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 II: the meta-analyses

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, [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

Conflict 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 consulting Global Medical Adviser. Jörg Kessler has received a lecture fee once from Neoventa Medical AB. Branka M. Yli has taught STAN 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 stated explicitly that they have no conflicts of interest in connection with this article.

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 II: the meta-analyses. Acta Obstet Gynecol Scand 2014; 93: 571–586.

Received: 16 February 2014
Accepted: 30 April 2014

DOI: 10.1111/aogs.12412

Abstract

We appraised the methodology, execution and quality of the five published meta-analyses that are based on the five randomized controlled trials which compared cardiotocography (CTG)+ST analysis to cardiotocography. The meta-analyses contained errors, either created *de novo* in handling of original data or from a failure to recognize essential differences among the randomized controlled trials, particularly in their inclusion criteria and outcome parameters. No meta-analysis contained complete and relevant data from all five randomized controlled trials. We believe that one randomized controlled trial excluded in two of the meta-analyses should have been included, whereas one randomized controlled trial that was included in all meta-analyses, should have been excluded. After correction of the uncovered errors and exclusion of the randomized controlled trial that we deemed inappropriate, our new meta-analysis showed that CTG+ST monitoring significantly reduces the fetal scalp blood sampling usage (risk ratio 0.64; 95% confidence interval 0.47–0.88), total operative delivery rate (0.93; 0.88–0.99) and metabolic acidosis rate (0.61; 0.41–0.91).

Abbreviations: BD, base deficit; BD_{blood}, base deficit in blood; BD_{ecf}, base deficit in extracellular fluid; CI, confidence interval; CS, cesarean section; CTG+ST, cardiotocography combined with fetal ECG ST interval analysis; CTG, cardiotocography; ECG, electro-cardiotocography; FBS, fetal scalp blood sampling; FD, fetal distress; IPD, individual participant (patient) data; MA, meta-analysis; ODFD, operative delivery for fetal distress; RCT, randomized controlled trial; RR, risk ratio.

Introduction

From 2012 to 2013, five meta-analyses (MAs) on the value of cardiotocography (CTG) combined with fetal ECG ST interval analysis (CTG+ST) have been published: an updated Cochrane Review (1), one by a European consortium involved in four of the five randomized controlled trials (RCTs) performed on CTG+ST vs. CTG alone (2) (denoted "European MA" in text and tables), one by North American authors (3) ("American MA"), one by a group in Stockholm, Sweden (4) ("Stockholm MA"), and an individual participant data (IPD) MA by the European consortium (5) ("IPD MA"). This monograph focuses on the methodologies employed in the MAs, the clinical outcomes considered, and the execution and quality of each individual MA. New MAs were performed in those events where we found critical differences between the RCTs [see the accompanying Part I review (6)], and when improper handling of RCT data or errors were found in the five MAs.

Five meta-analyses

Five RCTs on CTG+ST vs. CTG alone were considered for inclusion in the MAs: the "Plymouth RCT" published by Westgate et al. in 1993 (7), the "Swedish RCT" by Amer-Wåhlin et al. in 2001 (8), the "Finnish RCT" by Ojala et al. in 2006 (9), the "French RCT" by Vayssière et al. in 2007 (10), and the "Dutch RCT" by Westerhuis et al. in 2010 (11). After the original articles, revised data from the Swedish and Dutch RCTs were published in 2011 (12–14). Metabolic acidosis data from the Finnish RCT have been revised (see below), but not data from the Plymouth and French RCTs.

Statistical analyses

For supplementary statistical calculations, we used the MEDCALC® version 5.00.017 computer software (MedCalc Software, Mariakerke, Belgium). Two-sided statistics were used with a p-value <0.05 considered significant. For performing new MAs, we used the COCHRANE REVIEW MANAGER version 5.2.7 computer software (The Cochrane Collaboration, http://ims.cochrane.org/revman/download). This program assesses heterogeneity with Tau^2, I^2 and chi-square (Cochrane Q) statistics, where heterogeneity is regarded as substantial if I^2 exceeds 30% or the chi-square test p-value is <0.10. An analysis showing low heterogeneity can be presented with fixed-effect MA and an analysis showing high heterogeneity with random-effect MA; since the result is practically the same with the two models at low heterogeneity, in the text and forest plots we present the results as random-effect MAs.

