signs and/or symptoms of PPROM: Birth wt <2055 g, chorioamnionitis Asymptomatic, twins, at risk for PTL: PTB <34 wks, 32 wks, and 20 wks; GA at delivery; birth wt; steroids for fetal lung maturity* *NS outcomes favoring no knowledge:* Symptomatic, singleton, signs and/or symptoms of PTL: Maternal hospitalization*, steroids for fetal lung immaturity* Symptomatic, singleton, signs and/or symptoms of PPROM: Endometritis, neonatal infection Asymptomatic, twins, at risk for PTL: PTB<36 wks, maternal hospitalization for PTL*, tocolysis* *Pooled estimates favoring neither:* Perinatal death *RR <1 are listed as favoring knowledge of TVU-CL, which could mean less inappropriate care and lower utilization costs, but could also signify lack of appropriate care. See additional detail on these outcomes.	Pooled estimates of published and unpublished studies showed that management based on TVU-determined CL in pts w/ suspected PTL or at risk of PTL may reduce the incidence of PTB. The appropriateness of tocolysis and steroids may be increased by use of TVU-CL. *Limitations of SR:* Limited range of study quality criteria, i.e., no "other biases" considered, as in Alfirevic et al. (2010); authors did not clarify clinical meaning of outcomes; primary outcomes of selected studies were different from the primary outcomes of the SR. No studies satisfied all quality criteria; **Ness 2007** satisfied all criteria except blinding (but an independent sonographer conducted TVU-CL); small # trials may be the reason for

Authors/Study Design (best-quality study is bolded; studies *not* published in full are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR<1 favors TVU-CL; MD <1 or SMD <1 signifies decrease, i.e., favors TVU-CL for adverse outcomes)	Conclusions/Comments/ Limitations
	obstetrician; other studies randomized women to TVU or no TVU; no cluster trials; thresholds for short cervix were 15 mm, 20 mm, and 25 mm, depending on study.	respiratory distress syndrome, intraventricular hemorrhage, necrotizing enterocolitis, sepsis) Non-prespecified: Chorioamnionitis, endometritis, neonatal infection	*No eligible studies of these outcomes:* Fetal death, neonatal death, respiratory distress syndrome, intraventricular hemorrhage, necrotizing enterocolis, sepsis, NICU admission, NICU days, maternal wellbeing, economic analysis, cervical cerclage *Authors' conclusions:* Current evidence is insufficient to recommend routine TVU-CL screening of asymptomatic or symptomatic pregnant women. Nonsignificant results favoring TVU-CL for PTB suggest the need for further research. Future research should include all populations, cost-effectiveness analyses, and protocols for management based on TVU-CL results.	NS results; no evidence for asymptomatic women w/ singleton gestation
TVU-CL Knowledge vs No Knowledge (singletons w/ PTL)				
Ness, 2007; Palacio, 2006	n=242 (2 studies)	PTB <37 wks	RR 0.59 (CI, 0.26-1.32)(NS) Events rates (knowledge, no knowledge): 22.3%, 34.7% (reported by Berghella et al.) Absolute risk reduction: 12.4 % NNT: 8 *Results reported by Ness, 2007:* Event rates (knowledge, no knowledge): 13.0%, 36.3% (*P*=0.01) Absolute risk reduction: 23.2 % NNT=5	Primary outcome for SR.
Alfirevic, 2007; **Ness, 2007**; Palacio, 2006	n=256 (3 studies)	PTB<34 wks	RR 0.55 (NS)	
Alfirevic, 2007; **Ness, 2007**	n=137 (2 studies)	PTB <28 wks	No events in either arm	
Alfirevic, 2007; **Ness, 2007**; Palacio, 2006	n=290 (3 studies)	GA at delivery	MD 0.64 (CI, 0.03-1.25) (NOTE: Assumed unit is wks; differences reported in individual studies were 3 days, 1.2 wks, and 0.3 wks.)	Results primarily determined by **Ness, 2007**; FFN was used in combination w/ TVU-CL findings of 20 mm-29 mm to determine management.
Ness, 2007	n=70 (1 study)	Birth wt <2500 g	RR 0.71 (NS)	
Palacio, 2006	n=97 (1 study)	Perinatal death	No events in either arm	

Authors/Study Design (best-quality study is bolded; studies *not* published in full are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR<1 favors TVU-CL; MD <1 or SMD <1 signifies decrease, i.e., favors TVU-CL for adverse outcomes)	Conclusions/Comments/ Limitations
Ness, 2007	n=93 (1 study)	Maternal hospitalization	RR 2.94 (NS) Ness et al. reported that 19.6% women in knowledge grp and 4.1% in control grp (*P*=0.03) were admitted, 8 for PTL or medical indication. After eliminating women admitted against protocol w/ long CL, admission was still higher in the knowledge grp but difference was less significant (*P*=0.05). Palacio et al. reported a significant (*P*=0.004) 1-day reduction in maternal hospital stay, which was the study's primary outcome.	Appropriateness of hospitalization not assessed.
Alfirevic, 2007; **Ness, 2007**	n=102 (2 studies)	Tocolysis	RR 0.85 (NS)	Alfirevic et al. reported a RR of 0.16 (CI, 0.05-0.39) for *inappropriate* tx (tocolysis plus steroids given and delivery ≥7 days later or no steroids given and delivery <7 days late). Ness et al. reported that failure to give steroids prior to preterm delivery was *not* increased in the TUV-CL arm.
Alfirevic, 2007; **Ness, 2007**	n=141 (2 studies)	Steroids for fetal lung maturity	RR 1.72 (NS)	
TVU-CL Knowledge vs No Knowledge (singletons w/ PPROM)				
Carlan, 1997	n=92	Birth wt <2500 g	MD 31.0 (NS)	
Carlan, 1997	n=92	Chorioamnionitis	RR 0.72 (NS)	
Carlan, 1997	n=92	Endometritis	RR 1.39 (NS)	
Carlan, 1997	n=92	Neonatal infection	RR 1.18 (NS)	
TVU-CL Knowledge vs No Knowledge (twins at risk for PTL)				
Gordon, 2006	n=125	PTB <36 wks	RR 1.27 (NS)	
Gordon, 2006	n=125	PTB <34 wks	RR 0.62 (CI, 0.30-1.25)(NS)	Primary outcome for SR. Gordon et al. used life-table analysis to measure differences in delivery rates; significant at 35 wks (*P*=0.02).

Authors/Study Design (best-quality study is bolded; studies *not* published in full are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR<1 favors TVU-CL; MD <1 or SMD <1 signifies decrease, i.e., favors TVU-CL for adverse outcomes)	Conclusions/Comments/ Limitations
Gordon, 2006	n=125	PTB <32 wks	RR 0.56 (NS)	
Gordon, 2006	n=125	PTB <30 wks	RR 0.20 (NS)	
Gordon, 2006	n=125	GA at delivery	MD 0.20 (NS)	
Gordon, 2006	n=125	Birth wt <2500 g	MD 155.0 (NS)	
Gordon, 2006	n=125	Maternal hospitalization	RR 1.29 (NS)	May represent increase in unnecessary admission or increase in appropriate admission.
Gordon, 2006	n=125	Tocolysis	RR 1.34 (NS)	May represent increase in unnecessary tx or increase in appropriate tx.
Gordon, 2006	n=125	Steroids for fetal lung maturity	RR 0.79 (NS)	May represent reduction in unnecessary tx or failure to treat appropriately.

[1]Unless otherwise noted, significance levels for *P* values is 0.05 and CIs are 95%.
[2]Unless otherwise noted, calculations of NNT and absolute risk reduction were performed with data supplied in the article.

