Health Technology Clinical Committee
Findings and Coverage Decision
Topic: **Routine Ultrasound**
Meeting Date: **October 22nd, 2010**
Final Adoption: **December 10th, 2010**

Number and Coverage Topic

20101122B – Routine Ultrasound

HTCC Coverage Determination

Routine Ultrasound is a **covered benefit with conditions** consistent with the criteria identified in the reimbursement determination.

HTCC Reimbursement Determination

❖ Limitations of Coverage

Routine Ultrasound is a covered benefit for pregnant women, routine screening ultrasound is a covered benefit, with the following conditions:

1. One Ultrasound in week 13 or earlier
2. One Ultrasound in weeks 16 thru 22
3. Other Ultrasound subject to agency determination

- Note: the committee acknowledged that optimal timing, clinically, for routine ultrasounds are in a narrower window (e.g. between weeks 11 and 13 for first trimester and between 18 and 20 weeks second trimester) but a narrower payment policy might unintentionally limit access when applied.

❖ Non-Covered Indicators

- N/A

❖ Agency Contact Information

Agency	Contact Phone Number
Labor and Industries	1-800-547-8367
Public Employees Health Plan	1-800-762-6004
Health and Recovery Services Administration	1-800-562-3022

Health Technology Background

The Routine Ultrasound topic was selected and published in December 2009 to undergo an evidence review process. The evidence based technology assessment report indicates that increasing use on number of US scans per pregnancy (1.5 US scans in 1995-1997; 2.7 US scans in 2005-2006), which increase depending on low or high risk. No difference in Medicaid versus private insurance. "Keepsake videos" not FDA approved. Key conclusions include DUS screening in high-risk pregnancies improves outcomes (v. low quality evidence) but routine US in low-risk pregnancies does not (moderate-high); routine US in 2nd trimester is safe; 2nd trimester US is most likely to detect multiple pregnancy and reduce IOL; gestational age at time of US does not affect perinatal mortality; preliminary evidence suggests potential cost savings with particular strategies of US in pregnancy and existing guidelines do not address the issue of clinical utility.

In August 2010, the HTA posted a draft and then followed with a final report from a contracted research organization that reviewed publicly submitted information; searched, summarized, and evaluated trials, articles, and other evidence about the topic. The comprehensive, public and peer reviewed Routine Ultrasound report is 116 pages, and identified a relatively large amount of literature.

An independent group of eleven clinicians who practice medicine locally meet in public to decide whether state agencies should pay for the health technology based on whether the evidence report and other presented information shows it is safe, effective and has value. The committee met on October 22nd, reviewed the report, including peer and public feedback, and heard public and agency comments. Meeting minutes detailing the discussion are available through the HTA program or online at http://www.hta.hca.wa.gov under the committee section.

Committee Findings

Having considered the evidence based technology assessment report and the written and oral comments, the committee identified the following key factors and health outcomes, and evidence related to those health outcomes and key factors:

1. Evidence availability and technology features

The committee concludes that the best available evidence on routine ultrasound has been collected and summarized. The evidence is presented below:

1.1 The evidence based technology assessment report indicates that ultrasound (US) is used in prenatal care for monitoring fetal development and maternal well being, including the important objective of preventing preterm birth.

1.2 Alternatives for screening and monitoring pregnancy include biophysical profile; pre-natal exams; ultrasound; fetal and umbilical Doppler ultrasound (DUS); utero-placental DUS; cardiotography; and fetal echocardiography.

1.3 The evidence based technology assessment report indicates that routine use of US in pregnancy is increasing (for example from an average of 1.5 exams in 1996 to 2.7 in 2006) ; the imaging provides a information about pregnancy status and fetal status and is considered reasonably safe; however evidence is very limited on the clinical utility of US and how the routine use of US impacts fetal or maternal health outcomes.

1.4 Despite low evidence of impact on health outcomes with routine screening, US is commonly used in united states:

- 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 can be 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.

1.5. A systematic and critical assessment of literature for evidence about the clinical utility of routine or screening use of US was gathered for use US in pregnant women for screening, guiding patient management as compared with no screening, screening with other methods, or concealment of US findings; along with descriptive information on US accuracy. Two meta-analysis: which included 11 RCTs on routine US in early pregnancy, and 8 RCTs on routine US in late pregnancy. Other studies relating to accuracy are also described.

2. Is the technology safe?

The committee discussed multiple key factors and health outcomes that were important for consideration in their overall decision on whether the technology is safe. Summary of committee considerations follows.

2.1 The evidence based technology assessment report includes moderate-quality evidence from a 2009 meta-analysis (Torloni et al., 2009) that included 41 studies, mostly RCT's; 3 additional RCTs and 4 observational studies. For major outcomes US shown 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.

2.2 The evidence based technology assessment report indicates that the FDA considers US to be a safe technology. The FDA does, however, considers "keepsake videos" to be an unapproved US.

2.3 The evidence based technology assessment report indicates that 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.

2.4 The evidence based technology assessment report indicates that 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. According to low-quality evidence, use of TVU to measure cervical length does not increase the risk of infection.

2.5 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.