Types of meta-analysis

The Cochrane, European, American and Stockholm MAs used aggregated data (Table 1). The IPD MA analysed the original raw data from participants in four of the five RCTs. An IPD MA offers numerous statistical and clinical advantages over an aggregate data MA (15). For example, it increases the power to detect differential treatment effects across individuals in RCTs and allows adjustment for confounding factors in observational studies.

In the Cochrane, European and American MAs, the random-effect and fixed-effect MA models were used as appropriate, after testing for heterogeneity (Table 1). However, there is no consensus in the literature as to the ideal cut-off point for heterogeneity to be used for each model. For example, Reid (16) recommends the fixed-effect model at an I^2 of ≤25% and the random-effect model at ≥75%, but gives no certain recommendation for values in between. Devane (17) gives a somewhat more precise recommendation: at an I^2 of 0–40%, heterogeneity is not important; 30–60% represents moderate heterogeneity; 50–90% substantial heterogeneity; and 75–100% considerable heterogeneity. Several other interpretations can be found in the literature. The chi-square test has the lowest power to detect heterogeneity and a $p < 0.10$ indicates heterogeneity according to Devane (17). While the I^2 index quantifies the degree of heterogeneity in a MA, the chi-square only informs us about the presence or absence of heterogeneity (18). Devane (17) recommends that in the case of statistical heterogeneity, the reasons for this finding should be investigated and the statistical approach appropriately modified.

In the MAs included in the present review, the cut-offs for I^2 heterogeneity varied from 30% (Cochrane MA) to 40% (American MA) and 50% (European MA) (Table 1). In the American MA, when an I^2 was ≥85%, the authors chose to perform no MA, for example regarding fetal scalp blood sampling (FBS). In the other MAs the random-effect model was then used. The Tau^2 cut-off was set to >0 in the Cochrane and European MAs but was

Key Message

Published meta-analyses on studies comparing cardiotocography+ST analysis with cardiotocography only, contained errors in handling of original data, unwarranted inclusions/exclusions of trials, and variable definitions of outcomes. A revised meta-analysis showed reductions in fetal scalp blood sampling, total operative delivery rate, and metabolic acidosis rate in the CTG+ST arm.

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) 571–586

Table 1. Details of five meta-analyses (MAs) based on five randomized controlled trials (RCTs) on the value of cardiotocography (CTG) combined with fetal ECG ST interval analysis (CTG+ST) for fetal surveillance in labor.

Meta-analysis	Cochrane review Neilson (2012)	European MA Becker et al. (2012)	American MA Potti & Berghella (2012)	Stockholm MA Salmelin et al. (2013)	IPD MA Schuit et al. (2013)
Type of meta-analysis	Aggregate	Aggregate	Aggregate	Aggregate	Individual participant data
Data collection	5 RCTs, principal authors of Swedish and French RCTs contacted for missing data	5 RCTs, principal authors of Swedish, Finnish, French and Dutch RCTs among authors to European and IPD MAs	5 RCTs, only data used in original articles are used	4 RCTs, only data used in original articles are used	4 RCTs, IPD provided by principal investigators: all randomized cases from Swedish and French RCTs included, from Finnish RCT 11 exclusions, from Dutch RCT 14 exclusions
Measures of treatment effect	RR with 95% CI (fixed-/random-effect model as appropriate); fixed-effect when no heterogeneity	RR with 95% CI (fixed-/random-effect model as appropriate); fixed effect when no heterogeneity	RR with 95% CI (fixed-/random-effect models as appropriate); fixed-effect when no heterogeneity	RR with 95%CI; consistently random-effect, although tests for heterogeneity were performed	RR with an RR <1 indicating treatment benefit; random-effect log-binomial model; imputation of missing data
Assessment of heterogeneity (figures indicate substantial heterogeneity)	Tau^2 (>0), I^2 (>30%), chi-square for heterogeneity ($p < 0.10$)	Tau^2 (>0), I^2 (>50%)	I^2 (40–84%, if ≥85% no MA), chi-square for heterogeneity ($p < 0.10$)	Performed, but random-effects MA consistently used	I^2 (0% indicating no heterogeneity, 25% low, 50% moderate, 75% high)

