**Article:**

Kapaya, H., Jacques, R.M., Rahaim, N. et al. (1 more author) (2016) “Does Short-term variation in fetal heart rate predict fetal acidaemia?” A Systematic review and meta-analysis. Journal of Maternal-Fetal & Neonatal Medicine. ISSN 1476-7058

https://doi.org/10.3109/14767058.2016.1156670

**Title**

***"Does Short-term variation in fetal heart rate predict fetal acidaemia?" A Systematic review and meta-analysis***

**Authors**

Dr Habiba Kapaya, Dr Richard Jacques, Dr Nadia Rahaim, Professor Dilly Anumba

**Author affiliations**

**Dr Habiba Kapaya** MBBS PhD MRCOG FCPS DFSRH

Clinical Lecturer (NIHR)

Department of Human Metabolism

Academic Unit of Reproductive & Developmental Medicine

4th Floor Jessop Wing

Tree Root Walk, Sheffield S102SF, UK

Telephone number +44 114 2268195

Fax number +44 114 2268538

Email address: h.kapaya@sheffield.ac.uk

**Dr Richard Jacques** BSc, MSc, PhD

Dr Richard M Jacques

Medical Statistics Group

School of Health and Related Research (ScHARR)

University of Sheffield

Regent Court 30 Regent Street Sheffield, UK

S1 4DA

Telephone number +44 114 2220813

Fax number +44 114 2220749

Email address: r.jacques@sheffield.ac.uk

**Nadia S A Rahaim** MBBS MSc in reproductive and maternal medicine

Department of Human Metabolism

Academic Unit of Reproductive & Developmental Medicine

4th Floor Jessop Wing

Tree Root Walk, Sheffield S102SF, UK

Telephone number +44 114 2268317

Fax number +44 114 2268538

Email address: mdp11nr@sheffield.ac.uk

**Professor Dilly Anumba** MBBS.FWACS.FRCOG.MD.LL.M (Medical Law)

Chair of Obstetrics and Gynaecology,

Consultant in Obstetrics & Gynaecology/Subspecialist in Maternal and Fetal Medicine

Academic Training Programme Director

Sheffield Medical School & South Yorkshire and Humber Deanery

4th Floor Jessop Wing

Tree Root Walk, Sheffield S102SF, UK

Telephone number +44 114 2268317

Fax number: +44 114 2268538

Email address: d.o.c.anumba@sheffield.ac.uk

***Corresponding Author**

**Dr Habiba Kapaya**

Clinical Lecturer (NIHR)

Department of Human Metabolism

Academic Unit of Reproductive & Developmental Medicine

4th Floor Jessop Wing

Tree Root Walk, Sheffield S102SF, UK

Telephone no: +44 114 2268195

Email address: h.kapaya@sheffield.ac.uk

**Short Title**

Short-term variation in predicting fetal acidaemia

**Key words**

Computer analysis

Computerized cardiotocography

Fetal acidaemia

Fetal acid-base status

Fetal heart rate (FHR)

Neonatal acidaemia

Short-term variation (STV)

Umbilical blood gas analysis

## ABSTRACT

**Objective:** To evaluate the association of short-term variation (STV) of the fetal heart rate in predicting fetal acidaemia at birth.

**Methods**: The search strategy employed searching of electronic databases (MEDLINE**,** Web of Science, Scopus, and Google Scholar) and reference lists of relevant studies. Data was extracted from studies, adhering strictly to the following criteria: singleton pregnancy at ≥ 24 weeks gestation, computerised CTG (index test) and calculation of STV before delivery. The outcome measure was arterial pH assessed in cord blood obtained at birth.

**Results:** Meta-analysis showed moderate accuracy of STV in predicting fetal acidaemia with a sensitivity of 0.57 (95% CI: 0.45 to 0.68), specificity of 0.81 (95% CI: 0.69 to 0.89), positive likelihood ratio of 3.14 (95% CI: 2.13 to 4.63) and negative likelihood ratio of 0.58, (95% CI: 0.46 to 0.72). However, in intra-uterine growth restricted fetuses, a small improvement in detecting acidaemia was observed; with a sensitivity of 0.63, (95% CI: 0.49 to 0.75) and negative likelihood ratio of 0.50, (95% CI: 0.31 to 0.80).

**Conclusion:** STV appears to be a moderate predictor for fetal acidaemia. However, its usefulness as a stand-alone test in predicting acidaemia in clinical setting remains to be determined.

## Introduction

Stillbirth is a devastating pregnancy outcome affecting millions of families worldwide and with minimal decline in its incidence in 20 years[1]. Despite advances in obstetric care, identifying antenatal fetal surveillance tests that have the highest predictive accuracy for fetal risks remains a challenge[2-4]. Cardiotocography (CTG), also called electronic fetal heart rate (FHR) monitoring (EFM), is widely used to assess the *in utero* fetal condition. This assessment can give vital short-term information about fetal wellbeing. However, it is subjective and associated with high intra-observer and inter-observer variability[5]. More recently, various computerised systems for objective FHR analysis have been developed, one of the most validated and used being the Oxford-system[6], devised by Dawes, Redman and colleagues in 1982[7]. Currently in use today in many antenatal clinics and Day assessment Units, computerised systems eliminate observer variability, improve the reproducibility of EFM, and determine FHR parameters such as short-term variation (STV) that cannot be assessed visually[8].

The value of STV has been investigated on its own[9,10] or in combination with other modalities of fetal surveillance[2,11] in predicting fetal acidaemia at birth. Some studies suggest that STV is a powerful and reliable indicator of fetal acid base status, fetal hypoxemia, and stillbirth[12,13]. However other reports suggest that it is a poor predictor of perinatal outcome[9]. Possible reasons for the controversy include differences in study design and population heterogeneity, varied gestational ages at which studies were carried out, non-uniformity in defining acidaemia at birth and variable thresholds of STV employed to predict fetal-neonatal acidaemia. Given these seeming conflicts in study quality and data interpretation, a comprehensive and systematic review is required to determine the usefulness of STV in predicting fetal and neonatal compromise.

