AOGS
ACTA Obstetricia et Gynecologica Scandinavica

AOGS REVIEW ARTICLE

A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part I: the randomized controlled trials

PER OLOFSSON[1], DIOGO AYRES-DE-CAMPOS[2], JÖRG KESSLER[3,4], BRITTA TENDAL[5], BRANKA M. YLI[6] & LAWRENCE DEVOE[7]

[1]Department of Obstetrics and Gynecology, Institution of Clinical Sciences, Skåne University Hospital, Lund University, Malmö, Sweden, [2]Department of Obstetrics and Gynecology, Medical School – S. Joao Hospital, Institute of Biomedical Engineering, Porto University, Porto, Portugal, [3]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway, [4]Department of Clinical Sciences, Clinical Fetal Physiology Research Group, Bergen University, Bergen, Norway, [5]Danish Health and Medicines Authority, Copenhagen, Denmark, [6]Delivery Department, Mother and Child Clinic, Oslo University Hospital, Oslo, Norway, and [7]Department of Obstetrics and Gynecology, Medical College of Georgia, Georgia Regents University, Augusta, Georgia, USA

Key words
Cardiotocography, fetal surveillance, meta-analysis, metabolic acidosis, randomized controlled trial, ST analysis

Correspondence
Per Olofsson, Department of Obstetrics and Gynecology, Skåne University Hospital, S-20502 Malmö, Sweden.
E-mail: per.olofsson@med.lu.se

Conflicts of interest
Per Olofsson was co-author of the Swedish RCT and has cooperated with FBS equipment sales companies in Sweden and Denmark (Medexa Medicinsk Service AB, LiNA Medical A/S) and with the STAN® manufacturer Neoventa Medical AB, where he is currently a consulting Global Medical Adviser. Jörg Kessler once received a lecture fee from Neoventa Medical AB. Branka M Yli has taught at ST analysis courses arranged by SCAN-MED A/S, Norway. Lawrence Devoe is a paid US Medical Adviser to Neoventa Medical AB. Diogo Ayres-de-Campos and Britta Tendal have no conflicts of interest to declare.

Please cite this article as: Olofsson P, Ayres-de-Campos D, Kessler J, Tendal B, Yli BM, Devoe L. A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part I: the randomized controlled trials. Acta Obstet Gynecol Scand 2014; 93:556–569.

Received: 16 February 2014
Accepted: 30 April 2014

DOI: 10.1111/aogs.12413

Abstract

We reappraised the five randomized controlled trials that compared cardiotocography plus ECG ST interval analysis (CTG+ST) vs. cardiotocography. The numbers enrolled ranged from 5681 (Dutch randomized controlled trial) to 799 (French randomized controlled trial). The Swedish randomized controlled trial (n = 5049) was the only trial adequately powered to show a difference in metabolic acidosis, and the Plymouth randomized controlled trial (n = 2434) was only powered to show a difference in operative delivery for fetal distress. There were considerable differences in study design: the French randomized controlled trial used different inclusion criteria, and the Finnish randomized controlled trial (n = 1483) used a different metabolic acidosis definition. In the CTG+ST study arms, the larger Plymouth, Swedish and Dutch trials showed lower operative delivery and metabolic acidosis rates, whereas the smaller Finnish and French trials showed minor differences in operative delivery and higher metabolic acidosis rates. We conclude that the differences in outcomes are likely due to the considerable differences in study design and size. This will enhance heterogeneity effects in any subsequent meta-analysis.

Abbreviations: BD, base deficit; BD_{blood}, base deficit in blood; BD_{ecf}, base deficit in extracellular fluid; CI, confidence interval; CTG, cardiotocography; FBS, fetal scalp blood sampling; ITT, intention-to-treat; MA, meta-analysis; NICU, neonatal intensive care unit; ODFD, operative delivery for fetal distress; OR, odds ratio; RCT, randomized controlled trial; RR, risk ratio.

Introduction

From 2012 to 2013, five independent meta-analyses (MAs) of the value of intrapartum fetal surveillance with cardiotocography (CTG) plus ST interval analysis (CTG+ST) of the electrocardiogram compared with CTG alone were published (1–5). In the same time frame, international clinical experience with the CTG+ST analysis method increased, as noted in numerous observational studies (6–11). With the current attention focused on the CTG+ST fetal surveillance method, we believe that a thorough review of the quality of the original five randomized controlled trials (RCTs) (12–16) was warranted, as well as a critical review of the quality of the five MAs that have followed (1–5). This article addresses the quality of the RCTs, while a companion piece will focus on the MAs.

Five randomized controlled trials

Our review addressed the following issues in the RCTs: (i) power calculations; (ii) pre-study training, inclusion criteria, randomization and recruitment pace; (iii) intrapartum management protocols; (iv) intrapartum interventions; (v) cord blood and early neonatal metabolic acidosis; (vi) neonatal outcomes. For supplementary statistical calculations, we used the MedCalc® version 5.00.017 computer software (MedCalc Software, Mariakerke, Belgium). Two-sided statistics were performed with a p-value <0.05 considered significant.

The first RCT on CTG+ST analysis vs. CTG alone, the Plymouth trial, was published in 1993 (12), followed by the Swedish trial in 2001 (13), the Finnish trial in 2006 (14), the French trial in 2007 (15), and the Dutch trial in 2010 (16). After receiving criticism about quality control, revised data from the Swedish and Dutch RCTs were published in 2011 (17–19). Metabolic acidosis data from the Finnish RCT have been revised (see below), but data from the Plymouth and French RCTs have not been revised.

Power calculations of outcome variables

Neonatal metabolic acidosis was the primary outcome variable in the Plymouth, Swedish and Dutch RCTs but not in the Finnish and French RCTs. Table 1 shows that neither the Finnish nor French RCTs were adequately powered to address this outcome. Due to a 46.5 and 73% lower than expected incidence of metabolic acidosis in the Plymouth and the Dutch RCTs, respectively, these trials were also found to be underpowered for this outcome (Table 1). Therefore, the majority of the RCTs failed to attain their recruitment goals and/or were underpowered for their primary outcome. Estimation of the incidences of the primary endpoints was accurate in the control group (i.e. in the CTG-alone group) only in the Swedish RCT (metabolic acidosis) and in the Plymouth RCT (operative delivery for fetal distress, ODFD).

Pre-trial training, inclusion criteria, randomization and recruitment pace

While all RCTs offered some form of pre-trial training (Table 2), the Plymouth RCT included a 100-case test period before enrollment. The Swedish RCT required a 2-month practice period before enrollment started and there was re-training during the trial. The Dutch RCT required certification and a 2-month practice period before enrollment started. It therefore seems that these three RCTs dealt more carefully with the potential problem of staff proficiency in applying the ST analysis methodology to clinical care, thereby increasing the likelihood of its proper use.