IPD, individual participant data; RR, risk ratio.

not defined in the American MA; the chi-square *p*-value was <0.10 in the Cochrane and American MAs but not calculated in the European MA.

Choosing the right model for MA is particularly important for binary outcome variables because the fixed- and random-effect models give different results. When heterogeneity is present, a confidence interval (CI) around the random-effect pooled estimate is wider than the CI around a fixed-effect pooled estimate (19). Thus, larger series are required in the random-effect model to achieve the same statistical power as in the fixed-effect model (20). This is illustrated by the calculation of metabolic acidosis in the European MA, showing an I^2 of 33%: the random-effect model showed a non-significant decrease of metabolic acidosis in the CTG+ST group [risk ratio (RR) 0.72, 95% CI 0.43–1.19], but if the pre-defined cut-off for I^2 heterogeneity (50%) is used, the fixed-effect model will show a significant reduction (RR 0.68, 95% CI 0.48–0.97) (Tables 1 and 6).

In summary, fixed- and random-effect models pose different questions. The random-effect model addresses the question "what is the average intervention effect?" whereas the fixed-effects model addresses the question "what is the best estimate of the intervention effect?" (19). Since the random-effect model estimates the underlying distribution of effects and not a single effect, when the models do not coincide it may not reflect the actual effect in the particular population under study. When heterogeneity is present, the random-effect MA will award more weight to smaller trials than such studies would receive in a fixed-effect MA. Consequently, if the results of smaller trials are consistently different from those of larger ones, which is the case with the Finnish and French RCTs, the direction of the outcomes in the entire MA can be shifted. A random-effect MA as a rule gives a more conservative 95% CI.

Before presenting our evaluation of the individual MAs, it is important to recognize that random-effect analysis is not a solution for the difficulties inherent in translating the results of a MA to the realities of daily clinical practice. The Cochrane Handbook (19) states that the choice between a fixed-effect and a random-effect MA should never be made on the basis of a statistical test for heterogeneity. As will become evident, this recommendation was not uniformly applied to the MAs under consideration.

Inclusion and exclusion of RCTs and their relevant data in the meta-analyses

The Cochrane Review included all five RCTs in its MA and cited the revised versions of the Swedish and Dutch

RCTs (12,14), but it did not include the revised Swedish data in the final analysis (Table 2). The Cochrane MA chose to use base deficit (BD) in blood (BD_{blood}) and not BD in extracellular fluid (BD_{ecf}) for calculation of metabolic acidosis, but BD_{blood} metabolic acidosis was reported only in the Finnish and Dutch RCTs – the Plymouth, Swedish and French RCTs reported BD_{ecf} metabolic acidosis. Consequently, the Cochrane MA is a mixture of two different ways to calculate BD and its metabolic acidosis result is therefore not uniform, because the different BD calculation algorithms have a large impact on the incidence of metabolic acidosis [see below and the accompanying Part I review (6)].

The European MA authors also included all five RCTs in their analysis (Table 2). However, the Swedish RCT data presented are from the original so-called modified intent-to-treat analysis from 2001 (n = 4966) (8), not the revised data from the so-called standardized intention-to-treat from 2011 (n = 5049) (12). The European consortium authors aimed to calculate metabolic acidosis with BD_{ecf} data and converted the Finnish BD_{blood} data to BD_{ecf} data to be comparable with the other RCTs, but they included cases with missing blood gases in the denominators when calculating the metabolic acidosis rates (6/733 vs. 4/739 instead of 6/714 vs. 4/722). Thus, the European MA did not contain all relevant data from

Table 2. Details of RCTs included/excluded in the MAs, with special reference to calculation of neonatal metabolic acidosis.

