PROTOCOL

DIAGNOSTIC ACCURACY AND OUTCOMES OF ULTRASOUND IN THE FIRST TRIMESTER OF PREGNANCY FOR DETECTION OF COMPLICATIONS RELEVANT FOR AUSTRIAN POPULATION, EXCLUSIVE OF SCREENING FOR DOWN SYNDROME

Contents

1 Aim

The Austrian Pregnancy Screening includes two ultrasound examinations, one in the second and one in the third trimester.

Obstetric experts in Austria recommend a third screening ultrasound examination in the first trimester of pregnancy. Decision making about this involves the Austrian Ministry of Health (Bundesministerium für Gesundheit und Frauen), the highest medical consultant gremium (Oberster Sanitätsrat) and the Austrian Social Insurance (Österreichische Sozialversicherung).

In Austrian Social Insurance the department for Evidence Based Health Care decided to evaluate the medical indication for a third screening ultrasound in the first pregnancy trimester by a scientific review.

In Austria, the Ministry of Health and the Social Insurance are responsible for reimbursement of screening in pregnancy. Additional ultrasound investigations in high risk pregnancies are usually provided in obstetrical practices and are reimbursed by the Social Insurance.

The result of the planned review should provide the scientific basis for deciding whether or not to include a third ultrasound test into the Austrian screening program for pregnant women. The decision will be by the Ministry of Health as a legal enactment. The implementation of a third ultrasound examination will increase the costs for the screening programme in general. It has been suggested that some of the additional ultrasound examinations currently done in first trimester on indication will be included in the screening programme after implementation.

The ultrasound examination in the first trimester of pregnancy has to fulfil the usual criteria for implementation of screening programmes. This review will evaluate the medical indication for ultrasound screening in the first trimester, which means

- how accurate is ultrasound screening for endpoints described in paragraph 5
- which added value of the ultrasound screening in first trimester can be expected versus the ultrasound examination in second and third trimester for detection of endpoints described in paragraph 5

Consequences of the decision to implement an additional screening ultrasound also include discussions and agreement about standards for informing parents about risks, therapeutic options or options for termination of pregnancy after an abnormal diagnostic result. This has become particularly important after a High Court decision about the claim to maintenance for a Down syndrome baby against an ultrasound examiner during pregnancy[1].

[1] OGH Urteil vom 7.3.2006, Geschäftzahl 50b165/05h; Dokumentnummer JJT/20060307/OGH0002/0050OB00165/05H0000/000

