Accuracy of ultrasound scanning relative to reference tests for prenatal diagnosis of microcephaly in the context of Zika virus infection: a systematic review of diagnostic test accuracy

Ezinne C Chibueze, a Alex JQ Parsons, b Katharina da Silva Lopes, a Takemoto Yo, a Toshiyuki Swa, c Chie Nagata, d Nobuyuki Horita, e Naho Morisaki, f Olukunmi O Balogun, a Amarjargal Dagvadorj, g Erika Ota, h Rintaro Mori, a Olufemi T Oladapo i

a Department of Health Policy, National Center for Child Health and Development, 2-10-1 Okura, Setagaya ku, Tokyo, 157-853 Japan
b Department of Global Health Policy, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
c Graduate School of Human Sciences, Osaka University, Osaka, Japan
d Department of Education for Clinical Research, National Center for Child Health and Development, Tokyo, Japan
e Pulmonology, Graduate School of Medicine, Yokohama City University, Yokohama, Japan
f Department of Social Medicine, National Center for Child Health and Development, Tokyo, Japan
g Department of Health Informatics, Kyoto University, Yoshida Konoe-cho, Syako-ku, Kyoto, Japan
h St. Luke’s International University, Graduate school of Nursing, Global Health Nursing, Tokyo, Japan

Correspondence to Ezinne Chibueze (email: ezinne.chibueze@gmail.com).

(Submitted: 24 May 2016 – Published online: 25 May 2016)
Abstract

**Objective:** To assess the accuracy of ultrasound measurements of fetal biometric parameters for prenatal diagnosis of microcephaly in the context of Zika virus (ZIKV) infection.

**Methods:** We searched MEDLINE, EMBASE, CINAHL, CDTA and other bibliographic databases for studies published until 3rd March 2016. We extracted the numbers of true positives, false positives, true negatives and false negatives and performed meta-analysis to estimate group sensitivity and specificity, where possible. Predictive values for ZIKV-infected pregnancies were extrapolated from those obtained for pregnancies unrelated to ZIKV.

**Results:** Out of 111 eligible full texts, nine studies met our inclusion criteria. Pooled estimates from two studies showed that at 3, 4 and 5 standard deviations (SDs) < mean, sensitivities were 84%, 68% and 58% for head circumference (HC); 76%, 58% and 58% for occipito-frontal diameter (OFD); and 94%, 85% and 59% for biparietal diameter (BPD). Specificities at 3, 4 and 5 SDs below the mean were 70%, 91% and 97% for HC; 84%, 97% and 97% for OFD; and 16%, 46% and 80% for BPD. No study including ZIKV-infected pregnant women was identified. Extrapolated positive predictive values for ZIKV-infected pregnancies at 3, 4 and 5 SD < mean were 2.6%, 6.7% and 15.6% for HC; 4.3%, 15.6% and 15.6% for OFD; and 1.0%, 1.4% and 2.8% for BPD, respectively. Negative predictive values at 3, 4 and 5 SDs < mean were over 99% for HC, OFD, and BPD. Five out of seven studies with descriptive data reported high rates of false positives when ultrasound is applied for prenatal diagnosis of microcephaly.

**Conclusion:** Prenatal ultrasound appears more accurate in detecting the absence of microcephaly than its presence.

**Keywords:** Zika virus, microcephaly, ultrasound, DTA, diagnostic test accuracy, congenital abnormality, congenital infections
Introduction

Microcephaly is a sign of fetal brain abnormality in which there is a significantly small head size for gestational age and sex. Infants born with microcephaly are likely to present with variable clinical features ranging from subtle impairment in neurological development to serious intellectual disabilities on the longer term. It is generally a rare condition occurring in 5.8 to 18.7 per 100,000 pregnancies and often arising from a wide variety of conditions that can cause abnormal brain growth (1). In 2015, a 20-fold increase in neonatal microcephaly was observed in association with Zika virus (ZIKV) infections in pregnant women in Latin America (2). This observation prompted the World Health Organization (WHO) to declare the ZIKV outbreak in the Americas a *Public Health Emergency of International Concern* on 1st February 2016 (3). As part of its strategic framework, WHO is providing normative guidance to affected countries on conditions presumably associated with prenatal ZIKV infection with the aim of improving surveillance and clinical outcomes in at-risk populations. The WHO interim guidance recommends that pregnant women residing in areas of ongoing ZIKV transmission should have fetal ultrasound scans to exclude microcephaly or other brain abnormalities that have been reported in fetuses of women with prenatal ZIKV infection (4).