The purpose of this review is to evaluate current literature and establish the strength of association of STV with neonatal acidaemia.

## Materials and Methods

### *Search strategy*

This systematic review was carried out according to the standards set by the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) group [14]

We conducted an extensive search in Medline (PubMed and Ovid), Google Scholar, Web of Science and Scopus from inception until March 2015. For the Medline search we used a combination of MeSH headings, such as “Short-term variation” OR “STV” AND “Fetal heart rate” OR “FHR”.

The entire search strategy was limited to human studies published in English. The manuscripts were examined for duplicated populations. If any were found we selected the most recent and complete version. The search was performed by two physicians and a medical school librarian. Bibliographies of relevant retrieved studies and recent reviews were hand-searched to identify cited articles not captured by electronically. We scrutinised the abstracts identified by the electronic searches and obtained full manuscripts of all the citations that were thought to have met the inclusion criteria. Two reviewers (HK and RJ) independently examined the manuscripts to determine their eligibility for inclusion in the review and assessed their quality. Methodological quality was defined as the confidence that the study design, conduct, and analysis have minimised biases in estimation of the association. The articles were assessed using the complete STARD and QUADS checklists [15,16]. These are guidelines for reporting, and methodological quality of studies on diagnostic accuracy. We rated the study as being of high quality if it had at least four of the following items: adequate description of the population, adequate description of computerised CTG using Dawes and Redman criteria for the interpretation of FHR parameters in particular STV, duration of FHR tracing and record of last assessment prior to delivery, outcome measures defined appropriately, prospective recruitment, blinding of the investigators carrying out the outcome measure and the statement on assessing the value of STV in predicting perinatal outcome. When a study adhered to three or fewer of these criteria we considered it to be of medium or low quality.

***Data extraction***

We extracted data from the studies, adhering strictly to the following criteria: singleton pregnancy at or more than 24 weeks gestation, computerised CTG (index test) until the Dawes and Redman criteria for normality were met and calculation of STV before delivery; outcome measure being arterial or venous pH or base excess assessed in cord blood obtained at birth. Observational studies that allowed generation of a 2×2 table (true positives, false positives, false negatives, and true negatives) to compute an estimate of the association between STV and neonatal acidaemia were included. We excluded studies with five or fewer cases, because of unreliability.

***Data synthesis***

We used 2x2 tables to calculate point estimates and 95% confidence intervals of sensitivity (true positive rate), specificity (true negative rate), positive likelihood ratio (how many times more likely positive index test results were in the acidaemia group compared to the non-acidaemia group), and negative likelihood ratio (how many times less likely negative index test results were in the acidaemia group compared to the non-acidaemia group) for individual studies. When calculating the likelihood ratios, where 2×2 tables contained zero cells, 0.5 was added to each cell to enable the calculations to be carried out.

Following the guidance of the Cochrane Collaboration [17] we did not formally test for heterogeneity in sensitivity and specificity using the $I^2$ statistic. Instead, we assessed the magnitude of observed heterogeneity graphically by plotting the sensitivities and specificities in the receiver operating characteristic (ROC) space, examining how close the observed results lie to the summary ROC curve and the associated prediction ellipse.

We considered study design, study quality, population risk, gestational age at the time of index test and differences in cut off value of pH for diagnosing acidaemia as potential sources of heterogeneity. Pooled estimates of positive and negative likelihood ratios were calculated using the DerSimonian-Liard method [18]. A bivariate meta-analysis model was used to calculate the pooled estimates of sensitivity and specificity and fit a summary ROC curve [19]. Using a bivariate model allows the correlation that exists between sensitivity and specificity to be incorporated.

When reason for heterogeneity was identified, we carried out sub-group analyses. As clinical heterogeneity was present between studies, we used random effects models throughout.

Analysis was conducted in R (R Core Team) [20] using the mada package [21].

**Results:**

An initial search of 398 citations, identified 7 primary articles [10,11,22-26] reporting on diagnostic accuracy of STV and neonatal acidaemia, allowing generation of 2×2 tables of accuracy (true positives, false positives, false negative, true negatives) for individual studies.

Figure 1 displays the retrieval process of the relevant articles. The selection process started by screening the title and abstract to exclude the non-related papers then the duplicates were removed electronically. In accordance to the pre-set inclusion and exclusion criteria the two reviewers re-evaluated the recruited articles, agreed on the included papers and reached consensus regarding inconsistencies.

Table 1 details the study characteristics of the individual studies. There were 7 studies included [10,11,22-26] that used diagnostic accuracy to establish the value of STV in the prediction of neonatal acidaemia. The included studies totalled 780 pregnancies and reported on arterial cord pH, with thresholds ranging from <7.00 to 7.25. Computerised CTG was performed and STV was measured within 24 hrs of delivery in all except for one study [24] where no such record was found. Five studies were performed on high risk pregnancies [10,11,23,25,26], one on low-risk [24] and one included both high and low risk populations [22]. Amongst high-risk pregnancies, 4 studies observed a predictive relationship of STV for neonatal acidaemia in intra-uterine growth restricted (IUGR) fetuses [10,11,23,26]. However, all studies excluded fetuses with chromosomal and structural anomalies. There were 2 studies using retrospective [24,26], 2 consecutive [10,22], 1 prospective [11], 1 cross-sectional [25] and 1 undefined [23] patient recruitment.