2 Background

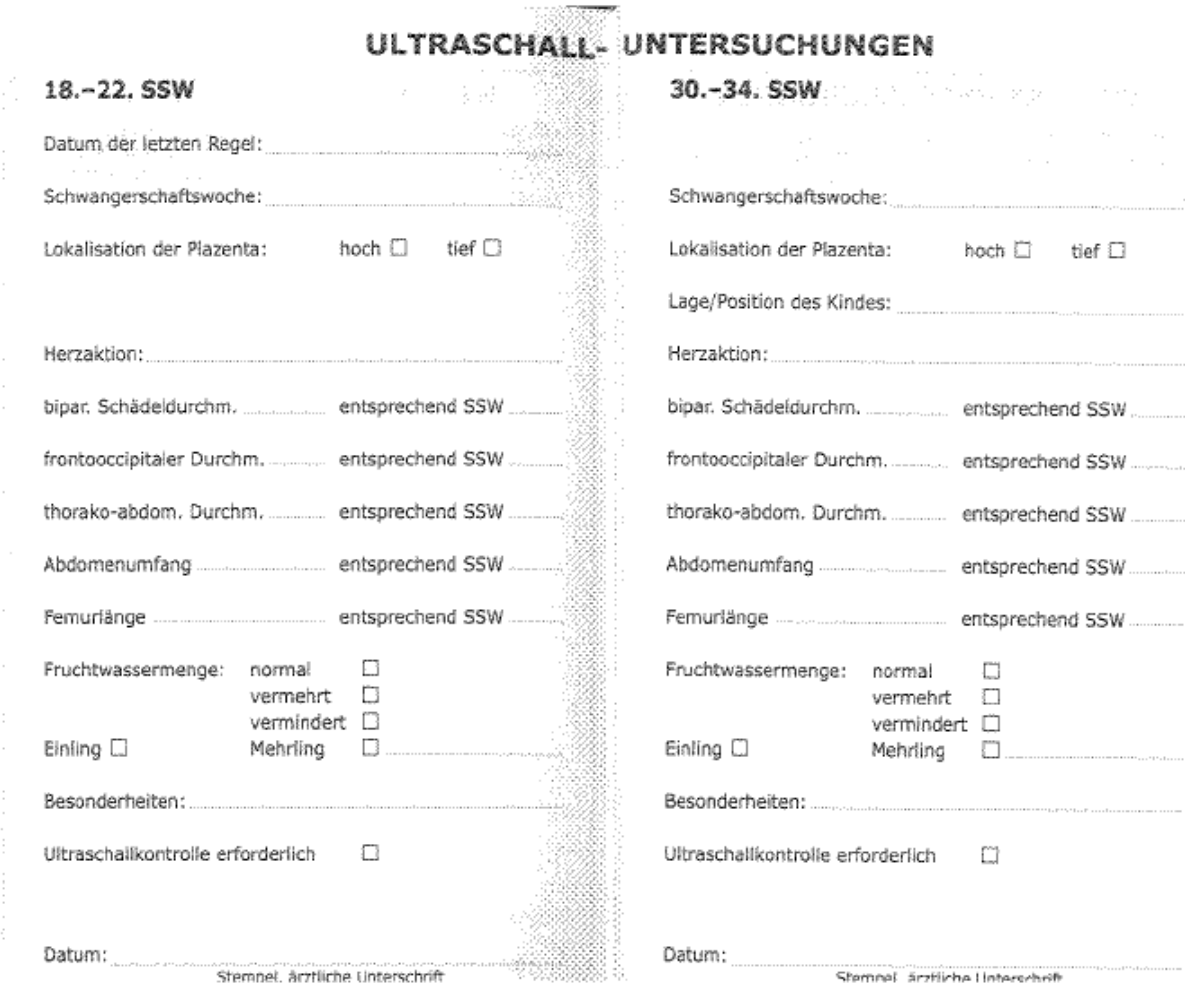
Mutter-Kind-Pass (mother and child) examinations currently only include ultrasound examinations in the second and third trimester of pregnancy. The benefit of including an additional ultrasound screening in the first trimester is unknown. A systematic review should provide information about accuracy and outcomes of ultrasound in this early period of pregnancy (until and including the 12th week) for the most relevant pregnancy complications in Austria.

An additional ultrasound examination in first pregnancy trimester is currently often done along with the *Combined Test (Triple test)* which adds four probabilities (nuchal translucency, two biomarkers and maternal age) for calculating the probability of Down syndrome of the foetus. This test – if positive – needs a more precise additional chromosomal test like chorion villus sampling (CVS) or amniocentesis. With the current policy it is only possible to select women with high risk for Down syndrome to undergo an amniocentesis or CVS; ultrasound screening is controversial because of its only consequence of abortion. The decision for abortion would be shifted from the pregnant mother towards the health system if a general screening for Down syndrome is implemented.

Whether an additional ultrasound screening in first trimester of pregnancy is of medical relevance should be scientifically evaluated in the planned review by addressing test accuracy and whether outcomes show an additional benefit for mother and child, because high costs for the health system will be involved.

The review about diagnostic accuracy of ultrasound in first trimester of pregnancy is prepared by the Team of Evidence Based Health Care of Hauptverband der österreichischen Soziaversicherungsträger in cooperation with Kleijnen Systematic Reviews Ltd.