Prenatal assessment of microcephaly has conventionally relied on ultrasound measurements of fetal biometric parameters such as the head circumference, biparietal diameter, and occipito-frontal diameter (5-7). The measurements of these parameters below a given threshold at the specific gestational age of assessment have been applied to diagnose fetal microcephaly (6, 8). However, there is currently no international consensus on the fetal biometric parameter or the threshold to confidently diagnose microcephaly *in utero*. In addition, as a result of the rarity of this condition and application of different parameters and thresholds, the risk of wrong or missed diagnosis is high (9-12).

In the context of ZIKV infection, an accurate prenatal diagnosis of microcephaly is critical for fetal prognosis and decision-making by health providers and families of women suspected or confirmed to have ZIKV infection. We conducted a systematic review to assess the diagnostic accuracy of ultrasound measurement of fetal biometric parameters compared to reference assessments at birth for prenatal diagnosis of microcephaly in the context of ZIKV infection. This review served as the evidence base for the revised WHO interim guidance on prenatal assessment of microcephaly in the context of ZIKV infection.
Methodology

Search strategies

We searched MEDLINE, EMBASE, CENTRAL, Cochrane Database for DTA studies, LILACS, and WHO Global Health Library for studies published until 3rd March 2016. Search terms related to the index tests, reference tests and target condition were employed in the search strategies as shown in Appendix 1. Searches for grey literature and bibliographies of existing systematic reviews on ultrasound in pregnancy were complemented with results of the search strategies. No restrictions were placed on search dates or language. Two review authors independently screened the titles and abstracts of studies identified by the search strategies. Full texts of potentially eligible studies were independently assessed by two review authors for relevant studies. Any disagreements were resolved through discussion or in consultation with a third review author.

Inclusion and exclusion criteria

Index tests, reference standard, and diagnosis of interest

We considered studies that compared prenatal ultrasound measurements (index test) with postnatal direct measurements of head size (reference test). We included studies which used any of the following biometric parameter as index tests: head circumference (HC), occipito-frontal diameter (OFD), biparietal diameter (BPD), or ratios of any of these with either abdominal circumference (AC) or femur length (FL). Microcephaly (diagnosis) was the condition of interest, reported either as the only condition or separately in addition to other fetal brain abnormalities.

Types of studies

We considered for inclusion studies of any design (randomized controlled trial, prospective or retrospective cohort studies, cross-sectional studies and case-control studies) comparing prenatal assessment of fetal biometric parameters with standard postnatal head size measurements for diagnosing microcephaly. Case series and conference proceedings reporting original data and with adequate information were also considered for inclusion.

Types of participants

Pregnant women who had ultrasound measurements of fetal biometric parameters for diagnosis of microcephaly (irrespective of the indications for ultrasound). We planned to separately assess pregnant women suspected of being at risk of, or confirmed with ZIKV infection.
Data extraction and synthesis

Two review authors independently extracted data on participants’ characteristics (ZIKV virus infection status, gestational age at the time of ultrasound assessment). We extracted data on the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) to determine the sensitivity and specificity (with 95% confidence intervals [CI]) of the index tests for each fetal biometric parameter. Results from studies that presented insufficient data for meta-analysis were qualitatively presented. In one study where ultrasound and magnetic resonance imaging (MRI) were employed,(13) only sonographic data was extracted for the review.

For studies considered sufficiently similar with respect to the research questions, study design and execution, we performed a meta-analysis using a random effects model on pooled data to estimate group sensitivity and specificity (with 95% CI). We generated HROC curves using a hierarchical summary receiver operating characteristic (HSROC) model. To gauge overall test accuracy, we calculated a diagnostic odds ratio (DOR) and an area under curve (AUC) using Der Simonian-Laird random-modeling and Holling’s proportional hazard model.(14)

Data on TP, TN, FP, and FN using cut-off values ranging from 3 SD to 5 SD below the mean were applied to estimate diagnostic test accuracy of fetal ultrasound. Pre-test probabilities based on the incidence of microcephaly in unclassified pregnancy (0.0285%) (1) and ZIKV-infected pregnancies (0.95%) (1) were applied to estimate positive and negative predictive values. We registered this review in PROSPERO, the international prospective register of systematic reviews of the University of York and the National Institute for Health Research, under the number CRD42016039365.

Risk of bias assessment

Accuracy Studies (QUADAS-2) tool (15, 16) in Review Manager (RevMan Version 5.3.). We provided a rating for risk of bias and applicability concerns based on the presence or absence of indicators index and reference standard, flow and timing of prenatal and postnatal tests. We assessed as high risk where serious deficiencies in criteria were detected, and unclear or low risk where descriptions were inadequate or appropriate. For the meta-analyzed data, we assessed heterogeneity using the I² statistic (percentage of inter-study variation due to heterogeneity).