An interim analysis or safety committee watch was performed in all but the Finnish and French RCTs (Table 2). The inclusion criteria differed among the RCTs. The most important difference was noted in the French RCT, as the investigators recruited only cases considered to have suspicious or pathological CTGs (86% of enrollees) or thick meconium-stained amniotic fluid (7%), or both (7%) at the start-up of monitoring. Cases with a normal CTG and no decelerations were excluded. However, the inclusion of patients with a pathological CTG at start-up of ST analysis violated the CTG+ST analysis clinical guidelines (20). To enable establishment of a fetal electrocardiogram T/QRS ratio baseline, ST monitoring should be initiated while the fetus is still well oxygenated and the CTG is not pathological. The French RCT data included in the MA by Schuit et al. (5) indicate that in several cases deterioration of the fetal condition might have already occurred before enrollment. The basis for this requirement is that alerts for changes in the ST interval, signaling fetal hypoxia and impending metabolic acidosis, may not occur if the fetal condition has already deteriorated and myocardial reserve is exhausted. Consequently, the prerequisites for use of the CTG+ST analysis method were not fulfilled in many cases in the French RCT.

Key Message

Among the randomized controlled trials, the Plymouth, Swedish and Dutch trials have the most similar design and therefore should be the main source of information regarding the effectiveness of CTG+ST analysis for fetal surveillance in labor.

Table 1. Power calculations and recruitment success in the five RCTs on fetal electrocardiogram ST interval analysis vs. CTG alone. A reduction/increase was calculated as (|1.0 − (CTG+ST rate)/(CTG rate)|) from figures provided in the articles.

Trial	Plymouth RCT Westgate et al. (1993)	Plymouth RCT Westgate et al. (1993)	Swedish RCT Amer-Wåhlin et al. (2001, 2011)	Finnish RCT Ojala et al. (2006)	French RCT Vayssière et al. (2007)	Dutch RCT Westerhuis et al. (2010, 2011)
Primary outcome variable (power calculation)	Metabolic acidosis in extracellular fluid (one of two primary outcomes)	Operative delivery for fetal distress (one of two primary outcomes)	Metabolic acidosis in extracellular fluid	Umbilical artery pH < 7.10	Operative delivery for nonreassuring fetal status	Metabolic acidosis in extracellular fluid
Secondary outcome variables	FBS, Apgar score, neonatal resuscitation, NICU admission	See first column	Operative delivery for fetal distress, Apgar score, NICU admission, neonatal encephalopathy Sarnat & Sarnat grade 1–3	FBS, operative intervention, umbilical artery pH < 7.05, metabolic acidosis in blood	FBS, total rate operative deliveries, metabolic acidosis in extracellular fluid, Apgar score, NICU admission, neonatal convulsions	FBS, operative delivery, metabolic acidosis in blood, Apgar score, NICU admission, neonatal encephalopathy Sarnat & Sarnat grade 2–3
Beta	0.50	0.10	0.20	0.20	0.20	0.20
Alpha	0.05	0.05	0.05	0.05	0.05	0.05
Estimated difference	50% reduction from 2%	50% reduction from 11%	70% reduction from 1.3%	50% reduction from 6.4%	10% reduction from 50% and 40%, respectively	40% reduction from 3.5%
Estimated recruitments needed	1300 in each arm	450 in each arm	1600 in each arm	761 in each arm	400 in each arm	2319 in each arm
Obtained recruitments: CTG+ST vs. CTG	1219 vs. 1215	See first column	2519 vs. 2447 eligible randomized; 2228 vs. 2164 after exclusions (original data); 2565 vs. 2484 all randomized (revised data)	1483 randomized; 733 vs. 739 after exclusions; 714 vs. 722 for neonatal outcome data	399 vs. 400, total 799	2832 vs. 2849; 2827 vs. 2840 after exclusions
Obtained difference	62.6% reduction from 1.07%	45.1% reduction from 9.1%	50.4% reduction from 1.41% (imputed data)	21.3% increase from 4.7%	9.2% reduction from 37.0%	29.5% reduction from 0.95% (revised data)
Reasons for failure to show	Prevalence 46.5% lower than expected, study underpowered, recruitment not fulfilled	No failure	No failure	Prevalence higher in CTG+ST group (4.7%), recruitment not fulfilled	Prevalence 26% (or 9.3%) lower than expected, study underpowered	Prevalence 73% lower than expected, study underpowered
Statistics	Chi-squared test; Student's *t*-test; Kruskal–Wallis; Mann–Whitney; OR with 95% confidence interval	See first column	Chi-squared test; Fisher's exact test; relative risk with 95% confidence interval; imputation for missing data and neonatal data added in revised version	Variance and relative risk with 95% confidence interval	Chi-squared test; Mann–Whitney; relative risk with 95% confidence interval; Bayesian model averaging	Relative risk with 95% confidence interval adjusted for stratification variables (multivariable regression analysis); number needed to treat; multiple imputation method

RCTs, randomized controlled trials; CTG, cardiotocography; FBS, fetal scalp blood sampling for determination of pH; NICU, neonatal intensive care unit.

Table 2. Study performance.

Trial	Plymouth RCT	Swedish RCT	Finnish RCT	French RCT	Dutch RCT
Pre-study training	Yes, with 100 cases	Yes, during 2 months, certification of users	Yes, but time not reported	Yes, but time not reported	Yes, at least 2 months, certification of users
Interim analysis	Yes, after 1200 included cases	Yes, after 1600 included cases	No information	Not planned	Serious events monitored by Safety Committee
Inclusion criteria	>34 weeks, high-risk with indication for continuous CTG, breech included	≥36 weeks, scalp electrode decided because of increased risk	≥36 weeks, amniotomy decided	≥36 weeks, suspicious or pathological CTG, thick meconium	≥36 weeks, high-risk needing CTG monitoring, abnormal/ nonreassuring CTG accepted after normal FBS
Exclusion criteria	Gross fetal abnormality	Multiple pregnancy, non-cephalic, no indication for scalp electrode	Scalp electrode contraindicated, multiple pregnancy, non-cephalic presentation, start-up in second stage of labor	Multiple pregnancy, non-cephalic, cardiac malformation, contraindication scalp electrode, normal CTG with no decelerations, severely abnormal CTG at arrival	<18 years, multiple, non-cephalic, no indication for scalp electrode
CTG-only group monitoring (internal/ external)	Conventional fetal heart rate monitors, internal monitoring	STAN S21 prototypes (ST data blinded), internal monitoring	Conventional fetal heart rate monitors, internal or external monitoring	Conventional fetal heart rate monitors, internal or external monitoring not specified	Conventional fetal heart rate monitors, internal monitoring
Randomization	Sealed envelopes	Allocation by STAN monitor at start-up	Sealed envelopes	Sealed envelopes	Web-based computer program, stratified for center and parity
No. randomized	Data cannot be extracted from article	5049	1483	Data cannot be extracted from article	5681
No. in ITT analysis	2434	4966 (original data) 5049 (revised data)	1472 (1436 neonatal outcome, exclusions due to missing cord blood gas data)	799	5667
Type of ITT	Modified (only cases with full cord blood gas panel included?)	Standardized for metabolic acidosis (all randomized cases included) and modified (non-eligible cases excluded)	Modified (exclusions: protocol violations, missing patient records, study withdrawals; neonatal outcome: only cases with full cord blood gas data)	Modified (only cases with full cord blood gas panel included?)	Modified (14 non-eligible cases excluded, representing 0.25% of the series)
Number of centers	1	3	1	2	9
Months of study	18	18	14	27	30
Recruitments per center and month	135	94	106	15	21
Percent of total population included in study	36%	33%	33%	8%	Data cannot be extracted from article

CTG, cardiotocography; FBS, fetal scalp blood sampling for determination of pH; ITT, intention-to-treat; RCT, randomized controlled trials.