3 Current Austrian set of ultrasound examinations in pregnancy



ULTRASCHALL- UNTERSUCHUNGEN

18.–22. SSW

Datum der letzten Regel: .....

Schwangerschaftswoche: .....

Lokalisation der Plazenta: hoch ☐ tief ☐

Herzaktion: .....

bipar. Schädeldurchm. ..... entsprechend SSW .....

frontooccipitaler Durchm. ..... entsprechend SSW .....

thorako-abdom. Durchm. ..... entsprechend SSW .....

Abdomenumfang ..... entsprechend SSW .....

Femurlänge ..... entsprechend SSW .....

Fruchtwassermenge: normal ☐ vermehrt ☐ vermindert ☐

Einling ☐ Mehrling ☐

Besonderheiten: .....

Ultraschallkontrolle erforderlich ☐

Datum: .....

Stempel, ärztliche Unterschrift

30.–34. SSW

Schwangerschaftswoche: .....

Lokalisation der Plazenta: hoch ☐ tief ☐

Lage/Position des Kindes: .....

Herzaktion: .....

bipar. Schädeldurchm. ..... entsprechend SSW .....

frontooccipitaler Durchm. ..... entsprechend SSW .....

thorako-abdom. Durchm. ..... entsprechend SSW .....

Abdomenumfang ..... entsprechend SSW .....

Femurlänge ..... entsprechend SSW .....

Fruchtwassermenge: normal ☐ vermehrt ☐ vermindert ☐

Einling ☐ Mehrling ☐

Besonderheiten: .....

Ultraschallkontrolle erforderlich ☐

Datum: .....

Stempel, ärztliche Unterschrift

4 Objectives

After consideration of the PICO questions based on the health insurance data and the international literature the Peer Group (Dr. Gottfried Endel, Dr. Irmgard Schiller-Frühwirth, Mag. Ingrid Wilbacher) decided after discussions that the following questions are the most relevant ones to be answered in the planned review:

Determination of the accuracy of ultrasound investigation in the first pregnancy trimester (incl. 12th week) in diagnosing the following disorders:

- Other chromosomal anomalies **exclusive of** Down Syndrom (Chimäre 46,XX/46,XY, Chimäre 46,XX/46,XY with Hermaphroditismus verus, Hermaphroditismus verus with Karyotype 46,XX, Gonadendysgenesia, 46,XX with Streak-Gonades, 46,XY with Streak-Gonades, Fragile X-Chromosome, Syndrom of fragile X-Chromosome, ICD 10 Q99)
- accuracy of detection of chorionicity with ultrasound in the first trimester of pregnancy, gold standard is membrane check after delivery
- Increased risk of preterm birth (ICD 10 P 07)
- Gestational diabetes
- Determination of gestational age

Determination of the outcomes after ultrasound investigation in the first pregnancy trimester (incl. 12th week) versus ultrasound investigation in the second and/or third trimester for the following target disorders:

- Other chromosomal anomalies **exclusive of** Down Syndrom (Chimäre 46,XX/46,XY, Chimäre 46,XX/46,XY with Hermaphroditismus verus, Hermaphroditismus verus with Karyotyp 46,XX, Gonadendysgenesia, 46,XX with Streak-Gonades, 46,XY with Streak-Gonades, Fragile X-Chromosome, Syndrom of fragile X-Chromosome, ICD 10 Q99)
- accuracy of detection of chorionicity with ultrasound in the first trimester of pregnancy, gold standard is membrane check after delivery
- Increased risk of preterm birth (ICD 10 P 07)
- Gestational diabetes
- Determination of gestational age

5 Criteria for including and excluding studies

These criteria will be based on the above questions. Diagnostic accuracy studies will be included if they allow generation of 2 by 2 tables of ultrasound findings compared to a reference standard. For the reference standard we will only accept assessments according to definitions of the actual outcome, not other tests predicting such an outcome. For assessing gestational age studies using any reference standard will be included. For studies assessing outcomes of screening in the first trimester compared to later screening, we will include randomised trials and controlled observational studies with parallel control groups.