(Appendix III continued)

C. Systematic Review of Ultrasound-Predicated Versus History-Predicated Cerclage (Blikman et al., 2008)

Key: CI, confidence interval; CL, cervical length; grp(s), group(s); hx, history; IOL, induction of labor; ITT, intention-to-treat; MA, meta-analysis; NNT, number needed to treat; pt(s), patient(s); PTB, preterm birth; RCT, randomized controlled trial; RR relative risk; SR, systematic review; STL, 2nd trimester loss; TVU, transvaginal ultrasound; TVU-CL, cervical length measured by transvaginal ultrasound; US, ultrasound

Authors/Study Design	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors TVU-CL)	Conclusions/Comments/ Limitations
Blikman et al. (2008) Narrative SR to assess whether US surveillance w/ cerclage as indicated results in better pregnancy outcomes and/or fewer cerclage-related complications than hx-predicated cerclage.	n=653 (6 studies) Searched Cochrane Trials Register, MEDLINE, and EMBASE to July 2007. *Study inclusion criteria:* any study design except case study; cerclage based on TVU-CL vs hx-predicated cerclage. *Population of interest*: Singleton pregnancy and an obstetric hx of PTB *Study exclusion criteria:* RCTs w/out allocation concealment, cohort studies w/out correction for important confounders, >20% missing data. *Study characteristics:* Studies accepted pts w/ PTB or STL; non-RCTs also accepted pts w/ hx of cervical surgery, torn cervix, previous forced dilitation, or cone biopsy; pregnancies w/ major fetal anomalies generally excluded; frequency of TVU and criteria for cerclage varied somewhat	MA not considered because of study heterogeneity. *Quality assessment:* (allocation concealment, selection bias, comparable grps, well-defined grps, comparable tx, missing data, ITT analysis, correction for confounding, standardized protocol) Blinding was considered impractical and was not considered. *Outcomes:* PTB according to cutoff used in each study; pregnancy loss <24 wks; % pts in TVU-CL grp who underwent cerclage	1 study omitted because of >20% missing data and no correction for confounding factors. 2 additional studies excluded because only pts who actually underwent cerclage were included in the TVU-CL grp. *PTB:* RRs >1 and NS difference 5 studies; RR <1 and NS in prospective cohort study *Pregnancy loss <24 wks:* NS difference in 3 studies, including 1 RCT; NR in 1 RCT; difference favoring TVU-CL in prospective cohort study *% women who did not undergo cerclage (TVU-CL, hx-predicated):* 40-68, 0 *Cerclage-related complications:* NR except in Groom, 2004 (no complications in either grp) *Authors' conclusions:* Mid-trimester TVU-CL can be used to optimize selection of candidates for cerclage among women at risk of cervical insufficiency (I.e., previous PTB, 2nd trimester loss, or cervical surgery) including pts w/ a classic hx for cervical insufficiency (i.e., previous PTB or 2nd trimester loss accompanied by painless and progressive dilatation of the cervix),	2 RCTs showed that a policy of TV predicated cerclage, compared w hx-predicated cerclage, resulted i fewer cerclage procedures and di not increased PTB; 1 RCT showed there was no increase in pregnan loss. Non-RCT evidence was generally consistent w/ these findings, except for a prospective cohort study showing a substanti but NS decrease in pregnancy los *Limitations of SR:* No pooled estimates. *Limitation of studies according to Blikman et al.:* See following table rows *Other study limitations:* No data c cerclage-related complications; most CIs were very wide (impreci estimates), encompassing values that would reflect benefit as well substantial harm.. *Quality assessments, considering study design and using criteria supplied by Blikman et al.:* RCTs,

Authors/Study Design	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors TVU-CL)	Conclusions/Comments/ Limitations
				good; others, very poor to fair
Beigi et al. (2005) RCT	n=97	PTB <37 wks (cutoff ≤20 mm) Pregnancy loss <24 wks No cerclage (TVU-CL grp)	RR 1.25 (NS) RR 1.08 (NS) 46%	Good quality
Althusius et al. (2000) RCT	n=73	PTB <34 wks (cutoff ≤25 mm) Pregnancy loss <24 wks No cerclage (TVU-CL grp)	1.05 (NS) Not available 59%	Good quality
Higgins et al. (2004) Prospective cohort study	n=135	PTB <30 wks (cutoff ≤25 mm) Pregnancy loss <24 wks No cerclage (TVU-CL grp)	RR 0.14 (NS) 0 losses in TVU-CL grp and 8/97 losses in control grp 68%	Fair quality
Berghella et al. (2002) Retrospective cohort study	n=177	PTB <35 wks (cutoff ≤25 mm) Pregnancy loss <24 wks No cerclage (TVU-CL grp)	RR 1.10 (NS) Not Available 64%	Very poor quality Grps differed on majority of clinic variables
To et al. (2002) Retrospective cohort study	n-90	PTB <34 wks (cutoff ≤15 mm) Pregnancy loss <24 wks No cerclage (TVU-CL grp)	RR 1.43 (NS) RR 2.86 (NS) 40%	Poor quality High pregnancy loss could be attributable to the fact that 11% women in TVU-CL grp had bulging membranes and to less frequent surveillance than in other studies
Groom et al. (2004) Case-control study	n=81	PTB Pregnancy loss <24 wks No cerclage (TVU-CL grp)	Not available RR 0.83 (NS) 64%	Poor quality

[1]Unless otherwise noted, significance levels for *P* values is 0.05 and CIs are 95%.
[2]Unless otherwise noted, calculations of NNT and absolute risk reduction were performed with data supplied in the article.

APPENDIX IV. SYSTEMATIC REVIEWS EVALUATING THE EFFECTIVENESS OF ULTRASOUND IN LOW-RISK PREGNANCY

A. Systematic Review Evaluating the Effectiveness and Safety of Ultrasound >24 Weeks (Whitworth et al., 2010)

Key: 3-D, three-dimensional; 4-D, four-dimensional; BL, baseline; BPD, biparietal diameter; C-section, Cesarean section; CTG, cardiotocograph; DUS, DUS; f/u, follow up; GA, gestational age; grp(s), group(s); IOL, induction of labor; IUGR, intrauterine growth restriction; LOS, length of stay; MA, meta-analysis; MD, mean difference; NNT, number needed to treat; pt(s), patient(s); PTL, preterm labor; QOL, quality of life; RCT, randomized controlled trial; RR relative risk; SGA, small for gestational age; SMD, standardized mean difference; subgrp, subgroup; US, ultrasound; wk(s), week(s)