In the Plymouth, Swedish and Dutch RCTs, cases at increased risk of fetal hypoxia in labor were recruited (Table 2). Inclusion of cases with a nonreassuring fetal heart rate was also allowed in these RCTs, but the majority of cases recruited in the French RCT were at considerably higher risk. This is illustrated by the crude French RCT data presented in the MA by Schuit et al. (5), where a composite adverse outcome was two to four times more common in the French RCT than in the other RCTs. In the Finnish RCT recruitment was made consecutively after amniotomy. It is unclear whether amniotomy was a routine procedure for active management of labor or performed for specific indications.

The monitoring techniques in the CTG arm of the RCTs varied (Table 2). Only the Swedish RCT used STAN® S21 monitors (Neoventa Medical AB, Göteborg, Sweden) for both study groups, while the other RCTs used different monitoring systems, allowed external CTG recording or did not specify the methodology used in the control groups. In general, external fetal heart rate monitoring provides inferior CTG signal quality when compared with internal signals (21). This might have introduced bias in the CTG-only group in the RCTs that allowed this modality.

Robust methods for allocation of women were used in all five RCTs but standardized intention-to-treat (ITT) analyses, including patients later excluded for various reasons, were not generally available (Table 2). A revised report of the Swedish RCT that included all randomized cases (17) was the only effort to address this issue. However, the Dutch RCT excluded only 14 cases (0.25%) from the ITT.

Leipälä et al. (22) requested an explanation of why the Swedish trial was revisited. The original Swedish RCT (13) was reported as a per-protocol analysis and by modified ITT analysis, i.e. including only the 4966 eligible cases. After criticism for not including all randomized cases and for misclassification of cases, a so-called standardized ITT on the primary outcome parameter metabolic acidosis was performed (17). In the standardized ITT analysis (n = 5049), 83 allocated cases that did not fulfill the inclusion criteria were added. These were cases of delivery before 36 weeks of gestation, breech delivery, malformations, and cases included after the trial was closed (23). Furthermore, the standardized ITT analysis was extended to include not only validated umbilical cord blood samples, but all cord blood determinations performed, neonatal blood tests showing metabolic acidosis, and imputed data in cases with missing acid–base data.

Table 2 also indicates that there were considerable variations in duration, enrollment pace and inclusion criteria among trials. The Plymouth, Swedish and Finnish RCTs ran for a period of 18 months or less with 94–135 enrollments/center/month, whereas the French and Dutch RCTs took 2 and 2½ years to complete, respectively, with an enrollment pace of 15–21 patients/center/month. However, the Dutch RCT adjusted statistically for the stratified randomization by center (and parity). Less frequent use of the ST analysis methodology may have contributed to a slower gain in staff experience, so influencing clinical decisions and possibly affecting the study results. Differences in population sample size in relation to the total population were also considerable. These differences raise the concern, particularly in the French RCT, that the characteristics of the enrolled population contributed to outcomes that differed in degree and direction from those of the other RCTs.

Management protocols in labor

Fetal scalp blood sampling (FBS) for determination of pH was optional in all five RCTs but only the Plymouth and Dutch RCTs had guidelines for this (Table 3). The interpretation algorithm was similar in all five RCTs (scalp blood pH ≥ 7.25, normal; 7.24–7.20, suspicious/pre-acidosis, repeat FBS; <7.20, abnormal/acidosis, deliver or reveal cause of hypoxia; in the Finnish RCT a pH <7.20 was an indication of immediate delivery). The Plymouth RCT used an earlier version of the STAN® monitor, the S8801 model, and a CTG+ST interpretation algorithm that differed from the other RCTs (Table 3). The T/QRS ratio and ST interval changes were read manually, where a T/QRS ratio above a certain cut-off or a rapidly emerging change in the ST interval waveform was regarded as significant and warranted action (12,24). This situation may have reduced the reliability of the ST analysis when compared with the automated analysis used in the later RCTs.

Intrapartum interventions: FBS and operative delivery

The use of FBS ranged from 9.4 to 62% in the CTG-only groups (Table 4). FBS was reduced in the CTG+ST arm in all RCTs, but the reduction was only significant in those with the highest use of FBS, i.e. the Finnish, French and Dutch RCTs. The total rates of ODFD were reduced in the CTG+ST arm in all RCTs except the Dutch study, but the reductions were only significant in the Plymouth and Swedish RCTs (Table 4). Overall, the total operative delivery rate, including cesarean sections and instrumental vaginal deliveries, was significantly reduced only in the Swedish RCT.

Metabolic acidosis as an outcome parameter

Umbilical cord blood acid–base status at birth was an outcome parameter in all five RCTs (Table 5). Paired

2014 The Authors. *Acta Obstetricia et Gynecologica Scandinavica* published by John Wiley & Sons on behalf of Nordic Federation of Societies of Obstetrics and Gynecology. **93** (2014) 556–569

Table 3. Management protocols in labor, umbilical cord blood acid-base characteristics.

Trial	Plymouth RCT	Swedish RCT	Finnish RCT	French RCT	Dutch RCT
Fetal scalp blood sampling	Guidelines related to the CTG in both groups	Optional	Optional	Optional	Guidelines in ST group, optional in CTG group
ST analysis interpretation algorithm	Fixed T/QRS ratio cut-offs (>0.24 > 30 min; >0.5 > 15 min), ST changes >5 min	Progressive T/QRS ratio increases, ST changes	Progressive T/QRS ratio increases, ST changes	Progressive T/QRS ratio increases, ST changes	Progressive T/QRS ratio increases, ST changes
Cord blood samples	Artery + vein	Artery + vein	Artery + vein	Artery + vein	Artery + vein
Validation of cord blood samples[a]	No	V-A pH ≥ 0.3, A-V P_{CO_2} ≥ 1.0 kPa (revised article)	No	A-V P_{CO_2} > 0.5 kPa	V-A pH ≥ 0.3
Base deficit algorithm compartment	Extracellular fluid	Extracellular fluid	Blood	Extracellular fluid	Extracellular fluid and blood

CTG, cardiotocography; RCT, randomized controlled trials.
[a]V, umbilical cord vein; A, umbilical cord artery; V-A, venous-to-arterial difference; A-V, arterial-to-venous difference.

cord blood samples from artery and vein were used, but validation of arterial blood samples was only performed in the Swedish (revised data report), French and Dutch RCTs. However, the validation criteria were not uniform (Table 3). Base deficit (BD) was calculated in the extracellular fluid (BD_{ecf}) from measured values of pH and P_{CO_2} in all trials except the Finnish RCT, in which BD was calculated in blood (BD_{blood}). Hence, the incidence of metabolic acidosis in the Finnish RCT cannot be compared directly with that of the other RCTs, because, when calculated in blood, BD is considerably higher and metabolic acidosis subsequently more prevalent than when calculated in extracellular fluid (25,26).