Results

Search results

The search strategies yielded 2,258 citations from the databases. One hundred and eleven potentially eligible studies were identified after screening of titles and abstracts and removing
duplicates (Figure 1). Full texts of all potentially eligible studies were assessed and studies excluded for reasons shown (Appendix 2). Nine studies met our inclusion criteria. Two of these reported sufficient data that could be used in meta-analysis while the other seven presented incomplete data and were described.

**Characteristics of included studies**

All included studies were based on hospital records in the USA (5), Israel (2), France (1) and Canada (1). The study designs were either prospective cohort (17-20) or retrospective cohort (13, 21-24) with enrollment periods spanning between 1979 and 2014.

The thresholds for prenatal and postnatal diagnoses of microcephaly were pre-specified in some studies. It was defined as head measurements of < 2 SD below the mean (17), < 3 standard deviations (SD) below the mean (18, 19, 21), 3 SD below the mean (24), > 3SD below the mean (20), below the 5th percentile (13) and the 10th percentile (22) threshold. The threshold applied was unstated in one study (23).

Only two studies contained data appropriate for meta-analysis as they assessed similar parameters at same thresholds (18, 19). The hierarchical summary receiver operating characteristic (HSROC) curves are shown in Figures 2 - 5.

Fetal microcephaly was secondary to cytomegalovirus (CMV)-infection (13) and phenylketonuria (PKU) (20) in 2 studies and congenital or primary in the seven other studies (17-19, 21-24).

In three out of nine studies (13, 20, 21), the ultrasound device used for prenatal detection of fetal parameters was reported. These included Acuson 128 XP 10 (Siemens) (20), GE Voluson 730 (13) and a range of ultrasound machines in the third study (21): GE Voluson E8, 730 Expert and Voluson 730 Pro (all GE Healthcare).

**Accuracy of ultrasound measurements of HC (8 studies, Table 2, Figures 3 G - I)**

Eight studies reported on the diagnostic accuracies of HC. Synthesis of two studies (18, 19) (45 fetuses) with meta-analyzable data showed sensitivities of 84%, 68% and 58% and specificity of 70%, 91% and 97% at thresholds of 3, 4 and 5 SD below the mean for GA, respectively. Based on these two studies, HC measurements using 3 SD below the mean was relatively high in sensitivity (84%), specificity (70%), positive likelihood ratio (2.6), and negative predictive values for unspecified (99%) and ZIKV-infected pregnant populations (99%) (Table 2, Figures 3 G - I). As the SD below the mean for GA increases from 3 to 5, the sensitivity decreases while the specificity increases substantially.
Descriptive data was provided in the other six studies (13, 17, 21-24). Among 42 fetuses prenatally diagnosed with microcephaly, Leibovitz et al. (21) reported 24 true positives and 18 false positives, and a positive predictive value (PPV) of 57.1 at a HC of 3 SD below the mean for GA. In a study on 20 suspected cases of fetal microcephaly, Stoler-Poria et al. (17) confirmed five cases to be true positives and 15 false positives with the true positive cases having a head circumference of between 2 and 4.8 SDs below the mean for gestational age. Wong et al. (22) reported comparable z-scores for prenatal and postnatal correlations in 455 fetuses. A z-score threshold of $\leq 1.3$ below the mean (44.6% sensitivity, 35.1% specificity, 44.9% FP rate, 45.9% FN rate,) was more sensitive and specific relative to a z-score of $\leq 1.7$ below the mean (28.8% sensitivity, 21% specificity, 62.6% FP rate, 28.2% FN rate). Additionally, an area under the ROC curve of 0.6 suggested inaccuracy of ultrasound prenatal diagnosis of microcephaly.

One study (13) reported a sensitivity of 85.7% and specificity of 85.3% for microcephaly detection at a head circumference of <5th percentile for gestational age. In this study, prenatal and postnatal findings were more consistent in the absence of coexisting brain abnormality. In another study (24), 11 of 16 cases of prenatally diagnosed microcephaly at a threshold of 3 SD below the mean for GA were false positive when examined at birth, giving a sensitivity of 31%. Campbell et al. (23) reported accurate identification of all 10 cases of microcephaly suspected before 24 weeks gestation at postnatal examination. There were no false positives or false negatives.

**Accuracy of ultrasound measurement of OFD (2 studies, Table 3, Figures 3 D - F)**

Pooled data from two studies (18, 19) (45 fetuses) reported sensitivities of 76%, 58% and 58% and specificity of 84%, 97% and 97% at 3, 4, and 5 SDs below the mean for GA, respectively. Higher thresholds were more sensitive while lower thresholds were more specific.