Authors/Study Design	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2] (RR <1 favors US except for birth wt below a threshold; MD <1 or SMD <1 signifies decrease, i.e., favors US for utilization and control for GA/birth wt)	Comments/ Limitations
Whitworth et al. (2010) Cochrane Review MA of trials investigating US for fetal assessment during early pregnancy (<24 wks) for unselected pts Bennett 1982; Bakketeig 1984; Eik-Nes 1984; Waldenstrom, 1988; Ewigman 1990*; Saari-Kemppainen 1990; Ewigman 1993*; Salvesen 1993; Geerts 1996; Crowther 1999; Harrington 2006; van Dyk 2007 *Conducted in the United States	n= 37,505 women (11 studies) Searched MEDLINE, EMBASE, and Cochrane Controlled Trial Register to September 2009. No language restrictions. *Study inclusion criteria:* RCTs and quasi-randomized trials of US for unselected and selected pregnant pts *Study exclusion criteria:* Cross-over designs; US for high-risk pregnancy; comparison of 3-D US vs 4-D US for pregnancy; between-grp difference in US protocols; DUS *Pt inclusion criteria:* Studies including US performed <24 wks of gestation; indication for US were estimation of GA in order to improve timing of other screening tests and estimate the EDD, identify multiple pregnancies, and conduct limited examination of fetal morphology. *Pt exclusion criteria:* Pts who underwent >2 US during the study duration	*Analysis:* MA w/ fixed and random effects models; planned subgrp analysis by parity, by the timing of the early US (before 14 wks or after 14 wks) and by whether the control grp had US scans rather than selective US scans. Only subgrps for primary outcomes were examined. *Quality assessment:* Cochrane Handbook methods *Outcomes:* <u>Primary outcomes</u>: Detection of major fetal abnormality <24 wks gestation; detection of multiple pregnancy by 24 wk gestation; induction of labor for postterm pregnancy; perinatal death defined as stillbirth after trial entry, or death of a live born infant <28 days of age). *2ndary outcomes:* Detection of nonviable or ectopic pregnancy prior to clinical presentation, chorionicity; multiple pregnancy prior to labor; soft markers before 24 wks; major anomaly before birth.	*Significant RRs favoring routine US in early pregnancy:* Detection of multiple pregnancy before, between 24 wks and 26 wks or prior to delivery; detection of fetal anomaly; rate of induction of labor for postterm pregnancy and for any reason; mother not satisfied w/ care (worried about pregnancy). Termination of pregnancy for fetal abnormality was significantly increased. *Subgrp analyses:* Significant RRs US planned >14 wks for detection of multiple pregnancy between 24 wks and 26 wks and selective US for controls for detection of multiple pregnancy <24 wks. No subgroup analysis on parity conducted due to unavailability of information. Studies did not measure outcomes at different time points. Sensitivity analysis for induction for postterm pregnancy, perinatal death, and detection of multiple pregnancy and abnormality <24 wks' gestation was conducted. Results did not show change in findings after exclusion of quasi-randomized trials from analysis. *No eligible studies of these outcomes:* Detection of ectopic pregnancy; chorionicity of multiple pregnancy; laparoscopic management of ectopic pregnancy; surgical management of abortion. *Authors' conclusions:* US performed <24 wks improves the early detection of multiple pregnancies and improved gestational dating may result in fewer inductions for post maturity.	Pooled estimates of published and unpublished studies suggest that early US provides earlier detection of multiple pregnancies, better detection of clinically unsuspected fetal malformation at a time when termination of pregnancy is possible, and a reduction in the incidence of IOL. *Limitations of individual studies identified by Whitworth et al.:* Overall, well-designed studies but lack of blinding is a problem; pseudo-randomization in some trials (Bennett 1982; Neilson 1984) but sensitivity analysis suggested inclusion did not alter pooled estimates; allocation method usually not defined; lack of generalizability for some outcomes due to specific study inclusion criteria (Ewigman 1993) and for outcomes in general because of high-resource settings; publication bias was assessed but none was detected; heterogeneity for some outcomes. *Other limitations:* Only 2 trials in U.S.
Routine US vs control grp (selective/concealed ultrasound), detection of fetal abnormalities and multiple gestation				
Bennett 1982; Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Saari-Kemppainen 1990; Ewigman 1993	n= 295 (7 studies)	Failure to detect multiple pregnancy between 24 wks and 26 wks	RR 0.07 (CI, 0.03 – 0.17) Events rate (US vs control): 1%, 39% NNT: 3	One study w/both high and low risk pts

			Subgrp analyses: US planned <14 wks: RR 0.89 (CI, 0.05-16.36) NOTE: Significant differences were not demonstrated. US planned >14 wks: RR 0.06 (CI, 0.02-0.16) Events rate (US vs control) (%): 1, 4 NNT: 3	
Bennett 1982; Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Saari-Kemppainen 1990; Ewigman 1993	n= 295 (7 studies)	Failure to detect multiple pregnancy <24 wks (# not detected since concealed results)	RR 0.07 (CI, 0.03-0.17) Events rate (US vs control) (%): 1, 39 NNT: 3 *Subset analysis:* Concealed results for controls: RR 0.17 (CI, 0.01-2.92) Selective US for controls: RR 0.07 (CI, 0.02-0.17) Events rate (US vs control) (%): 1, 39 NNT: 3 NOTE: Significant differences were not demonstrated.	One study w/both high- and low-risk pts
Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Saari-Kemppainen 1990; Ewigman 1993	n=273 unselected pregnant pts (5 studies)	Failure to detect multiple pregnancies prior to delivery	RR 0.12 (CI, 0.03-0.54) Events rate (US vs control) (%): 0, 9 NNT: 12	One study w/both high- and low-risk pts
Eik-Nes 1984; Ewigman 1993	n=387 pts (2 studies)	Detection of fetal abnormalities <24 wks	RR: 3.46 (CI, 1.67-7.14) Event rate (US, no-US) (%): 16, 4 NNT: 9	Heterogeneity among studies in terms of low- and high-risk pts; in Ewigman large # of pts ineligible for inclusion (15,530/53,367 randomized); BL difference between tx and control grps in terms of smoking status leading to fewer fetal abnormalities in tx grp (64% vs 69%; *P*=0.02); results from included studies may not have current relevance due to technical advances in equipment, training and expertise of operators and more widespread use of US.
Eik-Nes 1984; Ewigman 1993	n=387 (2 studies)	Detection of major fetal	RR 3.19 (1.99-5.11)	One study w/both high- and

		anomaly before birth	Events rate (US vs control) (%): 32, 9 NNT: 5	low-risk pts; results from included studies may not have current relevance due to technical advances in equipment, training and expertise of operators and more widespread use of US.
Routine US vs control grp (selective/concealed ultrasound), change in pt management				
Saari-Kemppainen 1990; van Dyk 2007	n=9502 (2 studies)	# of antenatal visits	NS RR 0.16	
Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Geerts 1996; van Dyk 2007	n=17,785 (6 studies)	Antenatal hospital admission	NS RR 1.04	Pts w/multiple gestation included in analysis
Crowther 1999	n=602 (1 study)	Inappropriately timed serum screening tests	NS RR 0.89	
Crowther 1999	n=602 (1 study)	Inappropriately timed anomaly scan (18 wks to 22 wks	NS RR 0.77	
Routine US vs control grp (selective/concealed ultrasound), perinatal morbidity and mortality				
Bennett 1982; Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Saari-Kemppainen 1990; Ewigman 1993; Geerts 1996; Crowther 1999; van Dyk 2007	n=35,735 (10 studies)	Perinatal mortality (all babies)	NS RR 0.89 (CI 0.70-1.12)	
Bennett 1982; Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Saari-Kemppainen 1990; Ewigman 1993	n=34,331 (8 studies)	Perinatal death (excluding lethal malformations)	NS RR 0.96 (NS difference, i.e., CIs did not overlap, between this estimate and estimate for all babies.)	
Bennett 1982; Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Saari-Kemppainen 1990; Ewigman 1993; Geerts 1996; Crowther 1999	n=34,923 (9 studies)	Perinatal death (earlier and later scans)	NS RR 0.87 *Subgrp analysis:* Earlier scan (<14 wks): RR 0.73 Later scan (>14 wks): RR 0.87 NOTE: Significant differences were not demonstrated.	Pts w/multiple gestation included in analysis
Eik-Nes 1984; Waldenstrom 1988; Ewigman 1993; Geerts 1996; van Dyk 2007	n=23,213 (5 studies)	Mean birth wt	NS RR 10.67	
Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Geerts 1996; Crowther 1999; Ewigman 1990; Saari-	n=19,337 (8 studies)	Low birth wt (<2.5 kg)	NS RR 1.04 *Subgrp analysis:*	