Figure 2 shows a summary of the risk of bias and concerns about applicability of the included studies assessed using the QUADAS-2 checklist [15]. Overall study quality was variable, revealing some deficiencies in the reporting. In terms of patient selection, 43% (3/7) had a possibility of introducing bias because of a lack of information about how patients were enrolled [23-25]. One study did not verify the duration of STV and time of recording in relation to delivery and this raised concerns regarding its applicability to this meta-analysis [24]. There was a possible risk of bias due to the reference standard in 86% (6/7) studies due to poor reporting of whether the reference standard results were interpreted without knowledge of the results of the index test. However, all studies contained an adequate description of the performance of the STV and outcome measure. Overall, we had concern about the applicability of only one study and this was investigated further as part of sensitivity analyses [24].

Results for individual studies are summarised in Table 2, the pooled results for the positive and negative likelihood ratios are shown in Figure 3, and the pooled results for sensitivity and specificity are displayed on summary ROC plots in Figure 4. The pooled results show that STV predicts neonatal acidaemia with sensitivity of 0.57 (95% CI: 0.45 to 0.68) and specificity of 0.81 (95% CI: 0.69 to 0.89). The pooled estimate for the positive likelihood ratio was 3.14 (95% CI: 2.13 to 4.63) and the pooled estimate for the negative likelihood ratio was 0.58 (95% CI: 0.47 to 0.72).

A number of sensitivity analyses were also conducted to determine how robust the pooled results were to studies shown to be outlying on the summary ROC plots as well as to any studies that may have introduced bias. The first analysis removed the study of Galazios et al [24], as this was determined most likely to introduce bias and had the most concern about applicability. Removing this study gave pooled sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio estimates of 0.61 (95% CI: 0.51 to 0.71), 0.75 (95% CI: 0.67 to 0.82), 2.64 (95% CI: 1.94 to 3.60) and 0.52 (95% CI: 0.41 to 0.67) respectively. The second analysis removed the two outlying studies with sensitivity of 1, Guzman et al [10] and Anceschi et al [23]. Removing these studies gave pooled sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio estimates of 0.53 (95% CI: 0.39 to 0.66), 0.84 (95% CI: 0.69 to 0.93), 3.13 (95% CI: 1.88 to 5.21) and 0.62 (95% CI: 0.53 to 0.72) respectively.

Due to clinical heterogeneity between the populations studied, subgroup analysis was performed on the four studies with IUGR [10,11,23,26]. The pooled results from this high-risk population show that STV predicts neonatal acidaemia with a sensitivity of 0.63 (95% CI: 0.49 to 0.75), specificity of 0.72 (95% CI: 0.61 to 0.82), positive likelihood ratio of 2.62 (95% CI: 1.71 to 4.03), and negative likelihood ratio of 0.50 (95% CI: 0.31 to 0.80). Due to the small number of studies we were unable to conduct any sensitivity analyses for this subgroup. We also considered heterogeneity due to differences in cut-off value of pH for diagnosing acidaemia and difference in cut-off values for STV in predicting outcome. However, visual checks of sensitivity and specificity did not reveal any evidence of a threshold effect.

## Discussion

For the prediction of fetal acidaemia, STV was found to have an overall moderate predictive accuracy with a sensitivity of 0.57 (95% CI: 0.45 to 0.68), specificity of

0.81 (95% CI: 0.69 to 0.89), positive likelihood ratio of 3.14 (95% CI: 2.13 to 4.63) and negative likelihood ratio of 0.58 (95% CI: 0.46 to 0.72). However, in fetuses with IUGR, there was a small improvement in the sensitivity (0.63, 95% CI: 0.49 to 0.75) and negative likelihood ratio (0.50, 95% CI: 0.31 to 0.80) of detecting acidaemia, but small decrease in specificity (0.72, 95% CI: 0.61 to 0.82) and positive likelihood ratio (2.62, 95% CI: 1.71 to 4.03).

The strength of our review and the validity of our inferences lie in the methodology used. We complied with existing guidelines for the reporting of systematic reviews of diagnostic [27] and observational studies [14] evaluating causal association. Our literature searches were extensive in relevant databases and careful attention was paid to assessment of design quality and reporting. We developed this strategy in consultation with a medical school librarian, avoiding the chance of missing eligible publications. Our review provides the best available evidence, at the present time of the association between STV and neonatal acidaemia.

The studies we pooled had heterogeneity in terms of quality, population risk, threshold for STV and neonatal acidaemia. In order to tackle this problem, we carried out recommended analyses, including bivariate analysis and sub-group meta-analysis with pooled sensitivities and specificities [28,29].

The limitations of our review lie mainly in the lack of clear reporting within individual studies and in the residual heterogeneity despite sub-group analysis. It is accepted that poor study design and conduct may affect the estimates of diagnostic accuracy [30,31], but it is not entirely clear how individual aspects of quality may affect accuracy and to what magnitude. To overcome this problem, each individual paper must be assessed, ideally by meta-regression [32] using items of study quality. Due to small sample size of primary studies, it was not possible to apply meta-regression, thus sub-group analysis using a random effects model was performed to account for heterogeneity and underpowered studies and no significant difference was observed. Furthermore, we pooled sensitivities and specificities which are less susceptible to variations in prevalence [29] and are more useful in determining the probability of having a problem, following the test. These strategies enabled us to perform meta-analysis, despite the presence of heterogeneity among the selected studies.

The quality of primary studies varied. Anceshi et al[22] evaluated 195 singleton pregnancies and studied STV between 26 and 42 weeks gestation. They observed that within the group of pregnant women below 34 weeks of gestation, STV less than 5.1ms was a significant predictor of acidemia (Sensitivity of 100%, Specificity of 61%;P<0.05), whereas for the whole group, sensitivity was 62.5% and specificity was 78.5%. Similar results were observed in their study on 24 preterm IUGR fetuses [23], where STV less than 4.5ms predicted acidaemia with a sensitivity of 100% and specificity of 70%.