OFD measurement at a threshold of 3 SD below the mean for GA is more sensitive but measurements at 4 and 5 SDs are more specific. Given the extremely low incidence of microcephaly applied, the proportion of fetuses diagnosed with microcephaly based on 3, 4, and 5 SD thresholds who were correctly diagnosed (positive predictive values) was extremely low. Deduction of PPVs using 0.95% incidence of microcephaly among ZIKV infected women did improve the PPV (Table 3, Figures 3 D - F). However, the proportion of fetuses without microcephaly who were correctly diagnosed was close to 100% for the three thresholds for both unspecified and ZIKV-infected pregnancies.
Accuracy of ultrasound measurements of BPD (3 studies, Table 4, Figures 3 A - C)

Meta-analysis of two studies (18, 19), which included 51 fetuses reported a high sensitivity (94%) at 3 SD below the mean but lower sensitivities at 4 and 5 SDs. The specificity at 3 SD was very low but improved with lower cut-offs. The positive likelihood ratio for 3 SD suggests a slight increase in the likelihood of microcephaly but the confidence interval includes 1 (suggesting no change in the likelihood of microcephaly). The positive likelihood ratios for 4 and 5 SDs indicate a large and often conclusive increase in the likelihood of microcephaly with the ratios exceeding 1. The PPVs for unspecified and ZIKV-infected pregnancies were even much lower than for OFD measurements across the three thresholds.

One study (20) provided descriptive data. This study noted a low true positive and a high false negative frequency for 2nd trimester (3.2%; 29%) and 3rd trimester (42.9%; 57.1%) at a threshold of 3SD below the mean.

Accuracy of ultrasound measurements of the HC to AC ratio (3 studies)

We could not perform a meta-analysis for this parameter. Descriptive information on the accuracy of ratios of head perimeter to abdominal circumference for fetal biometry assessment was provided in only three studies (18, 19, 21). In one study (18), ultrasound detection of microcephaly with HC: AC ratio was consistently specific in diagnostic accuracy at all thresholds (3, 4 and 5 SDs) below the mean but varied in sensitivity, lower at 5SD (20%) and higher at 3SD (80%) both below the mean. Another study (19) accurately detected the absence of microcephaly at thresholds of 3, 4 and 5SD below the mean (specificity of 100%) with accuracy in sensitivity greatest at 3SD (80%) below the mean. The third study (21) identified a low sensitivity for the HC: AC ratio at <5th percentile for fetal suspicion (33.3%) and actual confirmation of microcephaly (37.5%).

Accuracy of ultrasound measurements of BPD to FL ratio (2 studies)

A meta-analysis was not possible for this parameter. In one study (18), the sensitivity and specificity of BPD: FL ultrasound measurements in detecting microcephaly were low at all thresholds measured (33 - 78%) but the specificity was high for measurements of 5 SD (87%) below the mean. Another study (24) noted the limitations of using the BPD: FL ratio for defining cases with or without microcephaly, and reported 5 true positives and 11 false positives.
Accuracy of ultrasound measurements of FL to HC ratio (3 studies)

Available studies could not be meta-analyzed. In one study (18), ultrasound measurement of FL:HC had a high sensitivity of 75 - 100% at 3-5 SDs and 87 - 100% specificity for ≤ 3SDs all below the mean. Another study(19) reported low sensitivity at 50 – 75% SD at all thresholds, highest at 1 SD (75%) and 85 - 100% specificity at ≤ 2 SD all below the mean for FL: HC parameter. Leibovitz et al. (21) showed that at <5th percentile, a HC: FL ultrasound measurement showed a low sensitivity for both suspected (52.4%) and confirmed microcephaly (50%).

Risk of bias assessment and applicability concerns (QUADAS-2)

The two studies included in meta-analysis were at risk of bias due to lack of pre-specified prenatal thresholds (18, 19) and inappropriate exclusions (19). Only one of these studies (19) had concerns regarding applicability due to limitation of study population to a short interval of <2 weeks between prenatal index scan and postnatal reference test.

In five of seven descriptive studies (13, 17, 20, 23, 24), a high risk of bias rating was assigned as they limited the study population of pregnant women to the following: CMV-infection and the availability of MRI and US diagnosis (13), Hebrew native language ability (17), mothers who presented with phenylketonuria (20), before 26 weeks gestation (23) or late trimester measurements (28 to 43 weeks) (24). The two other studies had a low risk of bias (21, 22).

Concerns regarding applicability were noted in two (13, 20) of the seven studies that provided only descriptive data. These studies included only high risk mothers infected with CMV (13) and having phenylketonuria (20). All other five studies (17, 21-24) had low concerns regarding applicability.