Metabolic acidosis was defined as an umbilical cord artery blood pH < 7.05 in combination with a BD_{ecf} > 12.0 mmol/L in all but the Finnish RCT. Using the pH and P_{CO_2} values obtained from blood gas analyzers, BD_{ecf} can be calculated post hoc with the algorithm (in SI units): $BD_{ecf} = -0.9149 \times (0.23 \times P_{CO_2} \times 10^{[pH-6.1]} - 24.1 + 16.21 \times [pH - 7.4])$ (25). This algorithm was originally derived from the work by Siggaard-Andersen (27,28). The Finnish RCT used a Chiron Diagnostics 348 blood gas analyzer to calculate BD_{blood}. By recalculating the Finnish RCT data with the BD_{ecf} algorithm, the originally reported metabolic acidosis rates of 1.7% (12/714) in the CTG+ST analysis group and 0.7% (5/722) in the CTG-only group declined to 0.8% (6/714) and 0.6% (4/722), respectively [K. Ojala, personal communication to Welin et al. (7)]. The distinction between BD calculated in blood and that calculated in extracellular fluid is important for the diagnosis of metabolic acidosis, since the incidence of BD > 12.0 mmol/L might differ by a factor of 4 when using different BD algorithms (26). Hence, for a correct comparison of BD values and metabolic acidosis rates, the same BD algorithm must be used in comparative studies and MAs.

In the perinatal period, BD_{ecf} should be used rather than BD_{blood} for determining metabolic acidosis, because the fetus/newborn has a relative increase in the size of the extracellular fluid compartment compared with that of the intravascular compartment (29–31). This makes BD_{ecf} more stable and less susceptible to momentary perturbations. The impact of different BD calculations was demonstrated in the Dutch RCT (19), showing a significant reduction in metabolic acidosis rate in the CTG+ST group with the BD_{blood} algorithm [risk ratio (RR) 0.63, 95% CI 0.42–0.94] but not with the BD_{ecf} algorithm (RR 0.70, 95% CI 0.38–1.28). With BD_{blood} the metabolic acidosis rates were 1.6 and 2.6%, and with BD_{ecf} they were 0.7 and 1.1%.

Neonatal outcome: metabolic acidosis, neonatal intensive care admissions

The proportion of missing cord blood gas data was available only in the Swedish (7.4%) and Finnish (2.4%) publications (Table 5), but the Dutch authors reported an estimated incidence of 20% missing values (16). Imputed data were calculated in the Dutch RCT and in the revised version of the Swedish RCT. It can be inferred from the thesis of Westgate (24) that cases with missing cord blood gas data were excluded from the analyses of neonatal variables, and possibly also of other variables in the Plymouth RCT. Thirty-six cases (2.4%) with missing blood gas data were excluded from analyses of neonatal outcome variables in the Finnish RCT [comparative data retrieved from Becker et al. (1)] and apparently cases with missing cord blood data were excluded from the ITT analyses in the French RCT.

Westgate et al. (12) presented the results of the Plymouth RCT as the OR of CTG alone vs. CTG+ST. For comparison with the other RCT results we recalculated these

Table 4. Intrapartum interventions. Statistics represent CTG+ST group vs. CTG-only group.

Trial	Plymouth RCT[a]	Swedish RCT	Finnish RCT	French RCT	Dutch RCT
FBS	93/1219 vs. 114/1215 RR 0.81 (0.63–1.06)[b]	234/2519 vs. 261/2447 RR 0.87 (0.74–1.03)	51/733 vs. 115/739 RR 0.45 (0.33–0.61)	108/399 vs. 248/400 RR 0.44 (0.36–0.52)	301/2827 vs. 578/2840 RR 0.52 (0.46–0.59)
Spontaneous vaginal delivery (not calculated in original RCTs)	875/1219 vs. 832/1215 RR 1.05 (0.995–1.10)[b]	2065/2519 vs. 1947/2447 RR 1.03 (1.003–1.059)[b]	616/733 vs. 625/739 RR 0.99 (0.95–1.04)[b]	183/399 vs. 179/400 RR 1.02 (0.88–1.19)[b]	2038/2827 vs. 2018/2840 RR 1.01 (0.98–1.05)[b]
Operative delivery, total (not calculated in all RCTs)	344/1219 vs. 383/1215 RR 0.90 (0.79–1.01)[b]	454/2519 vs. 500/2447 RR 0.88 (0.79–0.99)[b]	117/733 vs. 114/739 RR 1.03 (0.82–1.31)[b]	216/399 vs. 221/400 RR 0.98 (0.86–1.11)	789/2827 vs. 822/2840 RR 0.96 (0.87–1.06)
Cesarean section, total	Data cannot be extracted from article	210/2519 vs. 222/2447 RR 0.92 (0.77–1.10)[b]	47/733 vs. 35/739 RR 1.35 (0.86–2.07)	Data cannot be extracted from article	405/2827 vs. 391/2840 RR 1.02 (0.89–1.17)
Operative vaginal delivery, total	Data cannot be extracted from article	244/2519 vs. 278/2447 RR 0.85 (0.72–1.003)[b]	70/733 vs. 79/739 RR 0.89 (0.66–1.21)[b]	Data cannot be extracted from article	384/2827 vs. 431/2840 RR 0.90 (0.79–1.03)
Operative delivery for fetal distress, total	61/1219 vs. 111/1215 RR 0.55 (0.40–0.74)[b]	193/2519 vs. 227/2447 RR 0.83 (0.69–0.99)	51/733 vs. 63/739 RR 0.82 (0.57–1.16)[b]	134/399 vs. 148/400 RR 0.91 (0.75–1.10)	261/2827 vs. 237/2840 RR 1.10 (0.93–1.31)
Cesarean section for fetal distress	15/1219 vs. 30/1215 RR 0.50 (0.27–0.92)[b]	87/2519 vs. 97/2447 RR 0.87 (0.65–1.16) 43/2228 vs. 63/2164 RR 0.66 (0.45–0.97)[c]	15/733 vs. 15/739 RR 1.01 (0.50–2.05)	54/399 vs. 65/400 RR 0.83 (0.60–1.16)[b]	91/2827 vs. 70/2840 RR 1.31 (0.96–1.79)
Operative vaginal delivery for fetal distress	46/1219 vs. 81/1215 RR 0.57 (0.40–0.81)[b]	106/2519 vs. 130/2447 RR 0.79 (0.62–1.02)	36/733 vs. 48/739 RR 0.76 (0.50–1.15)	80/399 vs. 83/400 RR 0.97 (0.73–1.27)[b]	170/2827 vs. 167/2840 RR 1.02 (0.83–1.27)
Cesarean section for failure to progress/other reasons	Data cannot be extracted from article	123/2519 vs. 125/2447 RR 0.96 (0.75–1.22)[b]	32/733 vs. 20/739 RR 1.61 (0.93–2.79)[b]	Data cannot be extracted from article	314/2827 vs. 321/2840 RR 0.96 (0.58–1.61)
Operative vaginal delivery for failure to progress/other reasons	Data cannot be extracted from article	138/2519 vs. 148/2447 RR 0.91 (0.72–1.13)[b]	34/733 vs. 31/739 RR 1.11 (0.69–1.78)[b]	Data cannot be extracted from article	214/2827 vs. 264/2840 RR 0.82 (0.69–0.98)
Total operative delivery for failure to progress/other reasons	283/1219 vs. 272/1215 RR 1.04 (0.90–1.20)[b]	261/2519 vs. 273/2447 RR 0.93 (0.79–1.09)	66/733 vs. 51/739 RR 1.30 (0.92–1.85)[b]	82/399 vs. 73/400 RR 1.13 (0.85–1.49)[b]	528/2827 vs. 585/2840 RR 0.91 (0.81–1.02)