Guzman et al [10] and Serra et al [26] studied IUGR fetuses between 26 and 42 gestational age. In agreement with the above authors they found STV <3.5ms and <4.7ms as significant predictors of neonatal acidemia. In contrast to the above studies, Turan et al [11] attempted to integrate STV with the venous Doppler and the biophysical profile score in 58 IUGR fetuses. They found that although STV <3.5ms predicted neonatal acidemia with of sensitivity 47% and specificity of 83%, but when combined with Doppler studies, the sensitivity increased to 56% and specificity decreased to 79%.

Galazios et al [24] reported their results on 167 uncomplicated pregnancies. They found that STV <5ms predicted neonatal acidemia with sensitivity of 34% and specificity of 96.6%. Although they included a large number of cases, their definition of acidemia was considered at pH <7.25, which is much higher compared to other studies, and necessitating caution in interpreting their results.

Garcia et al [25] observed that STV threshold of 5.25ms predicted acidaemia with a sensitivity of 57.1% and a specificity of 85.2% in 41 pregnant women with hypertensive disorder. However, the study did not address and investigate the confounding and known influence of various drugs such as hydralazine and magnesium sulphate on FHR and its variation [33]. Therefore, it is difficult to establish whether the observed changes in STV truly reflected in-utero hypoxia or affected by other confounding factors.

It is difficult to draw definitive conclusion based on the studies included for this review and to answer the following questions:

- Does STV reliably predicts neonatal acidaemia or not?
- Is STV a better marker for detecting acidaemia in high-risk pregnancies, in particular IUGR, compared to low-risk pregnancies?
- Does STV predict acidaemia in high-risk preterm fetuses compared to term?
- Does STV performs better when combined with other modalities of fetal surveillance?

This review has highlighted gaps in literature and stresses the importance of exploring further areas of study design that will be vital to answer these questions.

FHR variability is known to depend on several factors such as gestational age [34]. It is well understood that STV increases as gestational age progresses, reflecting development of fetal autonomic nervous system (ANS) and maturation of vagal innervation to the fetal heart [34,35]. None of the studies, except for one [22] divided their study group into subgroups according to gestational age. Furthermore, all studies included in this review, except for two [10,26] measured STV for a duration of 40 minutes or even less. It is well known from previous research that a healthy fetus can remain in quiet sleep for up to 50 minutes with an unreactive, low FHR variation [5,36]. Studies of STV that include fetal monitoring for longer periods of time are therefore required to determine its true predictive accuracy for fetal acidemia. Moreover, the period of time between the determination of fetal heart STV and fetal delivery varied significantly between studies and subjects. Further studies are therefore required to determine the optimum timing of STV assessment for predicting fetal acidemia before its clinical utility and more widespread adoption can be ascertained.

In conclusion, the results of this meta-analysis highlight the need for high quality primary studies of STV predictive accuracy for fetal acidaemia. Furthermore, its place in assessing the fetus at risk of acidaemia in clinical care, singly or in conjunction with current surveillance techniques remains to be determined. Such studies need to include other surveillance techniques in clinical practice, as well as individual patient data enabling the test to be assessed at an individual level.

**Acknowledgements**

We would like to thank Anthea Tucker, librarian at the University of Sheffield, for her assistance in constructing the most effective and efficient search strategy for the systematic review.

**Declaration of interests**

The authors have no conflict of interest

## References


1. Mistry H, Heazell AE, Vincent O, Roberts T. A structured review and exploration of the healthcare costs associated with stillbirth and a subsequent pregnancy in England and Wales. BMC Pregnancy Childbirth 2013;13:236.
2. Bilardo CM, Wolf H, Stigter RH, Ville Y, Baez E, Visser GH, Hecher K. Relationship between monitoring parameters and perinatal outcome in severe, early intrauterine growth restriction. Ultrasound Obstet Gynecol 2004;23(2):119-25.
3. Baschat AA. Arterial and venous Doppler in the diagnosis and management of early onset fetal growth restriction. Early Human Development 2005;81(11):877-887.
4. Dawes GS, Moulden M, Redman CWG. Short-term fetal heart rate variation, decelerations, and umbilical flow velocity waveforms before labor. Obstetrics and Gynecology 1992;80(4):673-678.
5. Pardey J, Moulden M, Redman CW. A computer system for the numerical analysis of nonstress tests. Am J Obstet Gynecol 2002;186(5):1095-103.
6. Tsikouras P, Galazios G, Liberis V, Sigalas I, Kontomanolis E, Koutlaki N, Maroulis G. Short term variation in computerized CTG and fetal outcome. Bevilacqua G, editor2007. 787-792 p.
7. Huhn EA, Lobmaier S, Fischer T, Schneider R, Bauer A, Schneider KT, Schmidt G. New computerized fetal heart rate analysis for surveillance of intrauterine growth restriction. Prenat Diagn 2011;31(5):509-14.
8. Georgieva A, Papageorghiou AT, Payne SJ, Moulden M, Redman CWG. Phase-rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidaemia at birth. Bjog-an International Journal of Obstetrics and Gynaecology 2014;121(7):889-894.
9. Chang TC, Tan KT, Neow P, Yeo GS. Computerised analysis of foetal heart rate variation: prediction of adverse perinatal outcome in patients undergoing prostaglandin induction of labour at term. Ann Acad Med Singapore 1997;26(6):772-5.
10. Guzman ER, Vintzileos AM, Martins M, Benito C, Houlihan C, Hanley M. The efficacy of individual computer heart rate indices in detecting acidemia at birth in growth-restricted fetuses. Obstet Gynecol 1996;87(6):969-74.
11. Turan OM, Berg C, Moyano D, Bhide A, Bower S, Thilaganathan B, Gembruch U, Nicolaides K, Harman C, Baschat AA. Computerized fetal heart rate analysis, Doppler ultrasound and biophysical profile score in the prediction of acid-base status of growth-restricted fetuses. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 2007;30(5):750-6.
12. Street P, Dawes GS, Moulden M, Redman CW. Short-term variation in abnormal antenatal fetal heart rate records. Am J Obstet Gynecol 1991;165(3):515-23.
13. Dawes GS, Moulden M, Redman CW. Short-term fetal heart rate variation, decelerations, and umbilical flow velocity waveforms before labor. Obstet Gynecol 1992;80(4):673-8.
14. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283(15):2008-12.
15. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, Group Q-. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-36.
16. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC and others. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Fam Pract 2004;21(1):4-10.
17. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM Gatsonis C (editors), Cochrane Handbook for

Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.
18. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-88.
19. Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology 2005; 58: 982-990.
20. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna Austria. http://www.R-project.org/.
21. Doebler P (2015). mada: Meta-Analysis of Diagnostic Accuracy. R package version 0.5.7. http://cran.R-project.org/package=mada.
22. Anceschi MM, Piazze JJ, Ruozi-Berretta A, Cosmi E, Cerekja A, Maranghi L, Cosmi EV. Validity of short term variation (STV) in detection of fetal acidemia. Journal of Perinatal Medicine 2003;31(3):231-236.
23. Anceschi MM, Ruozi-Berretta A, Piazze JJ, Cosmi E, Cerekja A, Meloni P, Cosmi EV. Computerized cardiotocography in the management of intrauterine growth restriction associated with Doppler velocimetry alterations. Int J Gynaecol Obstet 2004;86(3):365-70.
24. Galazios G, Tripsianis G, Tsikouras P, Koutlaki N, Liberis V. Fetal distress evaluation using and analyzing the variables of antepartum computerized cardiotocography. Arch Gynecol Obstet 2010;281(2):229-33.
25. Garcia GS, Mariani Neto C, Araujo Jr E, Garcia RL, Nardozza LMM, Moron AF. Fetal acidemia prediction through short-term variation assessed by antepartum computerized cardiotocography in pregnant women with hypertension syndrome. Archives of Gynecology and Obstetrics 2008;278(2):125-128.
26. Serra V, Moulden M, Bellver J, Redman CW. The value of the short-term fetal heart rate variation for timing the delivery of growth-retarded fetuses. BJOG 2008;115(9):1101-7.
27. Khan KS, Dinnes J, Kleijnen J. Systematic reviews to evaluate diagnostic tests. Eur J Obstet Gynecol Reprod Biol 2001;95(1):6-11.
28. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58(10):982-90.
29. Altman DG. Practical Statistics for Medical Research, 1st edn. London: Chapman & Hall, 1997.
30. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282(11):1061-6.
31. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174(4):469-76.
32. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002;21(11):1525-37.
33. Nijhuis IJM, Ten Hof J, Mulder EJH, Nijhuis JG, Narayan H, Taylor DJ, Visser GHA. Fetal heart rate in relation to its variation in normal and growth retarded fetuses. European Journal of Obstetrics Gynecology and Reproductive Biology 2000;89(1):27-33.
34. Serra V, Bellver J, Moulden M, Redman CW. Computerized analysis of normal fetal heart rate pattern throughout gestation. Ultrasound Obstet Gynecol 2009;34(1):74-9.
35. Soncini E, Ronzoni E, Macovei D, Grignaffini A. Integrated monitoring of fetal growth restriction by computerized cardiotocography and Doppler flow velocimetry. Eur J Obstet Gynecol Reprod Biol 2006;128(1-2):222-30.
36. Visser GH, Dawes GS, Redman CW. Numerical analysis of the normal human antenatal fetal heart rate. Br J Obstet Gynaecol 1981;88(8):792-802.