5.1 Criteria for including studies

- accuracy studies
- studies that contain early screening vs. screening at a later date
- screening population
- scan in the first trimester, transvaginal + abdominal
- date of publication as of 1.1.1996
- comparison of screening with confirmation of the findings post partum/post abortum/post Amniocentesis (CVS)

5.2 Criteria for excluding studies

- Doppler and Echocardiography – happens only if there is a special indication or the technical conditions are given
- Down Syndrome – reference to a good review
- risk population
- combination with biochemical markers (question: affiliation of a third scan into the mother-child-booklet)
- studies where animal experiments were involved
- no scan in the first trimester

5.3 Miscellaneous

Regarding the person who conducts the examination – inclusion of all, but separately (GP, Gyn, US, technician) + description of the people who conduct examinations

Softmarker: abnormality without any evidence of events things inside the womb

Hardmarker: abnormality with evidence of unusual events inside the womb

6 Methods and Quality Assessment

Assessing relevance and inclusion: Studies will be screened for relevance independently by two reviewers, disagreements will be resolved by consensus. Studies which appear potentially relevant will be ordered and assessed for inclusion by one reviewer and checked by a second.

Quality assessment: quality assessment of accuracy studies will take place using the QUADAS instrument, adopted as appropriate. For example, questions about the reference standard where this is verifiable and indisputable (such as in objectives 1 to 4), can be ignored. Quality assessment for randomised trials and controlled observational studies will take place using the appropriate checklists from CRD Report 4. Quality assessment will be used for descriptive purposes of general study quality, and were possible as items in meta-regression analysis in order to investigate the influence of study quality on the estimates of diagnostic accuracy. Quality assessment forms will be developed using Microsoft Access or Excel. Separate forms will be developed for the different study designs included in the review. Quality assessment will be carried out by one reviewer and checked by a second.

The QUADAS tool[2]

Item		Yes	No	Unclear
1.	Was the spectrum of patients representative of the patients who will receive the test in practice? (= Standard)	(X)	( )	( )
2.	Were selection criteria clearly described? (Standard)	(X)	( )	( )
3.	Is the reference standard likely to correctly classify the target condition? (Goldstandard)	( )	( )	( )
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	( )
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis? (ITT)	( )	( )	( )
6.	Did patients receive the same reference standard regardless of the index test result?	( )	( ) different reference standard (with invasive Interventions)	( )
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	( )
8.	Was the execution of the index test described in sufficient detail to permit replication of the test? Subcategories: what exactly was screened. Examinor. Transvaginal/ tranabdominal	( )	( )	( )
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	( )
10.	Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	( )
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	( )
13.	Were uninterpretable/ intermediate test results reported?	( )	( )	( )
14.	Were withdrawals from the study explained?	( )	( )	( )

[2] http://www.biomedcentral.com/1471-2288/3/25

Box 5.8
Quality criteria for assessment of experimental studies

1. Was the assignment to the treatment groups really random (Screening yes/no)?
   Adequate approaches to sequence generation
   - o Computer-generated random numbers
   - o Random numbers tables

   Inadequate approaches to sequence generation
   - o Use of alternation, case record numbers, birth dates or week days

2. Was the treatment allocation concealed?
   Adequate approaches to concealment of randomisation
   - o Centralised or pharmacy-controlled randomisation
   - o Serially-numbered identical containers
   - o On-site computer based system with a randomisation sequence that is not readable until allocation
   - o Other approaches with robust methods to prevent foreknowledge of the allocation sequence to clinicians and patients

   Inadequate approaches to concealment of randomisation
   - o Use of alternation, case record numbers, birth dates or week days
   - o Open random numbers lists
   - o Serially numbered envelopes (even sealed opaque envelopes can be subject to manipulation)