CTG, cardiotocography; CTG+ST, cardiotocography plus ECG ST interval analysis; FBS, fetal scalp blood sampling for determination of pH; ITT, intention-to-treat; RCT, randomized controlled trials; RR, relative risk.

[a]Odds ratio CTG vs. CTG+ST analysis in original article recalculated to risk ratio (RR) (95% confidence interval) for CTG+ST analysis vs. CTG with MedCalc computer statistical software (MedCalc® Software, Mariakerke, Belgium).

[b]Calculated from available data with MedCalc computer statistical software.

[c]Calculated for modified ITT analysis.

Table 5. Neonatal outcome. Statistics represent cardiotocography plus ECG ST interval analysis (CTG+ST) group vs. CTG-only group.

Trial	Plymouth RCT[a]	Swedish RCT	Finnish RCT	French RCT	Dutch RCT
Percent missing blood gas data	Unclear, data reported only from full blood gas panel cases; selective analysis in first 400 cases (24)	376/5049 = 7.4% missing Imputed data reported (revised data)	36/1472 = 2.4% missing	Data cannot be extracted from article	20% found during ongoing trial, imputed data reported
Metabolic acidosis ($pH < 7.05 + BD_{ecf} > 12.0$ mmol/L)	5/1219 vs. 13/1215 RR 0.38 (0.14–1.07)[b]	15/2159 vs. 31/2079 RR 0.47 (0.25–0.86) (original data, modified ITT) 18/2565 vs. 35/2484 RR 0.50 (0.28–0.88) (revised, imputed data, standardized ITT)	Original data on BD_{ecf} not available, but according to Welin et al. (7) the figures are: 6/714 vs. 4/722 RR 1.52 (0.43–5.35)[c]	8/399 vs. 5/400 RR 1.60 (0.53–4.86)[c]	20/2827 vs. 30/2840 RR 0.70 (0.38–1.28) (original data) 19/2827 vs. 27/2840 RR 0.70 (0.38–1.28) (revised data)
pH < 7.15	110/1219 vs. 101/1215 RR 1.09 (0.84–1.41)[b]	–	–	–	–
pH < 7.10	–	–	41/714 vs. 34/722 RR 1.22 (0.78–1.90)	–	–
pH < 7.05	23/1219 vs. 25/1215 RR 0.92 (0.52–1.61)[b]	–	20/714 vs. 8/722 RR 2.53 (1.12–5.70)	12/399 vs. 11/400 RR 1.09 (0.49–2.45)[c]	47/2827 vs. 70/2840 RR 0.67 (0.46–0.97) (revised data)
pH < 7.00	–	–	–	–	18/2827 vs. 32/2840 RR 0.56 (0.31–1.01) (revised data)
Apgar score <4 at 1 min	–	36/2519 vs. 47/2447 RR 0.74 (0.48–1.14) 23/2228 vs. 38/2164 RR 0.59 (0.35–0.98)[b]	–	–	49/2827 vs. 40/2840 RR 1.25 (0.82–1.90)
Apgar score <7 at 5 min	20/1219 vs. 32/1215 RR 0.62 (0.36–1.08)[b]	26/2519 vs. 28/2447 RR 0.90 (0.53–1.53)	9/714 vs. 8/722 RR 1.14 (0.44–2.93) Incomplete data	6/399 vs. 6/400 RR 1.00 (0.33–3.08)[c]	42/2827 vs. 34/2840 RR 1.24 (0.79–1.95)
NICU admission	24/1219 vs. 31/1215 RR 0.77 (0.46–1.31)[b]	169/2519 vs. 181/2447 RR 0.91 (0.71–1.11)	26/714 vs. 26/722 RR 1.01 (0.59–1.72) Incomplete data	5/399 vs. 6/400 RR 0.84 (0.26–2.72)[c]	40/2827 vs. 45/2840 RR 0.89 (0.58–1.35)
Neonatal encephalopathy, Sarnat & Sarnat ≥ stage 1	–	3/2519 vs. 8/2447 RR 0.36 (0.10–1.37) 0/2228 vs. 6/2164 Fisher's exact test $p = 0.01$[b]	0/714 vs. 1/722 Fisher's exact test $p = 1.0$[c] Encephalopathy not defined, incomplete data	–	–
Neonatal encephalopathy, Sarnat & Sarnat ≥ stage 2	–	0/2519 vs. 3/2447 Fischer exact test $p = 0.1$[c]	Encephalopathy not defined	–	3/2827 vs. 1/2840 RR 3.01 (0.31–28.96)[c]
Neonatal seizures	–	Reported as Sarnat & Sarnat stage 2–3	0/714 vs. 2/722 Fisher's exact test $p = 0.50$[c] Incomplete data	1/399 vs. 1/400 RR 1.00 (0.06–16.00)[c]	Reported as Sarnat & Sarnat stage 2–3

Table 5. Continued

Trial	Plymouth RCT[a]	Swedish RCT	Finnish RCT	French RCT	Dutch RCT
Perinatal death	Data not reported, but 2/1219 vs. 0/1215 according to Westgate's thesis (24): Fisher's exact test $p = 0.5$[c]	3/2519 vs. 2/2447 RR 1.46 (0.24–8.71)[c]	0/714 vs. 0/722	0/399 vs. 1/400 Fisher's exact test $p = 1.0$[c]	3/2827 vs. 2/2840 RR 1.51 (0.25–9.01)[c]
Perinatal death corrected for lethal malformations	Data not reported	2/2519 vs. 1/2447 RR 1.94 (0.18–21.41)[c]	–	0/399 vs. 1/400 Fisher's exact test $p = 1.0$[c]	2/2827 vs. 0/2840 Fisher's exact test $p = 0.25$[c]

BD, base deficit; ITT, intention-to-treat; NICU, neonatal intensive care unit; RCT, randomized controlled trial; RR, relative risk.
[a]Odds ratio CTG vs. CTG+ST recalculated to risk ratio (95% confidence interval) CTG+ST vs. CTG with MedCalc computer statistical software.
[b]Calculated for modified ITT.
[c]MedCalc statistic.

figures to RR of CTG+ST vs. CTG alone (Table 5). Following RR calculation there was a 62% reduction in metabolic acidosis in the CTG+ST group, but the study was underpowered to demonstrate a significant difference in this outcome. Nonsignificant differences were also found for pH < 7.15, pH < 7.05, Apgar score <7 at 5 min, and admissions to the neonatal intensive care unit (NICU).