**Figure 1: Literature retrieval process**

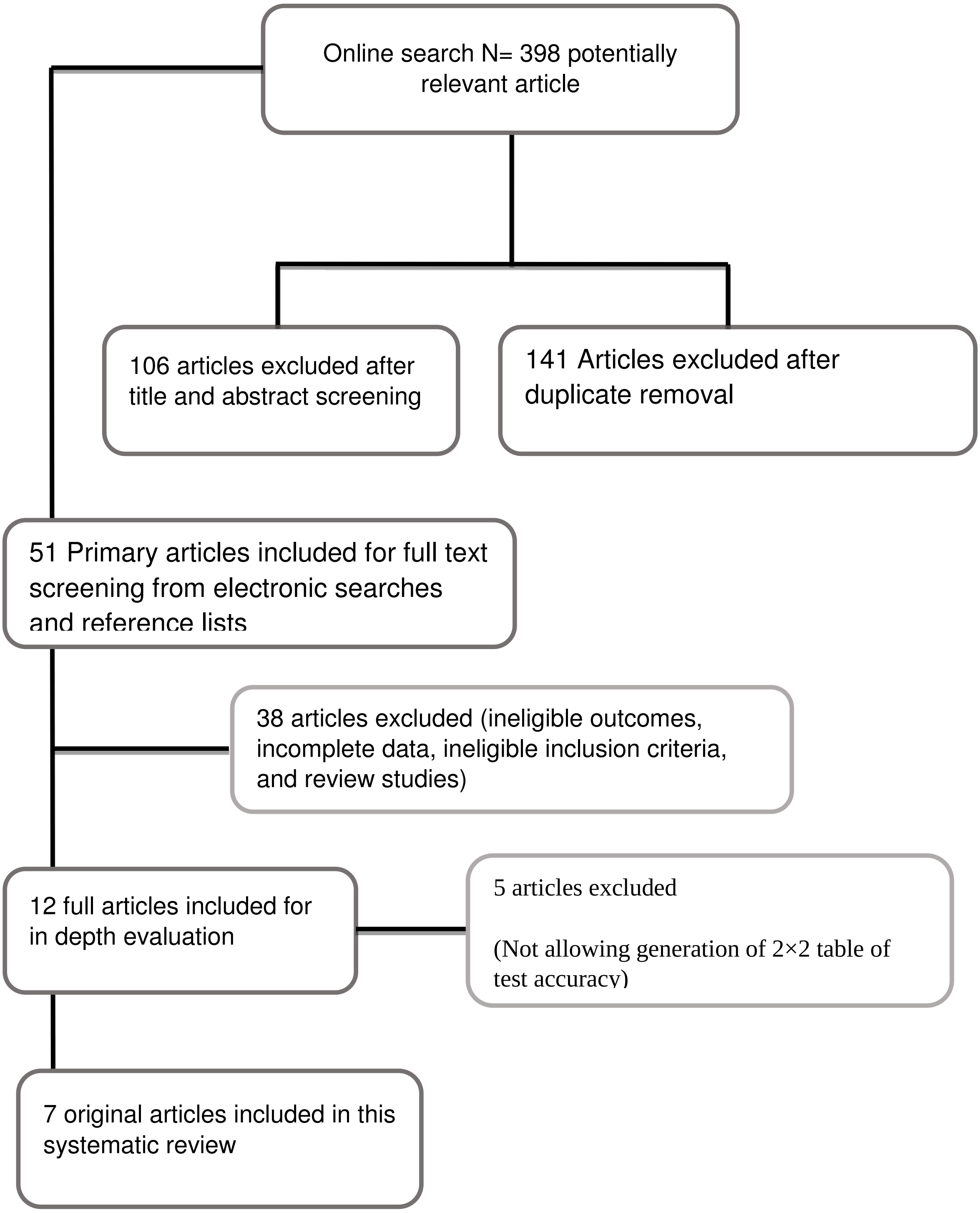


**Figure 2: Bar chart showing quality of