Kemppainen 1990; van Dyk 2007			Singleton: RR 0.83 All babies or not clear: RR 1.28 NOTE: Significant differences were not demonstrated.	
Geerts 1996; Crowther 1999	n=1584 (2 studies)	Very-low birth wt (<1500 g)	NS RR 1.26 *Subgrp analysis:* Singleton: RR not estimable All babies or not clear: NS RR 1.26	
Bakketeig 1984; Ewigman 1993; Geerts 1996	n=17,105 (3 studies)	Rate of SGA	NS RR 1.05	
Bakketeig 1984; Eik-Nes 1984; Ewigman 1990; Crowther 1999	n=3906 (4 studies)	Apgar <7 at 5 min	NS RR 0.76	Pts w/multiple gestation included in analysis
Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Geerts 1996; Ewigman 1990; Saari-Kemppainen 1990; Crowther 1999; van Dyk 2007	n=19,088 (8 studies)	Admission to neonatal intensive care unit	NS RR 0.95	Pts w/multiple gestation included in analysis
Routine US vs control grp (selective/concealed ultrasound), C-section, IOL, and delivery methods				
Bakketeig 1984; Waldenstrom 1988; Ewigman 1993; Harrington 2006; van Dyk 2007	n=22,193 (5 studies)	C-section	NS RR 1.05	
Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Ewigman 1990; Ewigman 1993; Geerts 1996; Harrington et 2006; van Dyk al., 2007	n= 25,516 unselected pregnant pts (8 studies)	Rate of induction of labor for postterm pregnancy	RR 0.59 (CI, 0.42 – 0.83) Events rate (US vs control) (%): 2, 3 NNT: 100 *Sub-grp analysis:* US <14 wks: NS RR 0.99 NOTE: Significant differences were not demonstrated. US >14 wks: RR 0.49 (CI, 0.31-0.77) Events rate (US vs control) (%): 2, 3 NNT: 25	Heterogeneity among studies; 1 study w/both high- and low-risk pts; high loss to f/u in 1 RCT
Bennett 1982; Eik-Nes 1984; Bakketeig 1984; Waldenstrom 1988; Ewigman 1990; Ewigman 1993; Harrington 2006	n=24,790 (7 studies)	Induction of labor for any reason	RR 0.78 (CI, 0.63-0.97) Events rate (US vs control) (%): 18, 19 NNT: 100	One study w/both high- and low-risk pts.
Routine US vs control grp (selective/concealed ultrasound), abortion rate				
Eik-Nes 1984; Saari-Kemppainen 1990; Ewigman 1993; Geerts 1996; van Dyk 2007	n=28,256 (5 studies)	Termination of pregnancy for fetal abnormality	RR 2.23 (CI, 1.10-4.54) Events rate (US vs control): 0.17, 0.07	One study w/both high- and low-risk pts; large differences in the detection rates between the 2 hospitals involved in Saari-Kemppainen

				et al.(1990) study, which suggests that difference in diagnostic expertise can effect overall performance of US.
Routine US vs control grp (selective/concealed ultrasound), growth and childhood development				
Salvesen 1993	n=1657 (1 study)	Impaired development at childhood f/u (screened using the Denver developmental screening test)	NS RR 0.95	
Salvesen 1993	n=1993 (1 study)	Poor oral reading at school	NS RR 1.02	
Salvesen 1993	n=1984 (1 study)	Poor reading comprehension at school.	NS RR 0.82	
Salvesen 1993	n=1982 (1 study)	Poor spelling at school.	NS RR 0.85	
Salvesen 1993	n=1993 (1 study)	Poor arithmetic at school	NS RR 0.90	
Salvesen 1993	n=1993 (1 study)	Poor overall school performance	NS RR 0.96	
Salvesen 1993	n=603 (1 study)	Dyslexia	NS RR 0.77	
Waldenstrom 1988; Salvesen 1993	n=5418 (2 studies)	Reduced hearing in childhood	NS RR 0.90	
Waldenstrom 1988; Salvesen 1993	n=5417 (2 studies)	Reduced vision in childhood	NS RR 0.83	
Waldenstrom 1988; Salvesen 1993	n=5331 (2 studies)	Use of spectacles	NS RR 0.88	
Waldenstrom 1988; Salvesen 1993	n=4715 (2 studies)	Non–right-handedness	NS RR 1.12	
Salvesen 1993	n=1663 (1 study)	Ambidexterity	NS RR 1.23	
Routine US vs control grp (selective/concealed ultrasound), maternal outcomes				
Crowther 1999	n=634 unselected pregnant pts (1 study)	Mother not satisfied w/ care (worried about pregnancy)	RR 0.80 (CI, 0.65-0.99) Events rate (US vs control) (%): 31, 39 NNT: 13	Missing data (<10%) for some outcomes.

[1]Unless otherwise noted, significance levels for *P* values is 0.05 and CIs are 95%.

[2] Unless otherwise noted, calculations of NNT and absolute risk reduction were performed with data supplied in the article.

(Appendix IV continued)

B. Routine Ultrasound in Late Pregnancy (>24 Weeks) (Bricker et al., 2008)

Key: BL, baseline; BPD, biparietal diameter; C-section, Cesarean section; CTG, cardiotocograph; DUS, DUS; f/u, follow up; GA, gestational age; grp(s), group(s); IOL, induction of labor; IUGR, intrauterine growth restriction; LOS, length of stay; MA, meta-analysis; MD, mean difference; NNT, number needed to treat; pt(s), patient(s); PTL, preterm labor; QOL, quality of life; RR relative risk; SGA, small for gestational age; SMD, ; subgrp, subgroup; US, ultrasound)

Authors/Study Design	Study Population	Methods/Treatment/ Outcome Measures	Results [1,2] (RR <1 favors US except for birth wt below a threshold; MD <1 or SMD<1 signifies decrease, i.e., favors US for utilization and control for GA/birth wt)	Comments/Limitations
Bricker et al. (2008) Cochrane Review MA of trials investigating routine US for fetal assessment during late pregnancy (>24 wks) *Studies included:* Bakketeig 1984; Neilson 1984 (quasi-randomized); Proud 1987; Salvesen et al 1992; Duff 1993; Ewigman et al 1993*; Newnham et al 1993; Eik-Nes 2000; McKenna 2003 *Conducted in the United States	n= 27,024 women (8 studies) Searched Cochrane Pregnancy and Childbirth Group's Trials Register until February 2008. No language restrictions. *Study inclusion criteria:* RCTs and quasi-randomized trials of US for unselected and low-risk pregnant pts; published or unpublished	*Analysis:* MA w/ fixed effect and random effects models. Planned subgrp analysis by low-risk vs unselected pts and purpose of US exam. Sensitivity analysis conducted based on quality of randomization. *Quality assessment:* Cochrane Handbook methods *Outcomes:* *Primary outcomes*: IOL; C-section, all deaths (perinatal, neonatal, and infant); preterm delivery <34 wks; neurodevelopment at age 2 yrs; maternal psychological effects.	ROUTINE US vs NO/CONCEALED/SELECTIVE US *Effect on primary outcomes:* None on IOL or C-Section; NS effect on preterm delivery <37 wks and no data on <34 wks; insufficient data on neurodevelopment; no data on maternal psychological effects. *Significant pooled estimates favoring routine US compared w/ no/concealed/selective US:* # hospital days, stillbirths (excluding congenital abnormalities), postterm delivery >42 wks' gestation *Nonsignificant pooled estimates of primary outcomes favoring US:* Perinatal mortality, preterm delivery <37 wks, IOL *Significant pooled estimates favoring control grps:* GA at delivery (slight difference), mean birth wt (slight difference) *Nonsignificant pooled estimates of primary outcomes favoring control:* Neonatal deaths, stillbirths (except when congenital abnormality was excluded), C-section	Pooled estimates of published studies suggest that late US does not lead to significant changes in pt management or outcomes. Exceptions are a small impact on # of days spent in hospital postdelivery, a small but statistically nonsignificant increase in the C-section rate in the screened pts, and possible reduction in the postterm delivery rate and stillbirth rate if placental grading is incorporated into routine 3rd trimester US, and, when US is used serially, a possible increase in low-birth wt infants.