- The evidence based technology assessment report indicated 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).

3. Is the technology effective?

The committee discussed multiple key factors and health outcomes that were important for consideration in their overall decision on whether the technology is effective. Summary of committee considerations follows.

3.1 *Accuracy*: the evidence based technology assessment report provided a descriptive review of accuracy. As a screening tool, US is often combined with other tests. However, review articles report sensitivities ranging from 40% to 99% depending on the target condition or use. Information on specificity, positive predictive value, and negative predictive value was not readily available.

3.2 *Overall Evidence, Routine/ Screening Use:* The evidence based technology assessment report indicated moderate- to high-quality evidence for 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.

3.3 The evidence based technology assessment report indicated 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.

3.4 The evidence based technology assessment report indicates 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.

3.5 *Gestational Age:* 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).

3.6 *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. 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).

3.7 *Multiple Gestation:* 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).

3.8 Attempts to assess differential effects according to multiple versus singleton pregnancy and maternal 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).

4. Special Populations?

4.1 The evidence based technology reported indicated 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.

4.2 *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 while routine US performed in late pregnancy is not. 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 pregnancies by 24 weeks to 26 weeks was greatly reduced when US was performed after 14 weeks, 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).

4.3 *Other Patient Characteristics or Evidence-Based Patient Selection Criteria:* Alfirevic et al. (2010) conducted subgroup analyses according to singleton versus multiple births and according to five risk factors (small for gestational age or IUGR, hypertension/preeclampsia, diabetes, prolonged pregnancy, and previous pregnancy loss).

4.4 *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.

4.5 *Provider Type, Setting, or Other Provider Characteristics*: A systematic review has shown that emergency department targeted ultrasonography (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). 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.

5. Is the technology cost-effective?

The committee discussed multiple key factors that were important for consideration in their overall decision on whether the technology has value and is cost-effective. Summary of committee considerations follows.

1.1. The evidence based technology report included 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.

1.2. The evidence based technology assessment report indicated 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.

1.3. Washington state agency utilization and cost information indicated costs of US of $20M for the past four years (average of $5M per year and per treatment costs averaging $336).

6. Medicare Decision and Expert Treatment Guidelines

Committee reviewed and discussed the expert guidelines as identified and reported in the technology assessment report.

6.1 Centers for Medicare and Medicaid Services currently cover ultrasounds. Ultrasound diagnostic procedures are listed on their website and are divided into two categories. Medicare coverage is extended to the procedures listed in Category I. Periodic claims review by the intermediary's medical consultants should be conducted to ensure that the techniques are medically appropriate and the general indications specified in these categories are met. Techniques in Category II are considered experimental and should not be covered at this time.

6.2 Guidelines – a search of the core sources and relevant specialty groups identified four guidelines for US (American Institute of Ultrasound Medicine, 2007; American College of Obstetricians and Gynecologists, 2009; American College of Radiologists, 2208-2009; and the Institute of Clinical Systems Improvement, 2010).

- 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 considers evidence pertaining to the long-term effects on child growth and development, differential effectiveness and safety, or cost-effectiveness.

Committee Decision

Based on the deliberations of key health outcomes, the committee decided that it had the most complete information: a comprehensive and current evidence report, public comments, and agency and state utilization information. The committee concluded that the current evidence on Routine Ultrasound demonstrates that there is sufficient evidence to cover with conditions the use of Routine Ultrasound. The committee considered all the evidence and gave greatest weight to the evidence it determined, based on objective factors, to be the most valid and reliable.

Based on these findings, the committee voted 8 to 0 to cover with conditions Routine Ultrasound.

Routine Ultrasound is a covered benefit for pregnant women, routine screening ultrasound is a covered benefit, with the following conditions:

4. One Ultrasound in week 13 or earlier
5. One Ultrasound in weeks 16 thru 22
6. Other Ultrasound subject to agency determination

Note: the committee acknowledged that optimal timing, clinically, for routine ultrasounds are in a narrower window (e.g. between weeks 11 and 13 for first trimester and between 18 and 20 weeks second trimester) but a narrower payment policy might unintentionally limit access when applied.

Health Technology Clinical Committee Authority

Washington State's legislature believes it is important to use a scientific based, clinician centered approach for difficult and important health care benefit decisions. Pursuant to chapter 70.14 RCW, the legislature has directed the Washington State Health Care Authority, through its Health Technology Assessment program to engage in a process for evaluation process that gathers and assesses the quality of the latest medical evidence using a scientific research company and takes public input at all stages. Pursuant to RCW 70.14.110 a Health Technology Clinical Committee (HTCC) composed of eleven independent health care professionals reviews all the information and renders a decision at an open public meeting. The Washington State Health Technology Clinical Committee (HTCC) determines how selected health technologies are covered by several state agencies (RCW 70.14.080-140). These technologies may include medical or surgical devices and procedures, medical equipment, and diagnostic tests. HTCC bases their decisions on evidence of the technology's safety, efficacy, and cost effectiveness. Participating state agencies are required to comply with the decisions of the HTCC. HTCC decisions may be re-reviewed at the determination of the HCA Administrator.