evidence on STV in predicting foetal acidemia**

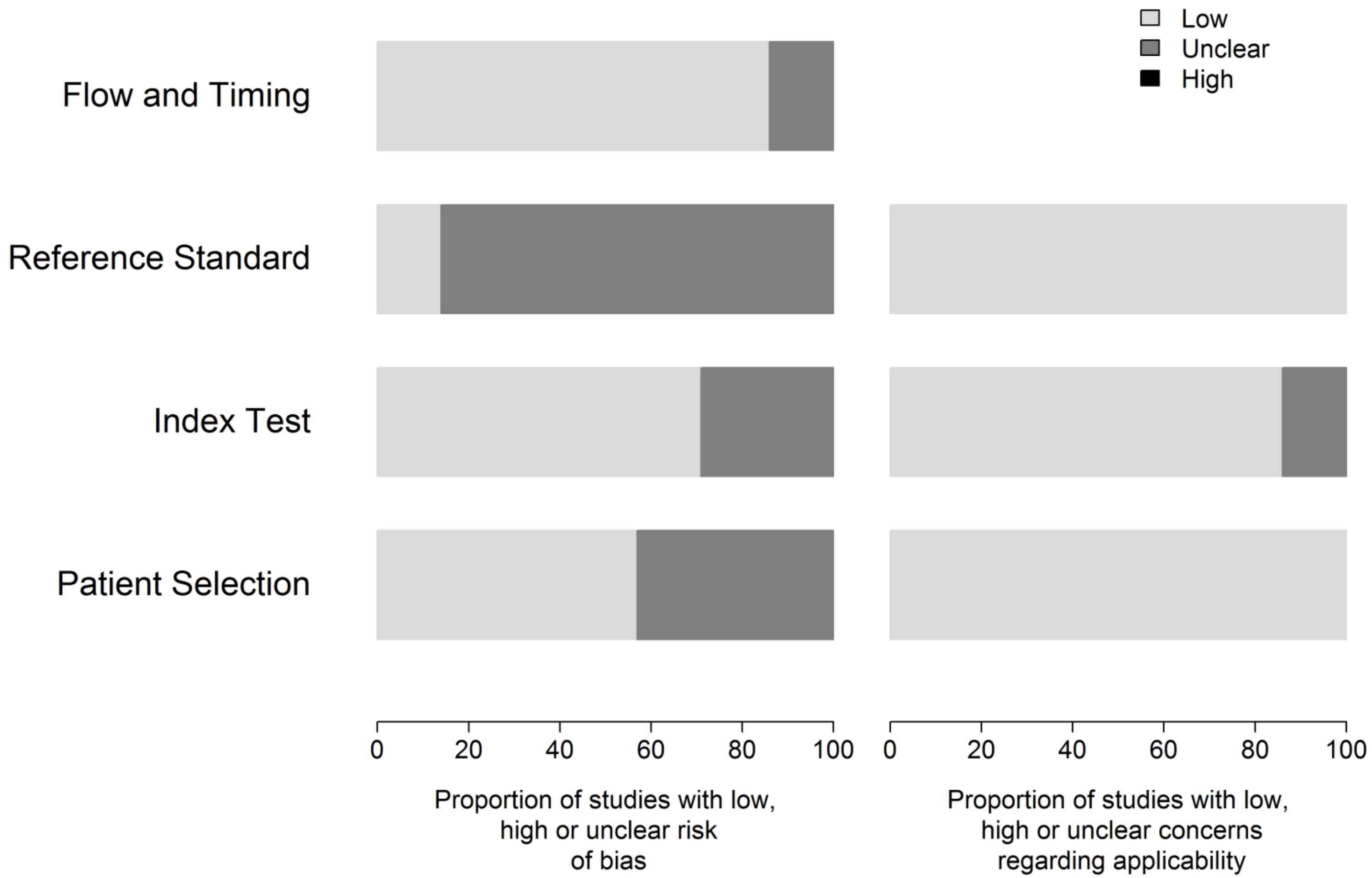
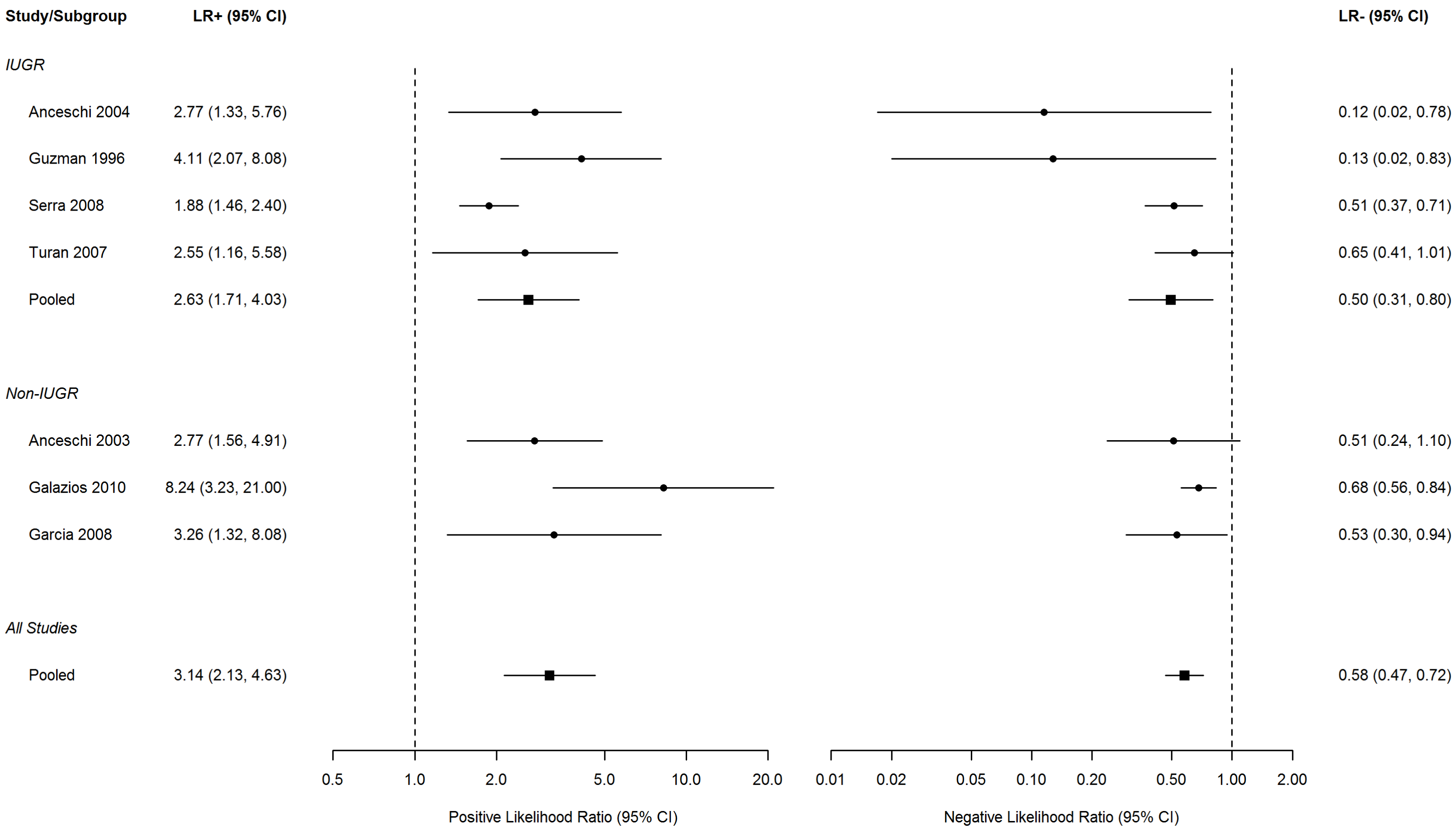