	Study exclusion criteria: Accuracy studies; US for high-risk pregnancy; comparison of 3-D US vs 4-D US for pregnancy; between-grp difference in US protocols *Pt inclusion criteria:* US performed >24 wks of gestation to assess fetal size, amniotic fluid volume, placental site, placental grading, fetal structural anatomy, fetal presentation *Pt exclusion criteria:* Pts which underwent >2 US during the study duration	*Secondary outcomes:* Antenatal admission to hospital; intention to deliver; GA at birth; acute neonatal problems; psychological effects (including stress, anxiety, depression, QOL, satisfaction); postterm delivery > 42 wks; birth wt <5th percentile; moderate and severe neonatal morbidity; perinatal mortality of twins	SERIAL US AND DUS vs ROUTINE US *Significant pooled estimates favoring serial US and DUS:* None *Nonsignificant pooled estimates of primary outcomes favoring serial US and DUS:* Perinatal mortality, stillbirths, neonatal deaths, C-section *Significant pooled estimates favoring control grps:* Birth wt <10th and 3rd percentiles *Nonsignificant pooled estimates of primary outcomes favoring control grps:* IOL (RR very close to 1) *Pooled estimates favoring neither US or control:* CTG (serial US and DUS vs selective US) Studies did not measure outcomes at different time points. Sensitivity analysis did not show change in findings after exclusion of quasi-randomized trials from analysis. Subgrp analyses were not performed due to the limited data. *Author's conclusions:* Routine US does not improve maternal and perinatal outcomes during late pregnancy in low-risk or unselected populations, except reduction in # of still births. It may be associated w/ a small increase in C-section rates. Placental grading in the 3rd trimester may be important in order to reduce labor induction rate for postterm delivery.	*Limitations of individual studies according to information provided by Bricker et al.:* Authors described the quality of studies as "satisfactory"; all studies lacked blinding. selection bias due to pseudo-randomized RCTs (Neilson 1984) but sensitivity analysis based on this study showed no change in effect estimates; reviewed studies did not report data for several 2ndary outcomes prespecified in the review protocol; US protocols varied between studies high levels of heterogeneity for some outcomes. The authors did not test for publication bias. *Other limitations:* Most of the studies were conducted in Europe rather than the United States. .
Routine US vs control grp (no/concealed/selective US), change in pt management				
Neilson 1984 (quasi-randomized)	n=877 (1 study)	# of hospital days	MD 0.10 (CI, 0.07-0.13)	No blinding; difference in BL characteristics between grps since more pts from social Class V in the tx grp
	n=5396 pts (4 studies)	Antenatal admission	RR 1.07 (NS)	Two studies included pts w/multiple gestation.
	n=2536 (2 studies)	Further US scans	RR 0.91 (NS)	
	n=2000 (1 study)	Cardiography	RR 1.02 (NS)	

Routine US vs control grp (no/concealed/selective US), reduction in perinatal mortality and morbidity (including intermediate outcomes)				
	n=24,276 (7 studies)	Perinatal mortality	RR 0.94 (CI, 0.55-1.61) (NS)	Two studies included pts w/multiple gestation.
	n=21,734 (5 studies)	Perinatal mortality (excluding congenital abnormalities)	RR 0.88 (NS); NS difference compared w/ overall estimate	Two studies included pts w/multiple gestation. Correction for abnormality increased heterogeneity because of Proud, 1987.
	n=314 (3 studies)	Perinatal mortality (twins)	RR 0.63 (NS); NS difference compared w/ overall estimate	One study included pts w/multiple gestation.
	n=21,708 (5 studies)	Still birth	RR 1.11 (CI, 0.25-4.26) (NS) One study (Peterborough, 1987) was unique in combining placental grading w/ US; RR was 0.05 (CI, 0-0.90).	One study included pts w/multiple gestation.
Neilson 1984 (quasi-randomized); Proud 1987	n=2877 (2 studies)	Stillbirths (excluding congenital abnormalities)	RR 0.05 (CI, 0.00-0.90); NS difference compared w/ overall estimate. Event rate (US, Control) (%): 0, 1 NNT: 100 Proud (1987) was unique in combining placental grading w/ US; this factor may be related to findings. Exclusion of Neilson et al. (1984) in sensitivity analysis did not affect results.	No blinding; difference in BL characteristics between grps in Neilson study since more pts from social Class V in the tx grp, no information provided regarding the social scale used.
	n=21,708 (5 studies)	Neonatal deaths	RR 1.04 (CI, 0.58-1.85) (NS)	One study included pts w/multiple gestation.
	n=2902 (2 studies)	Neonatal deaths (excluding congenital abnormalities)	RR 1.99 (NS); NS difference compared w/ overall estimate.	One study included pts w/multiple gestation.
	n=17,151 (2 studies)	Preterm delivery at <37 wks	RR 0.96 (NS)	
Bakketeig 1984; Proud 1987	n=17,151 (2 studies)	Postterm delivery rate>42 wks	RR 0.69 (CI, 0.59-0.81) Event rate (US, Control) (%): 3, 4 NNT: 100	No pt blinding
	n=5889 (4 studies)	Apgar score <7 at 5 min	RR 0.89 (NS)	One study included pts w/multiple gestation.
Neilson 1984 (quasi-randomized); Proud 1987	n=2877 (2 studies)	GA at delivery	MD -0.16 (CI, -0.26 to -0.06)	No blinding; difference in BL characteristics between grps since more pts from social Class V in the tx grp
	n=19,710 (4 studies)	Birth wt	MD -0.47 (NS)	One study included pts w/multiple gestation.

	n=4510 (3 studies)	Low birth wt <2.5 kg	RR 0.92 (NS)	One study included pts w/multiple gestation.
	n=2404 (2 studies)	Birth wt <5th percentile	RR 1.18 (NS)	
	n=20,298 (4 studies)	Birth wt <10th percentile	RR 0.98 (NS)	
	n=6539 (4 studies)	Admission to special care baby unit	RR 0.95 (NS)	One study included pts w/multiple gestation.
	n=15,281 (1 study)	Moderate neonatal morbidity	RR 0.97 (NS)	
	n=15,281 (1 study)	Severe neonatal morbidity	RR 1.03 (NS)	
	n=6533 (4 studies)	Neonatal resuscitation	RR 0.95 (NS)	One study included pts w/multiple gestation.
	n=3044 (2 studies)	Neonatal ventilation	RR 0.64 (NS)	One study included pts w/multiple gestation.
Routine US vs control grp (no/concealed/selective US), C-section, IOL, and delivery methods				
	n=22,663 (6 studies)	IOL	0.93 (CI, 0.81-1.07)(NS)	
	n= 21,035 pts (5 studies)	Rate of C-section (elective and emergency)	RR 1.06 (CI, 1.00-1.13) (NS)	One study included pts w/multiple gestation.
	n=5884 (4 studies)	C-section (elective)	RR 1.09 (NS)	One study included pts w/multiple gestation.
	n=5884 (4 studies)	C-section (emergency)	RR 1.11 (NS)	One study included pts w/multiple gestation.
	n=5884 (4 studies)	Instrumental delivery	RR 1.04 (NS)	
Routine US vs control grp (no/concealed/selective US), rate of abortion for fetal anomaly				
No studies				
Routine serial routine US and DUS vs selective US, change in pt management				
Newnham 1993	n=2834 (1 study)	Admission to special baby care unit	RR 0.95 (NS)	
Newnham 1993	n=2834 (1 study)	CTG	RR 1.01 (NS)	Lack of generalizability due to results limited to single study.
Routine serial routine US and DUS vs selective US, reduction in perinatal mortality and morbidity (including intermediate outcomes)				
Newnham 1993	n=2834 (1 study)	Perinatal mortality	RR 0.59 (NS)	
Newnham 1993	n=2834 (1 study)	Stillbirths	RR 0.84 (NS)	
Newnham 1993	n=2834 (1 study)	Neonatal deaths	RR 0.30 (NS)	
Newnham 1993	n=2834 (1 study)	Neonatal deaths (excluding congenital abnormalities)	RR 0.04 (NS); NS difference compared w/ overall neonatal deaths	
Newnham 1993	n=2834 (1 study)	GA at delivery	MD -0.10 (NS)	Lack of generalizability due to results limited to single study.
Newnham 1993	n=2834 (1 study)	Apgar <7 at 5 min	RR 0.77 (NS)	