Meta-analysis RCT	Cochrane MA[a] Neilson (2012)	European MA[b] Becker et al. (2012)	American MA[c] Potti & Berghella (2012)	Stockholm MA[d] Salmelin et al. (2013)	IPD MA[e] Schuit et al. (2013)
Plymouth RCT (Westgate et al., 1993)	Included with BD_{ecf} data for metabolic acidosis	Included	Included	Excluded because of non-computerized ST analysis method	Excluded because of non-computerized ST analysis method and no access to IPD
Swedish RCT original data (Amer-Wåhlin et al., 2001)	Included with BD_{ecf} data for metabolic acidosis	Included but incorrect data used in MA	Included	Included	IPD included
Swedish RCT revised data on metabolic acidosis (Amer-Wåhlin et al., 2011)	Article cited but revised data not used in MA	Article cited but revised data not used in MA	Not included, not cited	Included	IPD included
Finnish RCT original data (Ojala et al., 2006)	Included with BD_{blood} data for metabolic acidosis	Included	Included	Included	IPD included
Awareness of different calculation of metabolic acidosis in Finnish RCT?	No	Yes, but wrong denominators included in MA	No, included BD_{blood} in metabolic acidosis calculation	Yes, but included BD_{blood} in metabolic acidosis calculation	Yes, included Finnish BD_{ecf} data in metabolic acidosis calculation
French RCT (Vayssière et al., 2007)	Included with BD_{ecf} data for metabolic acidosis	Included	Included	Included	IPD included
Dutch RCT original data (Westerhuis et al., 2010)	Included	Included	Included	Included	IPD included
Dutch RCT revised data on metabolic acidosis BD_{ecf}, pH <7.05, pH <7.00 (Westerhuis et al., 2011)	Included with revised BD_{blood} data for metabolic acidosis	Included, correct data used for metabolic acidosis	Included, correct data used for metabolic acidosis but revised article not cited	Included, correct data used for metabolic acidosis	IPD included, correct data used for metabolic acidosis
Number of cases included	15 338	15 352 (≤15 338 included in analyses)	15 303	12 904	12 987

BD_{blood}, base deficit in blood; BD_{ecf}, base deficit in extracellular fluid.
[a]The Cochrane review aimed to analyse metabolic acidosis with BD_{blood}.
[b]The European MA aimed to analyse metabolic acidosis with BD_{ecf}.
[c]The American MA did not define the fetal compartment for calculation of BD.
[d]The Stockholm MA did not decide to calculate BD_{ecf} and BD_{blood} metabolic acidosis separately.
[e]The IPD MA aimed to analyse metabolic acidosis with both BD_{ecf} and BD_{blood} without mixing of data.

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) 571–586

the five RCTs. In a second sequence of the European MA, "sensitivity analyses" excluded the Plymouth RCT, as it used visual analysis of absolute T/QRS ratios and because biphasic ST interval changes were not yet part of the method (but this is not correct, see below).

The American MA included all five RCTs but with the original metabolic acidosis data instead of the revised data from the Swedish trial group (Table 2). Moreover, the use of BD_{blood} instead of BD_{ecf} to calculate metabolic acidosis in the Finnish RCT was not taken into account (see below). The Stockholm MA also disregarded the fact that the Finnish RCT reported BD_{blood} data. Thus, the American and Stockholm MAs on metabolic acidosis were mixtures of BD_{ecf} and BD_{blood} data (Table 2). The IPD MA aimed to analyse metabolic acidosis with both BD_{ecf} and BD_{blood} calculations of metabolic acidosis and the concepts were not mixed together.