3. Were the groups similar at baseline in terms of prognostic factors?
4. Were the eligibility criteria specified?
5. Were outcome assessors blinded to the treatment allocation?
6. Was the care provider blinded?
7. Was the patient blinded?
8. Were the point estimates and measure of variability presented for the primary outcome measure?
9. Did the analyses include an intention to treat analysis

Box 5.9[3]
Some quality criteria for assessment of observational studies
Cohort studies

- Is there sufficient description of the groups and the distribution of prognostic factors?
- Are the groups recruited at a similar time in their pregnancy?
- Is the intervention/treatment sufficiently described?
- Were the groups comparable on all important confounding factors?
- Was there adequate adjustment for the effects of these confounding variables?
- Was outcome assessment blind to ultrasound test?
- Was follow-up long enough for the outcomes to occur?
- What proportion of the cohort was followed-up?

[3] http://www.york.ac.uk/inst/crd/pdf/crd4_ph5.pdf

- Were drop-out rates and reasons for drop-out similar across intervention and unexposed groups?

Case-control studies

- Is the case definition explicit?
- Has the outcome definition of the cases been reliably assessed and validated?
- Were the controls randomly selected from the source of population of the cases?
- How comparable are the cases and controls with respect to potential confounding factors?
- Was ultrasound status assessed in the same way for cases and controls?
- How was the response rate defined? (Selection of study participants after question)
- Were the non-response rates and reasons for non-response the same in both groups?
- Is it possible that over-matching has occurred in that cases and controls were matched on factors related to exposure?
- Was an appropriate statistical analysis used (matched or unmatched)?

7 Literature Search

The following databases will be searched: Medline, Embase, DARE, HTA, Cinahl, Lilacs and the National Research Register from 1996 to current time. Furthermore, references in retrieved articles and systematic reviews will be checked, experts will be contacted, and the internet will be searched via general search engines such as Google for relevant studies. Identified references will be downloaded in Reference Manager software for further assessment and handling. The search strategies will be developed specifically for each database, and are available in the appendix.

7.1 Medline Outcomes

#	Search history	Results
1	exp Pregnancy Trimester, First/	9170
2	exp Ultrasonography, Prenatal/	16162
3	exp Clinical Trials/	193918
4	exp Research Design/	216276
5	exp Treatment Outcome/	291062
6	exp Double-Blind Method/	90532
7	exp Single-Blind Method/	10558
8	((single or double or triple) adj3 blind$3).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	120481
9	random$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	486959
10	controlled clinical trial.pt.	74768
11	clinical trial.pt.	455937
12	(clinical adj trial$1).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	586148
13	exp Epidemiologic Research Design/	472717
14	(control$3 adj trial$1).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	292591
15	randomi#ed controlled trial.pt.	233178
16	comparative study/	1343564
17	or/3-16	2381580
18	1 and 2 and 17	390
19	limit 18 to humans	390
20	limit 19 to yr="1996 - 2006"	334

7.2 Medline Accuracy

#	Search History	Results
1	exp Pregnancy Trimester, First/	9160
2	exp Ultrasonography, Prenatal/	16148
3	exp "Sensitivity and Specificity"/	216138
4	exp Diagnosis/	3914349
5	diagnos$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	1206136
6	sensitiv$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	732282
7	predict$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	471803
8	accura$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	229300
9	or/3-8	5076163
10	1 and 2 and 9	1257
11	limit 10 to humans	1256
12	limit 11 to yr="1996 - 2006"	987

7.3 Embase Outcomes

#	Search History	Results
1	exp Pregnancy Trimester, First/	7387
2	exp Ultrasonography, Prenatal/	191297
3	exp Clinical Trials/	402678
4	exp Research Design/	1057970
5	exp Treatment Outcome/	337526
6	exp Double-Blind Method/	61061
7	exp Single-Blind Method/	6068
8	((single or double or triple) adj3 blind$3).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	100919
9	random$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	343722