Newnham 1993	n=2834 (1 study)	Birth wt	MD -25.00 (NS)	Lack of generalizability due to results limited to single study.
Newnham 1993	n=2834 (1 study)	Low birth wt (<2.5 kg)	RR 1.14 (NS)	Lack of generalizability due to results limited to single study.
Newnham 1993	n=2834 (1 study)	Very low birth wt (<1.5 kg)	RR 1.27 (NS)	Lack of generalizability due to results limited to single study.
Newnham 1993	n=2834 (1 study)	Birth wt <10th percentile	RR 1.36 (CI, 1.10-1.68) Event rate (routine US/DUS vs selective US) (%):12, 9 NNT: 30	Low birth wt possibly attributable to US exposure since the tx grp received DUS and US at 18, 24, 34, and 38 wks compared w/control grp once only at 18 wks.
Newnham 1993	n=2834 (1 study)	Birth wt <3rd percentile	RR 1.66 (CI, 1.10 – 2.51) Event rate (routine US/DUS vs selective US) (%): 4, 3 NNT: 62	Low birth wt possibly attributable to US exposure since the tx grp received DUS and US at 18, 24, 34, and 38 wks compared w/control grp once only at 18 wks.
Newnham 1993	n=2834 (1 study)	Neonatal intraventricular hemorrhage	RR 0.80 (NS)	
Newnham 1993	n=2834 (1 study)	Neonatal resuscitation	RR 0.98 (NS)	
Newnham 1993	n=2834 (1 study)	Neonatal ventilation	RR 0.67 (NS)	
Routine serial routine US and DUS vs selective US, C-section and IOL				
Newnham 1993	n=2834 (1 study)	C-section	RR 0.89 (NS)	
Newnham 1993	n=2834 (1 study)	C-section (elective)	RR 0.95	Lack of generalizability due to results limited to single study.
Newnham 1993	n=2834 (1 study)	C-section (emergency)	RR 0.82	Lack of generalizability due to results limited to single study.
Newnham 1993	n=2834 (1 study)	IOL	RR 1.02 (NS)	Lack of generalizability due to results limited to single study.
Routine serial routine US and DUS vs selective US, rate of abortion for fetal anomaly				
No studies				

[1] Unless otherwise noted, significance levels for *P* values is 0.05 and CIs are 95%.
[2] Unless otherwise noted, calculations of NNT and absolute risk reduction were performed with data supplied in the article.

APPENDIX V. SYSTEMATIC REVIEW EVALUATING THE SAFETY OF ULTRASOUND

Systematic Review Evaluating the Safety of Routine Ultrasound (Torloni et al., 2009)

Key: Approx, approximately; C-section, Cesarean section; DUS, DUS; GA, gestational age; grp(s), group(s); IOL, induction of labor; LOS, length of stay; STROBE, Strengthening the Report of Observational Studies in Epidemiology; MA, meta-analysis; NNT, number needed to treat; pt(s), patient(s); OR odds ratio; RCT, randomized controlled trial; RES, reticuloendothelial system; SGA, small for gestational age; SR, systematic review; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; US, ultrasound; wt, weight

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
Torloni et al. (2009) Systematic Review SR w/ MA of trials investigating long-term and short-term effects of fetal and maternal exposure to routine US during pregnancy	41 studies: 16 RCTs (all but 2 were RCTs), 13 cohorts and 12 case-controls Searched MEDLINE and Cochrane RCT Register to October 2007. No language restrictions. *Study inclusion criteria:* RCTs, prospective and retrospective controlled observational studies that assessed any type of short- and long-term effects of ≥1 exposure to US during pregnancy among unselected pts; exposure to static or B-mode US alone or associated w/continuous or pulsed-wave DUS or DUS alone *Study exclusion criteria:* Continuous DUS for fetal heart monitoring *Pt inclusion criteria:* Pts w/low risk pregnancy; singleton pregnancy	*Analysis:* MA by type of study design w/ fixed and random effects models; planned subgrp analysis by type of US, # exposure during pregnancy, and GA at first exposure *Quality assessment:* RCTs judged according to adequacy of allocation concealment (A, adequate; B, unclear; C, inadequate); observational studies judged according to STROBE reporting guidelines *Outcomes:* Maternal outcome; adverse perinatal outcome; abnormal childhood growth and neurological development; non–right-handedness; childhood malignancy; and intellectual performance and mental disease	*Significant ORs:* low birth wt (<2500 g), preterm birth, Apgar at 5 min <7, congenital heart malformations, speech problems such as referral to speech therapist and delayed speech, non–right-handedness for boys only, subnominal intellectual performance adult men. *Authors' conclusions:* Diagnostic US should continue to be considered relatively safe for both mother and fetus. However, the long-term effects are still uncertain and available studies might not have addressed all possible biological effects. It is not possible at this time to define the safest use of US in terms of GA, duration and # of exposures, acoustic output and fetal position). Pts should be exposed to the least amount of US energy necessary to obtain diagnostic information. No RCTs using US during 1st trimester (1st vs 3rd trimester subgrp analysis not possible)	Pooled estimates of published and unpublished studies suggest that US is generally safe in the short-term. *Limitations of Studies:* For some outcomes, no evidence from RCTs; controlled studies ranked as A (n=4), B (n=9), and C (n=3);; quality grades of the observational studies ranged from 7 to 13, w/ 88% studies (22/25) scoring ≥8 of 16 points; quantification of the intensity of acoustic exposure and duration of US examination not documented in ~90% studies; no dose–response gradient available in most studies; protocol deviations in most studies; studies included had mostly US exposures before 1995, when the acoustic potency of the equipment used was lower than in modern machine; statistical heterogeneity (addressed w/ random effects models) in 23% of RCTs, 17% observational

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
	Pt exclusion criteria: Pts w/preexisting medical condition impacting pregnancy			studies. .
Maternal outcomes				
Stark 1984	n=806 (1 retrospective cohort)	Intrapartum complications	OR 1.10 (NS)	
Stark 1984	n=806 (1 retrospective cohort)	Postpartum complications	OR 1.04 (NS)	
Omtzigt 1980; Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Saari-Kemp 1990; **Ewigman 1993**; **French Doppler study, 1997;** Geerts 1996; McKenna et al 2003	n=25,200 (9 RCTs)	Maternal admission to hospital	OR 1.02 (NS)	
Mortality				
Omtzigt 1980; Secher 1986; Davies 1992; Duff 1993; **Ewigman 1993; French Doppler study, 1997;** McKenna 2003; Newnham 1993	n=32,962 (8 RCTs)	Fetal mortality	OR 1.11 (NS)	
Pastore 1999	n=689 (1 case-control study)	Fetal mortality	OR 1.36 (NS)	
Belfrage et al 1987; Fu et al 2000; Geerts 2004	n=17846 (3 cohorts)	Fetal mortality	OR 0.82 (NS)	
Omtzigt 1980; Secher et al 1986; Davies 1992; Duff 1993; **Ewigman 1993;** Newnham 1993; **French Doppler study, 1997;** McKenna 2003	n=30,942 (n=7 RCTs)	Neonatal mortality	OR 0.85 (NS)	
Omtzigt 1980; Eik-Nes 1984; Secher et al 1986; Waldenstrom 1988; Saari-Kemp 1990; Davies 1992; Duff 1993; **Ewigman 1993;** Geerts 1996; **French Doppler study, 1997;** Bakketeig 1984;	n=46,553 (13 RCTs)	Perinatal Mortality	OR 0.86 (NS)	
Sylvan 2005	n=209,726 (1 cohort)	Perinatal Mortality	OR 0.89 (NS)	
Apgar score and physical measurements				