The Stockholm and IPD MAs excluded the Plymouth RCT because of the non-computerized ST analysis methodology and, in the case of the IPD MA, because biphasic ST interval changes were not included in the ST analysis guidelines (Table 2). The latter claim is not entirely correct because negative T wave and ST interval depression with positive T waves were included in the Plymouth RCT management protocol [see Westgate et al., 1993 (7), Table II)]. In a response to a Letter to the Editor of the *American Journal of Obstetrics and Gynecology* by Rosén (21), the principal IPD MA author admitted that biphasic ST changes were incorporated in the Plymouth RCT management protocol, and that another reason for not including the Plymouth data was that they had no access to the IPD (22). The Plymouth RCT authors were contacted but could not provide the required data. This has affected the results of the IPD MA (and the Stockholm MA), since the Plymouth RCT contributed considerable weight, 16.2–17.0%, to the analyses of metabolic acidosis in those MAs that included it (1–3).

The IPD MA authors make an assertion that all RCTs had the same inclusion criteria, making them only "slightly different". However, the French RCT only included women with abnormal CTG in labor with or without meconium-stained amniotic fluid, but excluded normal CTG cases (10), criteria that in many cases are violations of the ST analysis clinical guidelines and recommendations (23,24). This fact alone should have invalidated the French RCT from inclusion not only in the IPD MA but also in the other MAs [for details, see the accompanying Part I review (6)].

Handling of missing data

Several of the variables evaluated in the MAs were not reported in the original RCTs, and we could not perform *post hoc* analyses of these variables. The Cochrane Review author contacted the authors of the original reports to provide further data. Representatives from all RCTs except the Plymouth RCT were co-authors of the European MA and IPD MA and could have provided missing data; the American and Stockholm MAs were performed without contributions from authors of the included RCTs.

Fetal scalp blood sampling: discrepancies in the meta-analyses

In all five RCTs, FBS was an adjunct diagnostic tool in both the CTG+ST group and CTG alone group. However, it is unclear why the Swedish RCT data were not available for the IPD MA (Table 3). In the Cochrane Review the rates of FBS in the Dutch RCT were tabulated as an outcome variable, but these data were not included in the MA. The Cochrane MA reported an RR of 0.61 (95% CI 0.41–0.91), but if the Dutch RCT data (302/2827 vs. 578/2840) are included, this results in an RR of 0.59 (95% CI 0.55–0.65) (788/7697 vs.1316/7641). Thus, inclusion of the large Dutch RCT series results in a narrower and more robust CI but no important change in RR. The American MA did not analyse FBS because of their calculation of high heterogeneity among studies.

A new meta-analysis of fetal scalp blood sampling

All four MAs that evaluated FBS usage showed significant reductions in the CTG+ST group, ranging from 39 to 51%, but in the Cochrane Review and the IPD MA the data were not complete (Table 3). As discussed in the accompanying Part I review (6) and elsewhere in the present review, the French RCT should not be pooled in an MA with the other RCTs because of methodological discrepancies. Our MA including the four other RCTs showed a significant reduction in FBS usage by 36% in the CTG+ST group (RR 0.64, 95% CI 0.47–0.88) (Figure 1, Table 4).

Operative delivery: discrepancies in the meta-analyses

It is not possible to determine the total cesarean and operative vaginal delivery rates in the Plymouth and French RCTs. Imputed data for the Cochrane Review were provided by the original RCT authors. For reasons that are unclear, data on total operative vaginal delivery rate from the Dutch RCT were not included in the Cochrane Review (Table 3). The Cochrane MA showed an RR of 0.89 (95% CI 0.81–0.98), which after inclusion of

Table 3. Interventions in labor. Calculations are CTG+ST analysis vs. CTG alone, presented as RR (95% confidence interval).