10	controlled clinical trial.pt.	0
11	clinical trial.pt.	0
12	(clinical adj trial$1).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	447864
13	exp Epidemiologic Research Design/	622451
14	(control$3 adj trial$1).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	135023
15	randomi#ed controlled trial.pt.	0
16	comparative study/	81848
17	or/3-16	2141891
18	1 and 2 and 17	570
19	limit 18 to humans	565
20	limit 19 to yr="1996 - 2006"	487

7.4 Embase Accuracy

#	Search History	Results
1	exp Pregnancy Trimester, First/	7387
2	exp Ultrasonography, Prenatal/	191297
3	exp "Sensitivity and Specificity"/	28907
4	exp Diagnosis/	1709819
5	diagnos$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	1333466
6	sensitiv$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	573444
7	predict$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	404802
8	accura$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	286509
9	or/3-8	2798767
10	1 and 2 and 9	1727

11	limit 10 to humans	1699
12	limit 11 to yr="1996 - 2006"	1394

7.5 Cinahl Outcomes

#	Search History	Results
1	exp Pregnancy Trimester, First/	424
2	exp Ultrasonography, Prenatal/	1144
3	exp Clinical Trials/	38903
4	exp Research Design/	174536
5	exp Treatment Outcome/	30633
6	exp Double-Blind Method/	0
7	exp Single-Blind Method/	0
8	((single or double or triple) adj3 blind$3).mp. [mp=title, subject heading word, abstract, instrumentation]	11110
9	random$.mp. [mp=title, subject heading word, abstract, instrumentation]	50991
10	controlled clinical trial.pt.	0
11	clinical trial.pt.	18196
12	(clinical adj trial$1).mp. [mp=title, subject heading word, abstract, instrumentation]	34432
13	exp Epidemiologic Research Design/	0
14	(control$3 adj trial$1).mp. [mp=title, subject heading word, abstract, instrumentation]	11503
15	randomi#ed controlled trial.pt.	0
16	comparative study/	35719
17	or/3-16	220517
18	1 and 2 and 17	35
19	limit 18 to humans [Limit not valid in: CINAHL; records were retained]	35
20	limit 19 to yr="1996 - 2006"	35

7.6 LILACS Accuracy

#	Search History	Results
1	("gravidez/" and primeiro trimestre) or ("embarazo/" and primero trimestre) or "pregnancy tests" or "pregnancy trimester, first/" or "pregnancy, first trimester/"	139
2	"ultrasonografia fetal/" or "ultrasonografia pre-natal/" or "ultrasonografia prenatal/" or "ultrasonography, fetal/" or "ultrasonography, prenatal/"	346
3	"diagnosis" or "diagnosis, prenatal" or "diagnosis, prenatal/" or "diagnostico intra-uterino/" or "diagnostico intrauterino/" or "diagnostico por imagem/" or "diagnostico por ultra-som/" or "diagnostico por ultrasonido/" or "diagnostico pre-natal" or "diagnostico pre-natal por ultra-som/" or "diagnostico pre-natal ultra-sonico/" or "diagnostico pre-natal/"	62947
4	1 and 2 and 3	17

7.7 Cinahl Accuracy

#	Search History	Results
1	exp Pregnancy Trimester, First/	424
2	exp Ultrasonography, Prenatal/	1144
3	exp "Sensitivity and Specificity"/	8634
4	exp Diagnosis/	236756
5	diagnos$.mp. [mp=title, subject heading word, abstract, instrumentation]	76030
6	sensitiv$.mp. [mp=title, subject heading word, abstract,	24173

	instrumentation]	
7	predict$.mp. [mp=title, subject heading word, abstract, instrumentation]	34556
8	accura$.mp. [mp=title, subject heading word, abstract, instrumentation]	13889
9	or/3-8	293881
10	1 and 2 and 9	108
11	limit 10 to humans [Limit not valid in: CINAHL; records were retained]	108
12	limit 11 to yr="1996 - 2006"	106