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
Bakketeig 1984; Eik-Nes 1984; Secher 1986; Saari-Kemp 1990; Duff 1993; **Ewigman 1993; French Doppler study, 1997;** Kieler 1998; McKenna 2003	n=27,299 (10 RCTs)	Apgar at 1 min <7	OR 1.06 (NS)	
Omtzigt 1980; Bakketeig 1984; Eik-Nes 1984; Waldenstrom 1988; Davies 1992; **Ewigman et al 1993;** Kieler 1998; **Mason 1993;** French Doppler study, 1997; **Crowther 1999;** Duff 1993; McKenna 2003	n=22,150 (12 RCTs)	Apgar at 5 min <7	OR 0.91 (NS)	
Belfrage 1987; Sylvan 2005	n=213,138 (2 cohort studies)	Apgar at 5 min <7	OR 0.83 (CI, 0.76-0.92) Event rate (Routine US, selective US) (%): 0.9, 1 NNT: 562	
Saari-Kemp 1990; Bakketeig 1984; Eik-Nes 1984; Secher 1986; Waldenstrom 1988; **Ewigman 1993;** Geerts 1996; **Crowther 1999;** Newnham 2004	n=24271(9 RCTs)	Low birth wt (<2500 g)	OR 1.06 (NS)	
Smith 1984; Stark 1984; Lyons; Moore 1988; Fu et al 2000; Geerts 2004;	n= 18,622 (6 cohorts)	Low birth wt (<2500 g)	OR 1.11 (NS)	
Grisso 1994	n=12,546 (1 case control study)	Low birth wt (<2500 g)	OR 1.38 (CI 1.25-1.51) Event rate (Routine US, selective US) (%): 20, 20 NNT:22	Timing of US exposure revealed that the excess risk appeared to be restricted to the pts undergoing US during the late pregnancy and that it disappeared when only full-term, uncomplicated pregnancies were examined.
Geerts 1996; **Crowther 1999**	n=1509 (2 RCTs)	Very-low birth wt (<1500 g)	OR 1.26 (NS)	
Geerts 2004	n=2291 (1 cohort)	Very-low birth wt (<1500 g)	OR 1.24 (NS)	
Omtzigt 1980; Secher et al 1986; Duff 1993; **Ewigman 1993; Mason 1993;** Geerts 1996; **French Doppler study, 1997;** Kieler 1998; Newnham 2004	n=35,894 (9 RCTs)	Mean birth wt	OR 0.78 (NS)	

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
Smith 1984; Stark 1984; Geerts 2004; Belieni 2005	n= 2496(4 cohorts)	Mean birth wt	OR −8.20 (NS)	
Kieler 1998; Newnham 2004	n=7431 (2 RCTs)	Mean length at birth	OR −0.07 (NS)	
Kieler 1998; Newnham 2004	n=7393 (2 RCTs)	Mean head circumference at birth	OR −0.07 (NS)	
Omtzigt 1980; Wladimiroff 1980; Bakketeig 1984; Secher et al 1986; Waldenstrom 1988; Walenstrom 1992;Davies 1992; Duff 1993; Geerts 1996; Kieler; LeFevre; McKenna 2003; Newnham 2004	n=44,745 (13 RCTs)	SGA	OR 0.99 (NS)	
Cochlin 1984; Fu 2000; Geerts 2004; Sylvan 2005	n=225,190 (4 cohorts)	SGA	OR 1.04 (NS)	
Perinatal morbidity				
Omtzigt 1980; Secher et al 1986; Waldenstrom 1988; Davies 1992; Walenstrom 1992; **Ewigman 1993; French Doppler study, 1997;** Kieler 1998; Newnham	n=34,049 (10 RCTs)	Preterm birth	OR 0.99 (CI, 0.90-1.08)	
Fu 2000	n=12,143 (1 cohort study)	Preterm birth	OR 0.50 (CI, 0.40-0.63) Event rate (Routine US, selective US) (%): 2, 4 NNT:48	
Bakketeig 1984; Eik-Nes 1984; Davies 1992; Waldenstrom 1988; Duff 1993; **French Doppler study, 1997;** Mckenna 2003;	n=17,151(8 RCTs)	Neonatal resuscitation	OR 0.94 (NS)	
Omtzigt 1980; Bakketeig 1984; Secher 1986; Waldenstrom 1988; Saari-Kemp 1990; Davies 1992; Duff 1993; **Mason 1993; Ewigman; French Doppler study, 1997;** Geerts 1997; **Crowther 1999;** Mckenna 2003	n=33,408 (13 RCTs)	Neonatal special care unit	OR 1.00 (NS)	
Geerts 2004	n=2291 (1 cohort)	Neonatal special care unit	OR 1.33 (NS)	
Omtzizigt 1980; Waldenstrom 1988	n=6459 (2 RCTs)	Neonatal seizures	OR 0.69 (NS)	

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
Ewigman 1993	n=15,281 (2 RCTs)	Congenital malformations	OR 1.14 (NS)	
Tikkanen 1992	n=1160 (1 case-control study)	Congenital malformations	OR 1.40 (CI, 1.06-1.85)	
Hellman et al 1970; Fu et al.,2000	n=12,451 (2 cohort studies)	Congenital heart malformations	OR 1.80 (CI, 1.16-2.78) Event rate (Routine US, selective US) (%): 0.9, 5 NNT: 261	
Childhood cancers				
Bunin 1984; Cartwright 1984; Stalberg 2007	n=1909 (3 case-control studies)	Central nervous system neoplasms	OR 0.94 (NS)	
Cartwright 1984; Kinnier 1984; Shu 1994; Naumburg 2000; Shu 2002	n=6334 (5 case-control studies)	Leukemia	OR 0.97 (NS)	
Kinnier 1984; Sorahan 1995	n=3413 (2 case control studies)	Solid tumors	OR 0.96 (NS)	
Cartwright 1984; Sorahan 1995; Shu 2002	n=2594 (3 case-control studies)	Other tumors	OR 0.87 (NS)	
Sorahan 1995	n=424 (1 case-control study)	RES neoplasms	OR 1.03 (NS)	
Bunin 1984; Cartwright 1984; Kinnier 1984; Shu 1994; Sorahan 1995; Naumburg 2000; Shu 2002; Stalberg 2007	n=14,057 (8 case-control studies)	All malignancies	OR 0.94 (NS)	
Growth and neurological development during childhood				
Salvesen 1994	n=2016 (1 RCT)	Speech development: abnormal score on speech test	OR 0.70 (NS)	
Salvesen 1994; Kieler 1998	n=5347 (2 RCT)	Speech development: stuttering	OR 1.07 (NS)	
Salvesen 1994; Kieler 1998	n=5316(2 RCT)	Speech development: limited vocabulary	OR 0.98 (NS)	
Salvesen 1994; Kieler 1998	n=5193 (2 RCTs)	Speech development: delayed speech	OR 0.96 (NS)	
Campbell 1993	n=214(1 case-control study)	Speech development: delayed speech	OR 2.72 (CI 1.52-4.88) Event rate (Routine US, selective US) (%): 50, 23 NNT: 5	