**Figure 3: Forest Plot of STV to predict foetal acidemia with sub-group analysis.**




LR+, positive likelihood ratio; LR-, negative likelihood ratio; CI, confidence interval

Squares represent pooled results, circles represent individual studies.

**Figure 4: Summary Receiver Operating Characteristic (ROC) Curves for the bivariate analysis of the accuracy of STV to predict foetal acidemia.**

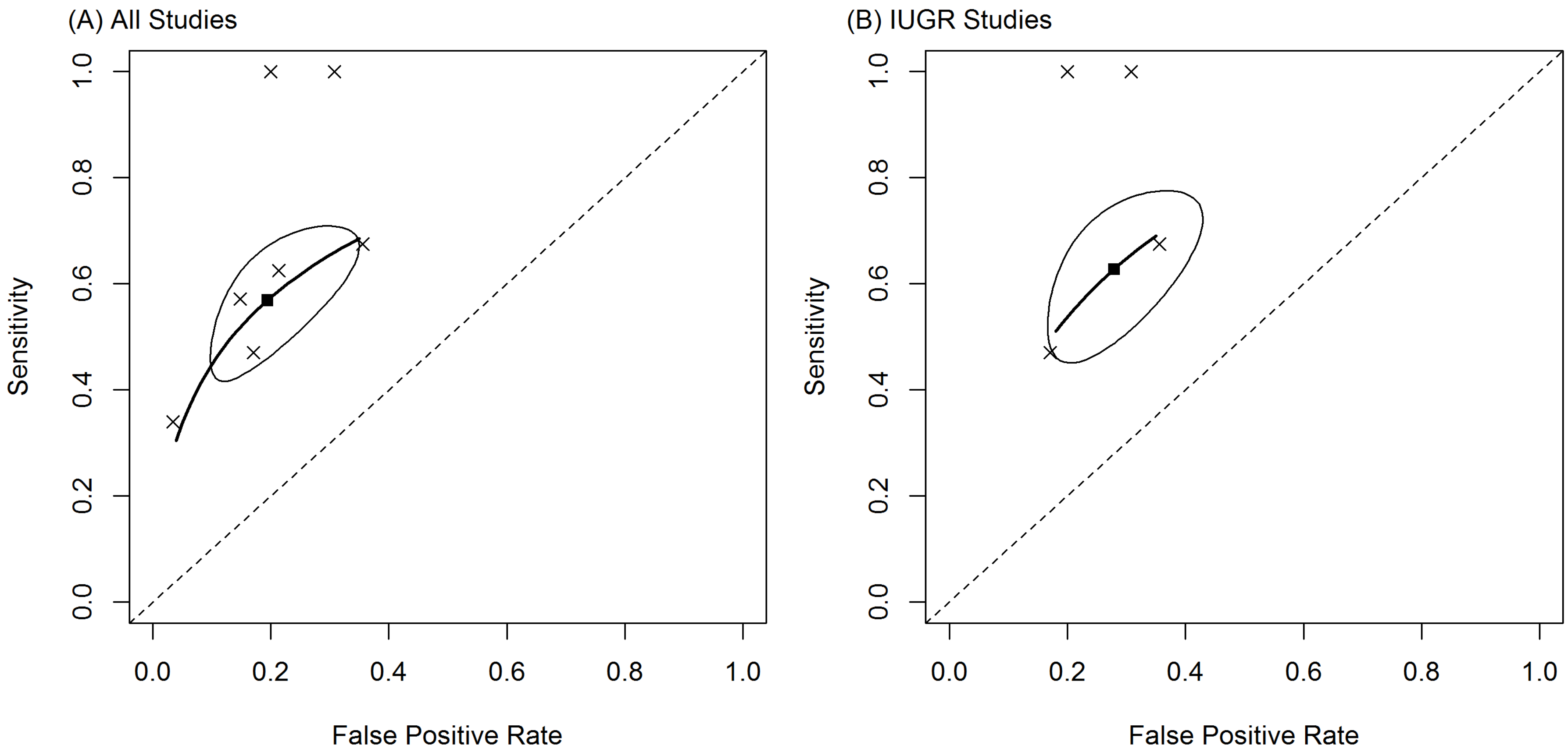


**Table 1: Studies included in the systematic review of association between short-term variation in fetal heart rate and fetal acidaemia**

| Study (First author and year) | Study design | Population studied | No of women analysed | GA at the time of test (weeks) | Details of index test | Umbilical artery acidaemia |
|---|---|---|---|---|---|---|
| Anceschi et al 2003 | Consecutive | Low and high risk population | N=195<br>Divided into subgroups according to GA<br><34 weeks (n=31), 35-37 weeks (n=37) and >37 weeks (n=127) | 26-42 | cCTG performed for 40 min, 4 hrs prior to elective-C-section | pH<7.00 |
| Anceschi et al 2004 | Not defined | High risk with IUGR fetuses | N=24 | 24-35 | cCTG performed for 40 min, 2 hrs prior to C-section | pH<7.00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Galazios et al 2010 | Retrospective | Low-risk population | N=167 | 38-40 | No mention of how long CTG was performed and no record of how long before delivery, | pH<7.25 |
| Garcia et al 2008 | Cross-sectional | High risk with hypertensive disorder | N=41 | 27-41 | Duration 20 min, 24 hrs of C-section | pH<7.20 |
| Guzman et al 1996 | Consecutive | High risk with IUGR fetuses | N=38 | 26-37 | Duration of recording 1 hr within 4 hrs of C-section | pH<7.20 |
| Serra et al 2008 | Retrospective | High risk with IUGR fetuses | N=257 | 26-42 | cCTG performed for 60 min within 24 hrs of delivery | pH<7.20 for pre-labour C-section and pH<7.12 for vaginal birth and emergency C-section |
| Turan et al 2007 | Prospective | High risk with IUGR fetuses | N=58 | 26-38 | Minimum duration of recording was 30 min on the day of C-section | pH<7.20 |

**Table 2: Analysis on STV in predicting fetal acidaemia for individual studies**

| Study | pH | STV | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Positive Likelihood Ratio (95% CI) | Negative Likelihood Ratio (95% CI) |
|---|---|---|---|---|---|---|---|---|---|---|
| Anceschi, 2003 | 7.00 | 5.10 | 5 | 40 | 3 | 147 | 0.63 (0.31, 0.86) | 0.79 (0.72, 0.84) | 2.77 (1.56, 4.91) | 0.51 (0.24, 1.10) |
| Anceschi, 2004 | 7.00 | 4.50 | 11 | 4 | 0 | 9 | 1.00 (0.74, 1.00) | 0.69 (0.42, 0.87) | 2.77 (1.33, 5.76) | 0.12 (0.02, 0.78) |
| Galazios, 2010 | 7.25 | 5.00 | 17 | 4 | 33 | 113 | 0.34 (0.22, 0.48) | 0.97 (0.92, 0.99) | 8.24 (3.23, 21.00) | 0.68 (0.56, 0.84) |
| Garcia, 2010 | 7.20 | 5.25 | 8 | 4 | 6 | 23 | 0.57 (0.33, 0.79) | 0.85 (0.68, 0.94) | 3.26 (1.32, 8.08) | 0.53 (0.30, 0.94) |
| Guzman, 1996 | 7.20 | 3.50 | 8 | 6 | 0 | 24 | 1.00 (0.68, 1.00) | 0.80 (0.63, 0.91) | 4.11 (2.07, 8.18) | 0.13 (0.02, 0.83) |
| Serra, 2008 | 7.20 | 4.70 | 54 | 63 | 26 | 114 | 0.68 (0.57, 0.77) | 0.64 (0.57, 0.71) | 1.88 (1.46, 2.40) | 0.51 (0.37, 0.71) |
| Turan, 2007 | 7.20 | 2.50 | 8 | 7 | 9 | 34 | 0.47 (0.26, 0.69) | 0.83 (0.69, 0.92) | 2.55 (1.16, 5.58) | 0.65 (0.41, 1.01) |

STV, short-term variation; TP, true positives; FP, false positives; FN, false negatives; TN, true negatives; CI, confidence interval