8 Data management

8.1 Data extraction

Data extraction forms will be developed using Microsoft Access or Excel, these will be piloted independently on a small selection of studies and adjusted as necessary. Studies will be data extracted by one reviewer and checked by a second. The following information will be extracted for all studies: study details (identifier, aim, study design, location, setting). In addition, data will be extracted on test details (test evaluated, gold standard, details of test performance, at which gestational age, methods, time between tests), participant details (number of participants, number of imaging tests performed, age, sex, inclusion criteria) and results (data to construct 2 x 2 table).

8.2 Analysis

Tests will be grouped further according to the specific tests or test combinations reported in the literature. If combinations of tests have been evaluated these will be analysed as a test combination. Studies reporting similar combinations of tests will be grouped together. Tests will be grouped according to when these were performed and results of tests performed before and after a gestational age of 12 weeks will be compared.

For each test, or combination of tests, the range in sensitivity, specificity and likelihood ratios (of both positive and negative tests results) will be calculated and discussed, together with possible ranges in positive and negative predictive values which will be calculated based on a number of different estimates of disease prevalence. Diagnostic odds ratios (DOR) will be calculated. These have the advantage of being a single indicator of diagnostic accuracy in contrast to most of the other measures, which have to be judged in pairs. The DOR can take values between 0 and infinity, with high values indicating good test performance. It is calculated as the Odds(Sensitivity)/Odds(1-specificity), ad/bc in a 2 x 2 table.

Heterogeneity of the sensitivity, specificity, likelihood ratios and DOR will be investigated using the Q statistic and through visual examination of Galbraith plots of study results. This involves plotting the log DOR (D) divided by its standard error against the inverse of the standard error. D can be plotted as a regression line through the origin, with lines 2 standard errors either side representing the 95% confidence level boundaries around D, points lying outside these lines are suggestive of heterogeneity in study results. If studies are homogenous in terms of sensitivity and specificity then the pooled sensitivity, specificity and likelihood ratios will be calculated using a random effects model. If either one of these measures shows evidence of heterogeneity then further analyses will be conducted using D. If study

homogeneity cannot be rejected, D will be pooled using a random effects model to calculate a summary receiver operating characteristic (sROC) curve, separately for each separate study type. Where there is evidence of heterogeneity this will be investigated further, if possible depending on the amount of data, using random-effects meta-regression analysis.

Outcomes based on studies with our inclusion criteria	
Participants characteristics	• Gestational age when the ultrasound was performed • Maternal: Ethnicity, Weight, age • Gestational age measured by date of LMP [last menstrual period] or CRL [crown-lump length measurement] or BPD [biparietal diameter] • Gestation conceived spontaneously or by IVF • Single or multiple pregnancies • Others?
Studies characteristics	• Retrospective/prospective study with 2x2 table • Diagnosis confirmed by pathology/autopsy • Karyotype performed • % of participants with high risk (age >35, maternal diseases, previous malformation) described • Dubious exam confirmed by expert • US performed on the 2nd trimester as well • Dichotomous/continuous data
Type of intervention	• Transvaginal ultrasound (details) • Transabdominal ultrasound (details) • Diagnostic test performed by: doctors, radiologist, nurse, midwife, others • Years of experience with performing US • Where radiologists/others trained specifically to perform the 1st trimester screening?

9 Timeplan

	Month 1				Month 2				Month 3				Month 4				Month 5			
Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Detailed protocol																				
Literature Search																				
Consolidation Lit																				
Define to include																				
Milestone Search			x																	
Organise full text																				
Critical appraisal																				
Consolidation Appr																				
Milestone Text													x							
Summarization																				
Report																				
Prepare to publish																				
Endreport																				x