In a re-examination of the original data from the Swedish database (17), neonates with single vessel cord blood acid–base values, those with missing cord blood data but with neonatal blood gases and/or lactate indicating an affected acid–base status during the first hour of life, and imputed data in the group with missing data were included in the ITT analysis of metabolic acidosis. The revised Swedish RCT is the only study that report standardized ITT analyses as it included all randomized cases, irrespective of eligibility and availability of cord blood gas data. The significant difference in rates of metabolic acidosis between the CTG+ST and CTG-only groups remained (the original figures of 0.69% vs. 1.49% were recalculated to 0.66% vs. 1.33%). The original RCT (13) showed an RR for metabolic acidosis of 0.47 with 95% CI 0.25–0.86 ($p = 0.015$). After correction for misclassified cases, the RR for metabolic acidosis was 0.48 with 95% CI 0.24–0.96 ($p = 0.038$) while the standardized ITT yielded an RR for metabolic acidosis of 0.50 with 95% CI 0.28–0.88 ($p = 0.019$) (17).

As mentioned, the Finnish RCT used a different equation to calculate BD. After communication with the principal author, Welin et al. (7) reported the incidence of metabolic acidosis when the original BD_{blood} was recalculated to BD_{ecf}, showing a reduction from 12 to 6 among the 714 cases in the CTG+ST group and from 5 to 4 among the 722 cases in the CTG-alone group. Using the same BD_{ecf} algorithm as in the other RCTs, the total number of cases with metabolic acidosis was reduced from 17 to 10. The RR (95% CI) for metabolic acidosis in the CTG+ST arm of 2.43 (0.86–6.85) was reduced to 1.52 (0.43–5.35) (Table 5). This does not alter the original conclusion that there was no significant difference between the groups, but it reduces the differences between the results of the Finnish RCT and those from Plymouth, Sweden and the Netherlands. This recalculation also reduces the reported heterogeneity in the incidence of metabolic acidosis.

In the French RCT, Vayssière et al. (15) reported a total of seven cases of $BD_{ecf} > 12.0$ mmol/L with validated blood samples in the CTG+ST group (7/399, 1.75%) but presented eight cases of metabolic acidosis (8/399, 2.0%), defined as pH < 7.05 and $BD_{ecf} > 12.0$ mmol/L. This divergence remains unexplained. Moreover, the higher metabolic acidosis rate in the CTG+ST group compared with the CTG-alone group (2.0% vs. 1.25%) is

2014 The Authors. *Acta Obstetricia et Gynecologica Scandinavica* published by John Wiley & Sons on behalf of Nordic Federation of Societies of Obstetrics and Gynecology. **93** (2014) 556–569

contradicted by the trend towards a lower rate of BD_{ecf} > 12.0 mmol/L in this arm (1.75% vs. 3.0%, Fishers exact test $p = 0.098$). Among cases with validated cord blood samples there were 19 cases with BD_{ecf} > 12.0 mmol/L, but crude data retrieved from the individual participant data MA by Schuit et al. (5) showed a total of 123 cases with a BD_{ecf} > 12.0 mmol/L in the French RCT. These figures indicate that the cord blood gas samples must have been of inferior quality, because only 15.4% (19/123) of the samples with BD_{ecf} > 12.0 fulfilled the validation criterion that the cord artery-to-vein P_{CO_2} difference should be >0.5 kPa. Furthermore, abnormally high BD_{ecf} values were reported in 15.4% (123/799) of cases, compared to 2–4% in the other RCTs. The striking differences in the composition of the French population sample might have limited the ability of the ST analysis to prevent metabolic acidosis, as also pointed out by Schuit et al. (5). After quality control, revised data from the Dutch RCT concerning metabolic acidosis were published in 2011 (18,19). The number of cases with metabolic acidosis was corrected from 50 to 46, but the original RR of 0.70 (95% CI 0.38–1.28) remained unchanged in the CTG+ST arm (16,19).

Admissions to the NICU were reported in all five RCTs (Table 5). Reductions in NICU admissions by 9–23% in the CTG+ST groups were reported in four RCTs and an increase by 1% in one RCT; none of these differences were statistically significant. Neonatal encephalopathy and/or seizures were reported in all trials except in the Plymouth RCT; no significant differences were found except for encephalopathy stage 1–3, which was significantly less common in the CTG+ST group in the Swedish RCT. However, neonatal encephalopathy was not uniformly defined; only the Swedish and Dutch RCTs defined this outcome according to Sarnat & Sarnat criteria stage 1–3 (32). The Swedish RCT reported on stage ≥1 and stage ≥2 separately, while the Dutch RCT reported only on stage ≥2. Westgate provided more details on neonatal outcome in the Plymouth RCT in her thesis (24), but they do not allow for the retrospective classification of neonatal encephalopathy. The Finnish RCT reported more cases of neonatal seizures than cases with a diagnosis of encephalopathy, which could be in conflict with the Sarnat & Sarnat definition where seizure is defined as stage 2 encephalopathy. The French RCT did not report encephalopathy. Therefore, the impact of CTG+ST analysis on neonatal encephalopathy cannot be adequately determined for all cases included in the trials.

Perinatal mortality was reported in all RCTs except for the Plymouth study. Data retrieved from the thesis by Westgate (24) reveal two perinatal deaths in the CTG+ST group and none in the CTG-alone group. No RCT showed a significant difference in perinatal mortality between the study and control groups (Table 5), but all studies were underpowered to evaluate this outcome.

Summary of major strengths and weaknesses

Plymouth trial

This initial RCT of the CTG+ST methodology introduced the essential criteria for intervention used in the subsequent trials. The trial used an older ST analysis methodology than the other trials, and supports the hypothesis that ST analysis reduces metabolic acidosis and operative delivery.

Strengths.

- Single-center RCT, suggesting lower risk of inconsistent management
- Power calculation related to metabolic acidosis and ODFD
- Well-defined inclusion criteria, strict FBS guidelines related to the CTG pattern
- Interim analysis
- Short study period, high recruitment pace, large trial

Weaknesses.