Discussion

This review provides a thorough overview of available information on prenatal application of ultrasound for diagnosis of microcephaly. HC and OFD measurements at 4 and 5 SD below the mean had high diagnostic odds ratio (25.3 to 48.0) and positive likelihood ratios (7.6 to 19.3) with wide 95% confidence intervals. Negative predictive values for unspecified and those extrapolated to ZIKV infected pregnancies at these standard deviations were also consistently high at close to 100%, although these values were derived from a relatively small number of fetuses. Thresholds of 4 and 5 SDs below the mean for OFD and HC showed a
tendency to consistently “rule in” the diagnosis of fetal microcephaly with reasonable level of confidence.

Our study indicates that the overall diagnostic test accuracy of ultrasound for predicting microcephaly at birth is limited as it varied with the applied cutoffs. Given the low incidence of microcephaly (1), fetal ultrasound seems not to have a large effect on the probability of identifying true cases of microcephaly. Inclusion of the presence of co-existing abnormalities such as intrauterine growth restriction and a detailed family history has been shown to improve the predictive value of ultrasound diagnosis (21). Therefore, setting an SD threshold to increase accuracy of microcephaly detection in any and ZIKV infected pregnancies should be informed by a balance of expert opinion, detailed history and analysis of other associated fetal anomalies (25).

Variation in sensitivity and specificity for all fetal head biometric measurements (BPD, HC, OFD) observed in all studies may have been due to trimester-specific changes in fetal growth, differences in ultrasound device, techniques and patient characteristics (congenital or acquired microcephaly, presence of other anomalies) (11, 26). Growth appreciably slows in the third trimester in fetus affected with microcephaly and influence of autosomal recessive inheritance in some cases may also play a role.

Fetuses with microcephaly are often miscarried, terminated or result in stillbirths which may explain the absence of comparative studies in ZIKV infected pregnant women. Comparisons with postmortem or pathological samples derived from such scenarios introduce some form of bias (27). In such cases, the estimated accuracy should be interpreted with caution.

Prenatal diagnostic accuracy of structural abnormalities affords informed maternal and health provider decisions, on whether to continue, terminate or institute fetal therapies. Potential misdiagnosis can be a source of emotional trauma during pregnancy, hence a review of growth standards employed and agreement with postnatal measurements can help eliminate or decrease the incidence of misdiagnosis.

To the best of our knowledge, this is the first systematic review to determine the diagnostic accuracy of prenatal ultrasound for fetal detection of fetal microcephaly. Our comprehensive search strategy and lack of date or language restrictions likely identified all studies. Our study had limitations. Primary data was from a limited number of fetuses and reported by studies with unclear or high risk of bias. The nature of studies included in quantitative synthesis and an overall high risk of bias rating limit the confidence in extrapolated results for Zika-infected pregnant women.

Trimester specific variation in fetal morphology visible on ultrasound measurements also limits the use of fetal biometric parameters in isolation (28). This proposes a need for
incorporating presenting features and a detailed history from the pregnant woman. Variation in thresholds, ultrasound device and timing of assessment during pregnancy adds potential flow and timing bias.

In conclusion, we provide evidence for the diagnostic accuracy of ultrasound in detection of fetal microcephaly. Ultrasound diagnostic accuracy of HC and OFD parameters at 4 and 5 SD below the mean was better at ruling in fetal microcephaly with high diagnostic odds ratio, sensitivity, specificity and positive likelihood ratio. The relative improvement in technology and skill suggests the need for new studies on the topic. It is reasonable to assume that the technical improvement of ultrasound machines in the last 20 years should contribute to improved diagnostic accuracy which was lacking in the studies published over 20 years ago. Research on diagnostic test accuracy based on present-day ultrasound devices are needed improve confidence in fetal microcephaly diagnosis.

Declaration of interest

The review authors have no conflicts of interest to declare.

Acknowledgements

We are grateful to Chiemi Kataoka and Miwako Segawa for their help with full text retrieval. The preparation of this review was financially supported by the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Reproductive Health and Research, World Health Organization.

The manuscript represents the views of the named authors only.