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
Salvesen 1994	n=2040 (1 RCT)	Speech development: referral to speech therapist	OR 0.50 (CI 0.30-0.84) Event rate (Routine US, selective US) (%): 2, 4 NNT: 49	Other speech and language tests did not find any association between in utero exposure to US and speech and language development, it is plausible the findings of this RCT are by chance.
Salvesen 1992; Kieler 1992	n=5417 (2 RCTs)	Childhood visual problems: reduced vision	OR 0.82 (NS)	
Salvesen 1992; Kieler 1992	n=5331 (n=2 RCTs)	Childhood visual problems: use of glasses	OR 0.87 (NS)	
Salvesen 1992	n=1971 (1 RCT)	Childhood visual problems: loss of visual acuity on test	OR 0.84 (NS)	
Kieler 1997	n=3265 (1 RCT)	Childhood visual problems: referral to ophthalmologist	OR 0.82 (NS)	
Salvesen 1993; Newnham 2004	n=3779 (2 RCTs)	Impaired neurological development (Denver scale)	OR 1.01 (NS)	
Salvesen 1992	n=603 (2 RCTs)	Dyslexia in childhood	OR 0.75 (NS)	
Stark 1984	n=806 (1 cohort)	Dyslexia in childhood	OR 1.78 (NS)	
Salvesen 1992	n=1993 (1 RCT)	School performance: poor overall school performance	OR 0.93 (NS)	
Salvesen 1992	n=1993(1 RCT)	School performance: poor arithmetic scores	OR 1.00 (NS)	
Salvesen 1992	n=1981(1 RCT)	School performance: poor spelling	OR 0.94 (NS)	
Salvesen 1992	n=1984(1 RCT)	School performance: poor reading comprehension	OR 0.99 (NS)	
Salvesen 1992	n=1993(1 RCT)	School performance: poor oral reading	OR 0.98 (NS)	
Salvesen 1992	n=2100 (1 RCT)	Deficit in attention, motor control and perception	OR 0.83 (NS)	
Kieler 1998	n=3235 (1 RCT)	Late motor development	OR 1.05 (NS	
Salvesen 1993; Kieler 1998	n=4715 (2 RCTs)	Non–right-handedness for both genders	OR 1.13 (NS)	

Authors/Study Design (studies with quality rating 'A' are bolded; nonrandomized trials are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2]	Conclusions/Comments/ Limitations
Kieler 1998; Salvesen 1999	n=2422 (2 RCTs)	Non–right-handedness for boys only	OR 1.26 (CI 1.03-1.54) Events rate(Routine US, selective US) (%): 10, 9 NNT: 58	
Kieler et al, 2001; Kieler 2002	n=40,841 (2 cohorts)	Non–right-handedness for boys only	OR 1.17 (CI 1.07-1.27) Event rate (Routine US, selective US) (%): 10, 9 NNT: 58	
Subnormal intellectual performance or mental diseases after childhood				
Kieler 2005	n=167,059 (1 cohort study)	Subnominal intellectual performance among adult men	OR 1.19 (CI 1.12-1.27) Event rate (Routine US, selective US) (%): 23, 20 NNT: 34	There was a weak association between US screening and subnominal intellectual performance among adult men; lack of randomization
Stalberg 2007	n=370,945 (1 study)	Schizophrenia >12 years among men and women	OR 1.47 (NS)	
Stalberg 2007	n=370,945 (1 study)	Other psychoses >12 years among men and women	OR 1.03 (NS)	

APPENDIX VI. SYSTEMATIC REVIEW EVALUATING THE EFFECTIVENESS OF ULTRASOUND IN THE EMERGENCY DEPARTMENT

Systematic Review of Emergency Department Targeted Ultrasonography in the Evaluation of 1st-Trimester Pelvic Pain and Bleeding (McRae et al., 2009)

Key: CI, confidence interval; dx, diagnosis; ED, emergency department; EDTU, emergency department targeted US; grps, group(s); IOL, induction of labor; IUP, intrauterine pregnancy; LOS, length of stay; MA, meta analysis; SR, systematic review; US, ultrasound

Authors/Study Design (good-quality studies are bolded; studies *not* fully published are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2, 3]	Conclusions/Comments/ Limitations
McRae et al. (2009) SR of diagnostic accuracy and clinical utility of EDTU for detection of IUP in 1st trimester of pregnancy; MA of sensitivity and specificity (detection of IUP rules out ectopic pregnancy)	1778+ (8 studies; 1 study did not report sample size) *Study inclusion criteria:* Comparison of EDTU w/ US performed in radiology departments or by gynecology consultants for women presenting to ED w/ 1st trimester pain and bleeding; objective outcomes measured; published in full and abstracts *Study/pt characteristics:* Most of the studies analyzed only pts w/ ectopic pregnancy	*Outcomes:* Surgical rupture, ED LOS, time to dx, time to surgery. (Diagnostic accuracy and costs were also assessed, but that evidence is not presented here.)	All measured outcomes were significantly reduced by use of EDTU: *Mean time to dx:* By 139 min (39 pts, 1 study) *Mean time to surgery:* By 145-211 min (113 pts, 2 studies) *Mean ED LOS:* By 59-149 min (>1647, 5 studies) *Proportion of ectopic pregnancies in which rupture occurred during surgery:* From 21.4% to 2.5% *Authors' conclusions:* The evidence strengthens the argument for the routine use of EDTU in the evaluation of pelvic pain or bleeding in the 1st trimester of pregnancy.	EDTU in women w/ 1st trimester bleeding or pain may reduce the time required to diagnose and treat ectopic pregnancy, which may result in better surgical outcome. *Study limitations according to information supplied by McRae et al.:* Only 1 controlled trial, which was not published in full; other studies primarily retrospective; no fully published studies of mean time to dx or time to surgery; LOS results derived primarily from pts w/ IUP.
Mateer 1996 Retrospective chart review	n=340	Proportion of pts who found to have ectopic pregnancy and who ruptured during surgical exploration	Reduced from 21.4% to 2.5% ($P<0.05$).	Historical control grp and inadequate description of pt identification and chart abstraction. All pts erroneously discharged from the emergency department and later found to have ectopic pregnancy, were

Authors/Study Design (good-quality studies are bolded; studies *not* fully published are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2, 3]	Conclusions/Comments/ Limitations
				either diagnosed as having an abnormal IUP or had a nondiagnostic EDTU, which is against established guidelines.
Rodgerson 2001 Retrospective chart review of pts w/ ruptured ectopic pregnancy.	n=37	Mean time to dx Mean time to surgery	Reduced by 139 min (*P*<0.0001) Reduced by 211 min (*P*<0.0001).	Possible selection bias, between-grp baseline differences. Not applicable to populations that include IUP.
Blaivas and Bell 2000 Retrospective chart review of pts w/ ectopic pregnancy who require surgery	n=94	Mean time to surgery	Reduced by 145 min (CI, 1 hr 41 min to 3 hr 9 min)	Published in abstract form only; insufficient information to judge quality. Not applicable to populations that include IUP.
Pierce 2001 Controlled trial	n=29	ED LOS	Reduced by 2.1 hrs (172 min) (*P*<0.0001)	Published in abstract form only; insufficient information to judge quality.
Shih 1997 Prospective observational study of women w/ IUP	n=115 (Pts w/ ectopic pregnancy were excluded)	ED LOS Use of US depended on whether attending ED physician was credentialed for EDTU.	Reduced by 120 min (*P*<0.001)	Unclear whether grps were similar. Not applicable to populations that include ectopic pregnancy.
Blaivas 2000 Retrospective chart review of, women w/ IUP	n=1419	ED LOS Use of US depended on whether attending ED physician was credentialed for EDTU.	Reduced by 59 min (CI, 49 to 77) (*P*=0.0001)	Unclear whether grps were similar. Not applicable to populations that include ectopic pregnancy.
Burgher 1998 Retrospective chart review	n=84	ED LOS Use of US depended on whether attending ED physician was credentialed for EDTU.	Reduced by 70 min (*P*<0.0003)	Methods for pt identification and chart abstraction were not described.

Authors/Study Design (good-quality studies are bolded; studies *not* fully published are underlined)	Study Population	Methods/Treatment/ Outcome Measures	Results[1, 2, 3]	Conclusions/Comments/ Limitations
Jang and Aubin 2003 Retrospective chart review	NR	ED LOS	Reduced by 149 min (P<0.05)	Published in abstract form only; insufficient information to judge quality.

[1]CIs are 95%.