- Underpowered to evaluate metabolic acidosis, recruitment goal not achieved
- Recruitments started from 34 weeks of gestation
- Not standardized ITT analysis, missing data and exclusions not clear
- Neonatal encephalopathy, seizures, deaths not reported

Swedish trial

The original trial was criticized for the exclusion from the ITT analysis of randomized cases that did not fulfill the inclusion criteria. This was addressed in a revised and standardized ITT analysis, i.e. inclusion of all randomized cases irrespective of eligibility.

Strengths.

- Power calculation related to metabolic acidosis
- STAN® S21 monitors in both trial arms, only internal monitoring
- Short trial period, high recruitment pace, large trial
- Interim analysis
- Revised article published with single vessel, neonatal and imputed cord blood gas data included, standardized ITT analysis addressing metabolic acidosis
- The only trial exposed to external review of crude data

Weaknesses.

- No clear guidelines for FBS use
- Errors in number of metabolic acidosis in original article

Finnish trial

This is a small trial, aimed to show differences in the incidence of severe fetal acidemia instead of metabolic acidosis. When metabolic acidosis was calculated, a different BD algorithm was used.

Strengths.

- Single-center RCT, suggesting lower risk of inconsistent management
- Short trial period, high recruitment pace

Weaknesses.

- Primarily not aimed to show metabolic acidosis difference
- Recruitment goal not achieved for primary outcome variable "cord artery pH < 7.10"
- Inclusion criterion, "amniotomy," provides no information on fetal risk
- No clear guidelines for FBS use
- Internal or external FHR monitoring in CTG group
- Not standardized ITT, missing data and exclusions not clear
- BD and metabolic acidosis calculated using a different algorithm from the other RCTs

French trial

The eligibility criteria of this trial allowed inclusion of cases with a CTG pattern that indicated fetal hypoxia before the CTG+ST monitoring was started, violating the ST analysis clinical guidelines and different from the other trials. It may be questioned if this trial should be pooled together with the other trials in comparisons.

Strength.

- Confirmed that ST analysis had no benefit in addition to conventional CTG in labors with pre-existing evidence of fetal compromise

Weaknesses.

- Underpowered for evaluation of metabolic acidosis
- Inclusion criterion "abnormal CTG" was a violation of clinical guidelines for use of CTG+ST monitoring
- Long trial period, low recruitment pace, small trial
- Not standardized ITT, missing data and exclusions not clear
- No clear guidelines for FBS use
- Internal or external monitoring in CTG group not specified
- Poor quality of umbilical cord blood samples, unreliable blood gas data

Dutch trial

The main weakness of the Dutch RCT is its low recruitment pace. This may have influenced the learning curve and the results of the trial, but the impact cannot be determined.

Strengths.

- Power calculation related to metabolic acidosis
- Safety committee watch of serious adverse events
- Well-defined inclusion criteria
- Strict guidelines for FBS use related to CTG+ST pattern (but not to CTG pattern alone)
- Internal electronic monitoring in CTG-alone group
- Largest RCT
- Not completely standardized ITT but only 14 cases (0.25%) excluded from ITT
- Revised data article published with corrected and imputed cord blood gas data

Weaknesses.

- Underpowered for metabolic acidosis, recruitment goal not achieved
- Long trial period, low recruitment pace, many centers involved
- Errors in number of metabolic acidosis in original article

Conclusions

The perfect RCT to evaluate the CTG+ST methodology remains to be performed, though "perfect" is probably an unachievable goal. While the larger Plymouth, Swedish and Dutch RCTs point towards reduced metabolic acidosis rates and reduced operative interventions in the CTG+ST arm, the smaller Finnish and French RCTs point towards higher metabolic acidosis rates and minor differences in operative delivery rates. Because the Finnish RCT calculated BD in blood rather than in extracellular fluid, this led to a falsely elevated rate of metabolic acidosis in comparison with the other RCTs. An adjusted analysis points to a less divergent result. The French RCT uses an inclusion criterion - abnormal CTG - which is in conflict

2014 The Authors. *Acta Obstetricia et Gynecologica Scandinavica* published by John Wiley & Sons on behalf of Nordic Federation of Societies of Obstetrics and Gynecology. **93** (2014) 556–569

with the premise for the use of CTG+ST, as ST events may not appear de novo when fetal hypoxia is already present. Moreover, the quality of cord blood samples used in the study is low.

All RCTs, except for the Swedish one, were underpowered for the primary outcome of metabolic acidosis; only the Plymouth RCT was adequately powered for the primary outcome of ODFD. The inclusion criteria varied considerably among the trials, rendering any comparison of the primary outcomes problematic with such widely varying a priori risk factors. This finding alone could explain some of the differences.

Finally, the conduct of the RCTs, including recruitment pace and total enrollment, interim analysis, and accounting for ITT, reflect the difficulties inherent in the undertaking of large clinical trials. These variations contribute to the drawbacks of any subsequent MAs, as they will enhance heterogeneity effects. This is addressed further in a separate appraisal of the previously published MAs (33).

Funding

No specific funding was received for this review.