Abbreviations

MEDLINE Medical Literature Analysis and Retrieval System Online
EMBASE Excerpta Medica Database
CINAHL Cumulative Index of Nursing and Allied Health Literature
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-analyses
CDTA Cochrane Database of Diagnostic Test Accuracy
Authors’ contributions

OTO conceived the study. Hand-searching, screening, data extraction, was performed by ECC, AJQP, KSL, YT, CN and NM. OBO and AD assisted with full text screening and data extraction. TS assisted with search strategy and conducted the search. HN conducted the meta-analysis. CN, EO, RM and OTO gave guidance on project design, data extraction, meta-analysis and methodological assessments. ECC and AJQP drafted the manuscript. CN, EO, RM, and OTO revised the manuscript. All authors read and approved the final manuscript for publication.
References


Figure 1. PRISMA flow diagram

Records identified through database searching (n = 1532)
- MEDLINE – 563
- EMBASE -942
- CINAHL- 33
- WHOGL - 24

Records identified through DTA Ultrasonography database (n = 384)
- MEDLINE DTA - 320
- EMBASE DTA - 45
- Cochrane CDSR - 0
- Cochrane DARE - 0
- WHOGL DTA- 15

Records screened after duplicates removed (n = 1946)

Records excluded (n = 1839)

Full-text articles assessed for eligibility (n = 111)

Full-text articles excluded, with reasons (irrelevance to review objective or lack of prenatal or postnatal data) (n = 102)

Studies included in qualitative synthesis (n = 9)

Studies included in quantitative synthesis (Meta-analysis) (n = 2)
### Table 1: Characteristics of included studies

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Country</th>
<th>Enrolment period</th>
<th>Setting (e.g. facility, medical records)</th>
<th>Study design</th>
<th>Participant information</th>
<th>Index test</th>
<th>Reference test</th>
<th>Reported outcomes</th>
<th>Ultrasound device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chervenak 1984</td>
<td>USA</td>
<td>July 1 1979 to July 1 1983</td>
<td>Medical Center, medical records</td>
<td>Prospective</td>
<td>16 fetuses (initially 18, but 2 were excluded as they were stillbirths); Microcephaly was diagnosed if the head circumference was &lt; -3 SDs below the mean for gestational age.</td>
<td>HC OFD BPD HC:AP BPD:FL FL:HC</td>
<td>BPD OFD HC</td>
<td>Microcephaly</td>
<td>Not reported</td>
</tr>
<tr>
<td>Campbell 1983</td>
<td>USA</td>
<td>1978 to June 1983</td>
<td>Hospital records</td>
<td>Retrospective</td>
<td>10 cases correctly detected on the basis before 26 weeks gestation, with no false positives and no false negatives based on 2 parameters with prenatal and postnatal confirmation implied</td>
<td>Head circumference Abdominal circumference</td>
<td>Head circumference Abdominal circumference</td>
<td>Microcephaly</td>
<td>Not provided</td>
</tr>
<tr>
<td>Chervenak 1987</td>
<td>USA</td>
<td>1983-1986</td>
<td>Medical Center</td>
<td>Prospective</td>
<td>Prenatal diagnosis was done for 24 fetuses using different biometrical parameters.</td>
<td>HC OFD BPD HC:AP FL:HC</td>
<td>HC OFD BPD HC:AP FL:HC</td>
<td>Microcephaly Deaths Stillbirths, Encephalocele</td>
<td>Not reported</td>
</tr>
<tr>
<td>Wilson 1989</td>
<td>Canada</td>
<td>1982 to 1985</td>
<td>Hospital, medical records</td>
<td>Retrospective</td>
<td>16 cases identified prenatally were assessed for abnormalities.</td>
<td>Head circumference (HC)</td>
<td>HC (postnatal assessment)</td>
<td>Microcephaly</td>
<td>Not provided</td>
</tr>
<tr>
<td>Harvey L 1996</td>
<td>USA</td>
<td>Unknown</td>
<td>Maternal PKU Collaborative Study (MPKUCS) database</td>
<td>Prospective</td>
<td>31 fetuses in 2\textsuperscript{nd} trimester and 20 in the 3\textsuperscript{rd} trimester, all from pregnant mothers diagnosed with phenylketonuria (PKU) and limited to live births</td>
<td>BPD</td>
<td>BPD</td>
<td>Microcephaly</td>
<td>Acuson 128 XP 10 (Mountain View, CA, U.S.A.) scanner with a 3-5 or 5 MHz variable focus transducer.</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Year</td>
<td>Data Source</td>
<td>Study Design</td>
<td>Number of Fetuses</td>
<td>Description</td>
<td>Equipment</td>
<td>Abbreviations</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Benoist 2008</td>
<td>France</td>
<td>2000-2007</td>
<td>Hospital, medical records</td>
<td>Retrospective</td>
<td>49</td>
<td>49 fetuses of CMV infected mothers, prenatal ultrasound investigations were compared to postnatal investigations (both autopsy and live births). 38 live births, 10 terminations of pregnancy and 1 fetal death.</td>
<td>Serial targeted transabdominal or transvaginal ultrasound of the head circumference (every fortnight from diagnosis until delivery).</td>
<td>GE Voluson 730 ultrasound examinations with high-frequency probes (transabdominal for breech presentation (4–8 MHz) and transvaginal for normal presentation (5–9 MHz) routes) GE Medical Systems, Ultrasound and Primary Care Diagnostic, Gif sur Yvette, France.</td>
<td></td>
</tr>
<tr>
<td>Stoler-Poria 2010</td>
<td>Israel</td>
<td>2001 to 2005</td>
<td>Medical center</td>
<td>Prospective</td>
<td>20</td>
<td>20 fetuses were included and followed up for neurodevelopment outcomes.</td>
<td>Head circumference (HC)</td>
<td>Developmental outcome, neurological development, microcephaly</td>
<td>Not provided</td>
</tr>
<tr>
<td>Wong 2012</td>
<td>USA</td>
<td>January 2005 to July 2011</td>
<td>Hospital, medical charts</td>
<td>Retrospective</td>
<td>730</td>
<td>730 ultrasounds of 455 fetuses in 433 patients.</td>
<td>Head circumference</td>
<td>Microcephaly</td>
<td>Not provided</td>
</tr>
<tr>
<td>Leibovitz 2016</td>
<td>Israel</td>
<td>2007 to 2014</td>
<td>Hospital, medical records</td>
<td>Retrospective</td>
<td>42</td>
<td>42 fetuses were evaluated.</td>
<td>Head circumference (1.62 (BPD + HC)</td>
<td>Fetal Microcephaly Normocephaly</td>
<td>Voluson E8, Voluson 730 Expert, and Voluson 730 Pro ultrasound machines (GE Healthcare Ultrasound, Milwaukee, WI, USA).</td>
</tr>
</tbody>
</table>