References

1. Becker JH, Bax L, Amer-Wåhlin I, Ojala K, Vayssière C, Westerhuis MEMH, et al. ST analysis of the fetal electrocardiogram in intrapartum fetal monitoring. A meta-analysis. Obstet Gynecol. 2012;119:145–54.
2. Neilson JP. Fetal electrocardiogram (ECG) for fetalmonitoring during labour. Cochrane Database Syst Rev. 2012;4:CD000116. DOI: 10.1002/14651858.CD000116.pub3.
3. Potti S, Berghella V. ST waveform analysis versus cardiotocography alone for intrapartum monitoring: a meta-analysis of randomized trials. Am J Perinatol. 2012;29:657–64.
4. Salmelin A, Wiklund I, Bottinga R, Brorsson B, Ekman-Ordeberg G, Eneroth Grimfors E, et al. Fetal monitoring with computerized ST analysis during labor: a systematic review and meta-analysis. Acta Obstet Gynecol Scand. 2013;92:28–39.
5. Schuit E, Amer-Wåhlin I, Ojala K, Vayssière C, Westerhuis MEMH, Maršál K, et al. Effectiveness of electronic fetal monitoring with additional STanalysis in vertex singleton pregnancies beyond 36 weeks of gestation: an individual participant data meta-analysis. Am J Obstet Gynecol. 2013;208:187.e1–13.
6. Massoud M, Giannesi A, Amabile N, Manevy M, Geron G, Gaucherand P. Fetal electrocardiotocography in labor and neonatal outcome: an observational study in 1889 patients in the French center of Edouard Herriot, Lyon. J Matern Fetal Neonatal Med. 2007;20:819–24.
7. Welin A-K, Norén H, Odeback A, Andersson M, Andersson G, Rosén KG. STAN, a clinical audit: the outcome of 2 years of regular use in the city of Varberg, Sweden. Acta Obstet Gynecol Scand. 2007;86:827–32.
8. Kale A, Chong Y-S, Biswas A. Effect of availability of fetal ECG monitoring on operative deliveries. Acta Obstet Gynecol Scand. 2008;87:1189–93.
9. Norén H, Carlsson A. Reduced prevalence of metabolic acidosis at birth: an analysis of established STAN usage in the total population of deliveries in a Swedish district hospital. Am J Obstet Gynecol. 2010;202:e1–7.
10. Doret M, Massoud M, Constans A, Gaucherand P. Use of peripartum ST analysis of fetal electrocardiogram without blood sampling: a large prospective cohort study. Eur J Obstet Gynecol Reprod Biol. 2011;156:35–40.
11. Kessler J, Moster D, Albrechtsen S. Intrapartum monitoring of high-risk deliveries with ST analysis of the fetal electrocardiogram: an observational study of 6010 deliveries. Acta Obstet Gynecol Scand. 2013;92:75–84.
12. Westgate J, Harris M, Curnow JSH, Greene KR. Plymouth randomized trial of cardiotocogram only versus ST waveform plus cardiotocogram for intrapartum monitoring in 2400 cases. Am J Obstet Gynecol. 1993;169:1151–60.
13. Amer-Wåhlin I, Hellsten C, Norén H, Hagberg H, Herbst A, Kjellmer I, et al. Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial. Lancet. 2001;358:534–8.
14. Ojala K, Vääräsmäki M, Mäkikallio K, Valkama M, Tekaya A. A comparison of intrapartum automated fetal electrocardiography and conventional cardiotocography – a randomised controlled study. BJOG. 2006;113:419–23.
15. Vayssière C, David E, Meyer N, Haberstich R, Sebahoun V, Roth E, et al. A French randomized controlled trial of ST-segment analysis in a population with abnormal cardiotocograms during labor. Am J Obstet Gynecol. 2007;197:299.e1–6.
16. Westerhuis MEMH, Visser GHA, Moons KGM, van Beek E, Benders MJ, Bijvoet SM, et al. Cardiotocography plus ST analysis of fetal electrocardiogram compared with cardiotocography only for intrapartum monitoring. A randomized controlled trial. Obstet Gynecol. 2010;115:1173–80.
17. Amer-Wåhlin I, Kjellmer I, Marsál K, Olofsson P, Rosén KG. Swedish randomized controlled trial of cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram revisited: analysis of data according to standard versus modified intention-to-treat principle. Acta Obstet Gynecol Scand. 2011;90:990–6.
18. Westerhuis MEMH, Visser GHA, Moons KGM, Zuithoff NPA, Mol BWJ, Kwee A. Letter to the editor. Cardiotocography plus ST analysis of fetal

electrocardiogram compared with cardiotocography only for intrapartum monitoring: a randomized controlled trial. Obstet Gynecol. 2011;117:406–7.
19. Westerhuis MEMH, Visser GHA, Moons KGM, Zuithoff NPA, Mol BWJ, Kwee A. Corrections. Cardiotocography plus ST analysis of fetal electrocardiogram compared with cardiotocography only for intrapartum monitoring: a randomized controlled trial. Obstet Gynecol. 2011;117:412.
20. Sundström A-K, Rosén D, Rosén KG. Fetal surveillance. Gothenburg: Neoventa Medical AB, 2000.
21. Bakker PC, Colenbrander GJ, Verstraeten AA, Van Geijn HP. The quality of intrapartum fetal heart rate monitoring. Eur J Obstet Gynecol Reprod Biol. 2004;116:22–7.
22. Leipälä J, Ojala K, Stefanovic V, Mäkelä M. Cardiotocography and ST analysis for intrapartum fetal monitoring. Acta Obstet Gynecol Scand. 2012;91:518.
23. Amer-Wåhlin I, Kjellmer I, Maršál K, Olofsson P, Rosén KG. Cardiotocography and ST analysis for intrapartum fetal monitoring. Acta Obstet Gynecol Scand. 2012;91:519.
24. Westgate JA. An evaluation of electronic fetal monitoring with clinical validation of ST waveform analysis during labour. Thesis. University of Plymouth, 1993.
25. Wiberg N, Källén K, Olofsson P. Base deficit estimation in umbilical cord blood is influenced by gestational age, choice of fetal fluid compartment, and algorithm for calculation. Am J Obstet Gynecol. 2006;195:1651–6.
26. Mokarami P, Wiberg N, Olofsson P. An overlooked aspect on metabolic acidosis at birth: blood gas analyzers calculate base deficit differently. Acta Obstet Gynecol Scand. 2012;91:574–9.
27. Siggaard-Andersen O. An acid base chart for arterial blood with normal and pathophysiological reference areas. Scand J Clin Lab Invest. 1971;27:239–45.
28. Siggaard-Andersen O. The acid–base status of the blood, 4th edn. Copenhagen: Munksgaard, 1976.
29. Siggaard-Andersen O. Therapeutic aspects of acid–base disorders. In: Evans FT, Gray TC (eds). Modern trends in anaesthesia. Part 3. London: Butterworths, 1966. pp. 99–131.
30. Rooth G, Jacobson L. The value and validity of base excess ecf in perinatal acid–base studies. Scand J Clin Lab Invest. 1971;28:283–6.
31. Wimberley PD. Fetal hemoglobin, 2-3-diphosphoglycerate and oxygen transport in the newborn premature infant. Scand J Clin Lab Invest. 1982;42(Suppl. 160):1–149.
32. Sarnat HB, Sarnat MS. Neonatal encephalopathy following fetal distress. A clinical and electroencephalographic study. Arch Neurol. 1976;33:696–705.
33. Olofsson P, Ayres-de-Campos D, Kessler J, Tendal B, Yli BM, Devoe L. A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part II: the meta-analyses. Acta Obstet Gynecol Scand. 2014;93:571–86.

Reviewer comments:

The subject under study, fetal surveillance in labor, is definitely of utmost importance. The authors argue in a "summary of major strengths and weaknesses" that a single-center randomized controlled trial has to be seen as a strength. I do not agree. Neither do I see that the use of STAN monitors in both trial arms has to be a strength. Why is "No clear guidelines for fetal scalp blood sampling use" to be seen as a weakness? The major weakness for all studies seems to be nonadherence to intention-to-treat analyses.

Ove Axelsson
Department of Obstetrics and Gynecology, Uppsala University, Uppsala, Sweden

Reviewer comments:

The authors critically appraised the five randomized controlled trials comparing cardiotocography plus ECG ST interval analysis (CTG+ST) vs. CTG in fetal surveillance during labor published so far. This is an important message since differences in outcomes between the studies are likely to be explained by significant differences in study designs indicating heterogeneity. It is striking that only five such randomized controlled trials have been reported since 1993.

In conclusion, the authors argue that a proper randomized controlled trial needs to be done. They should take this one step further, i.e. how to design a proper randomized controlled trial. Inclusion criteria? Power calculation? End points? For instance, "neonatal intensive care unit" has a totally different meaning in different institutions, and crite-