**Abbreviations:** Microcephaly (MCP), Ultrasound (US), head circumference (HC), head circumference, occipitofrontal diameter (OFD), biparietal diameter (BPD), abdominal circumference (AC), Femur length (FL), standard déviation (SD).
Figure 2. QUADAS-2 summary of risk of bias and applicability concerns of included studies
Table 2. Diagnostic accuracy of ultrasound measurements of head circumference for prenatal assessment of microcephaly

<table>
<thead>
<tr>
<th></th>
<th>-3 SD</th>
<th>-4 SD</th>
<th>-5 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cohorts</td>
<td>2(18, 19)</td>
<td>2(18, 19)</td>
<td>2(18, 19)</td>
</tr>
<tr>
<td>Number of comparisons</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Diagnostic odds ratio</td>
<td>12.7 (2.1-76.5), I² = 0%</td>
<td>25.3 (3.7-171.6), I² = 0%</td>
<td>48.0 (4.8-481.5), I² = 0%</td>
</tr>
<tr>
<td>AUC</td>
<td>0.84</td>
<td>0.88</td>
<td>0.68</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.84 (0.36-0.98)</td>
<td>0.68 (0.33-0.90)</td>
<td>0.58 (0.30-0.82)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.70 (0.34-0.91)</td>
<td>0.91 (0.74-0.97)</td>
<td>0.97 (0.83-1.00)</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>2.6 (0.88-8.4)</td>
<td>7.6 (2.1-25.7)</td>
<td>19.3 (3.0-126.3)</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.24 (0.030-1.1)</td>
<td>0.35 (0.11-0.76)</td>
<td>0.43 (0.19-0.74)</td>
</tr>
<tr>
<td>PPV (general pregnancy)</td>
<td>0.00075</td>
<td>0.00215</td>
<td>0.00548</td>
</tr>
<tr>
<td>NPV (general pregnancy)</td>
<td>0.99993</td>
<td>0.99990</td>
<td>0.99988</td>
</tr>
<tr>
<td>PPV (ZIKV infected pregnancy)</td>
<td>0.0262</td>
<td>0.0676</td>
<td>0.1564</td>
</tr>
<tr>
<td>PPV (ZIKV infected pregnancy)</td>
<td>0.9978</td>
<td>0.9966</td>
<td>0.9959</td>
</tr>
</tbody>
</table>

Parentheses indicate 95% confidence interval (CI). Pre-test probabilities, i.e. incidence of microcephaly among general pregnancies and ZIKV-infected pregnancies were estimated as 0.0285% and 0.95%, respectively.
Table 3. Diagnostic accuracy of ultrasound measurements of occipito-frontal diameter for prenatal assessment of microcephaly

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>-3 SD</th>
<th>-4 SD</th>
<th>-5 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cohorts</td>
<td>2(18, 19)</td>
<td>2(18, 19)</td>
<td>2(18, 19)</td>
</tr>
<tr>
<td>Number of comparisons</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Diagnostic odds ratio</td>
<td>18.6 (2.8-124.2), $I_2 = 0%$</td>
<td>48.0 (4.8-481.5), $I_2 = 0%$</td>
<td>48.0 (4.8-481.5), $I_2 = 0%$</td>
</tr>
<tr>
<td>AUC</td>
<td>0.88</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.76 (0.17-0.98)</td>
<td>0.58 (0.30-0.82)</td>
<td>0.58 (0.30-0.82)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.84 (0.50-0.97)</td>
<td>0.97 (0.83-1.00)</td>
<td>0.97 (0.83-1.00)</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>4.8 (0.73-23.3)</td>
<td>19.3 (3.0-126.3)</td>
<td>19.3 (3.0-126.3)</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.29 (0.024-1.1)</td>
<td>0.43 (0.19-0.74)</td>
<td>0.43 (0.19-0.74)</td>
</tr>
<tr>
<td>PPV (general pregnancy)</td>
<td>0.00135</td>
<td>0.00548</td>
<td>0.00548</td>
</tr>
<tr>
<td>NPV (general pregnancy)</td>
<td>0.99992</td>
<td>0.99988</td>
<td>0.99988</td>
</tr>
<tr>
<td>Accuracy (ZIKV infected pregnancy)</td>
<td>0.84</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>PPV (ZIKV infected pregnancy)</td>
<td>0.0436</td>
<td>0.1564</td>
<td>0.1564</td>
</tr>
<tr>
<td>NPV (ZIKV infected pregnancy)</td>
<td>0.9973</td>
<td>0.9959</td>
<td>0.9959</td>
</tr>
<tr>
<td>Accuracy (ZIKV infected pregnancy)</td>
<td>0.84</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Parentheses indicate 95% CI. Pre-test probabilities, i.e. incidence of microcephaly among general pregnancies and ZIKV-infected pregnancies were estimated as 0.0285% and 0.95%, respectively.
Table 4. Diagnostic accuracy of ultrasound measurements of biparietal diameter for prenatal assessment of microcephaly

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>-3 SD (95% CI)</th>
<th>-4 SD (95% CI)</th>
<th>-5 SD (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cohorts</td>
<td>2(18, 19)</td>
<td>2(18, 19)</td>
<td>2(18, 19)</td>
</tr>
<tr>
<td>Number of comparisons</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Diagnostic odds ratio</td>
<td>1.6 (0.056-46.1), $\Gamma^2 = 0%$</td>
<td>4.7 (0.86-25.5), $\Gamma^2 = 0%$</td>
<td>4.7 (0.66-33.9), $\Gamma^2 = 0%$</td>
</tr>
<tr>
<td>AUC</td>
<td>0.888</td>
<td>0.77</td>
<td>0.66</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.94 (0.67-0.99)</td>
<td>0.85 (0.46-0.98)</td>
<td>0.59 (0.30-0.83)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.16 (0.06-0.37)</td>
<td>0.46 (0.14-0.81)</td>
<td>0.80 (0.21-0.99)</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>1.1 (0.82-1.5)</td>
<td>1.6 (0.70-4.5)</td>
<td>3.0 (0.59-46.0)</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.38 (0.047-2.5)</td>
<td>0.33 (0.045-1.9)</td>
<td>0.51 (0.21-2.6)</td>
</tr>
<tr>
<td>PPV (general pregnancy)</td>
<td>0.00032</td>
<td>0.00045</td>
<td>0.00084</td>
</tr>
<tr>
<td>NPV (general pregnancy)</td>
<td>0.99989</td>
<td>0.99991</td>
<td>0.99985</td>
</tr>
<tr>
<td>PPV (ZIKV infected pregnancy)</td>
<td>0.0106</td>
<td>0.0109</td>
<td>0.0275</td>
</tr>
<tr>
<td>NPV (ZIKV infected pregnancy)</td>
<td>0.9964</td>
<td>0.9969</td>
<td>0.995</td>
</tr>
</tbody>
</table>

Parentheses indicate 95% CI. Pre-test probabilities, i.e. incidence of microcephaly among general pregnancies and ZIKV-infected pregnancies were estimated as 0.0285% and 0.95%, respectively.
Figure 3. Hierarchical summary receiver operating characteristics curves for HC, OFD and BPD at 3, 4 and 5 SD below the mean. The size of each circle reflects weight, not confidence region. (Open arrow: Two circles had exactly same accuracy and weight. Filled arrow: Three circles had exactly same accuracy and weight).