Meta-analysis	Cochrane review Neilson (2012)	European MA Becker et al. (2012)	American MA Potti & Berghella (2012)	Stockholm MA Salmelin et al. (2013)	IPD MA Schuit et al. (2013)
Fetal scalp blood sampling	486/4870 vs. 738/4801 (9.98 vs. 15.37%) Random-effect (Tau^2 0.15, I^2 92%, chi-square $p < 0.00001$): RR 0.61 (0.41–0.91)	Included 5 RCTs, detailed data not provided Random-effect (statistics not provided): RR 0.59 (0.44-0.79) Fixed-effect model: RR 0.60 (0.55–0.65)	MA not reported	694/6478 vs. 1202/6426 (10.71 vs. 18.70%) Random-effect (Tau^2 0.09, I^2 92%, chi-square $p < 0.00001$): RR 0.55 (0.40–0.76)	460/(6524–2565) vs. 941/(6463–2484) (11.61 vs. 23.65%) RR 0.49 (0.44–0.55)
Comments on fetal scalp blood sampling	Dutch RCT data (302/2827 vs. 578/2840) for unclear reasons not included in MA	Narrowing of the 95% CI with fixed-effect model	MA not reported because of heterogeneity >85% (I^2 91%); detailed RCT data not presented	No errors found	Detailed IPD cannot be checked; for unknown reason Swedish RCT data were not available
Cesarean section, total	876/7697 vs. 878/7641 (11.38 vs. 11.49%) Fixed-effect (I^2 0%, chi-square $p = 0.87$): RR 0.99 (0.91–1.08)	Included 3 RCTs, detailed data not provided Random-effect (statistics not provided): RR 1.03 (0.87–1.2)	876/7697 vs. 878/7641 (11.38 vs. 11.49%) Unknown MA type (statistics not provided): RR 0.99 (0.91–1.08)	MA not performed	RR 0.99 (0.91–1.09)
Comments on total cesarean section analysis	Total cesarean section rates not reported in Plymouth and French RCTs – data in MA provided by RCT authors	Unclear which 3 RCTs were included in MA, but data were not reported in Plymouth and French RCTs	Detailed RCT data not provided; total cesarean section rates were not reported in Plymouth and French RCTs; unclear how data were obtained	–	Detailed IPD cannot be checked; error in addition of cases: sum ($n = 1534$) doesn't fit with additions of CSFD ($n = 507$) and CSFP ($n = 763$)
Cesarean section for fetal distress	MA not performed	Included 5 RCTs, detailed data not provided Random-effect (statistics not provided): RR 0.90 (0.67–1.2)	MA not performed	MA not performed	RR 0.99 (0.83–1.17)
Comments on cesarean section for fetal distress	–	Calculated from original RCT articles: 262/7697 vs. 277/7641 (3.40 vs. 3.63%) RR 0.94 (0.80–1.11)	–	–	Detailed IPD cannot be checked
Operative vaginal delivery, total	660/4870 vs. 731/4801 (13.55 vs. 15.23%) Fixed-effect (I^2 0.0%, chi-square $p = 0.49$): RR 0.89 (0.81–0.98)	Included 3 RCTs, detailed data not provided Fixed-effect ("among-study variance zero"): RR 0.88 (0.80–0.97)	1044/7697 vs. 1162/7641 (13.56 vs. 15.19%) Fixed-effect (I^2 0%, chi-square $p = 0.66$): RR 0.89 (0.83–0.97)	MA not performed	RR 0.90 (0.83–0.99)

Table 3. Continued

Meta-analysis	Cochrane review Neilson (2012)	European MA Becker et al. (2012)	American MA Potti & Berghella (2012)	Stockholm MA Salmelin et al. (2013)	IPD MA Schuit et al. (2013)
Comments on operative vaginal delivery	For unclear reasons Dutch RCT data were not included in MA; Plymouth and French RCT data not in articles – data in MA provided by RCT authors	Random-effect model yielded same results; unclear which 3 RCTs were included in MA, but data not reported in Plymouth and French RCTs	Plymouth and French RCT data not in articles, unclear how these data were obtained	–	Detailed IPD cannot be checked; error in addition of cases: sum (n = 1732) doesn't fit with additions of OVDFD (n = 821) and OVDFP (n = 652)
Operative vaginal delivery for fetal distress	MA not performed	Included 5 RCTs, detailed data not provided Fixed-effect ("among-study variance zero"): RR 0.86 (0.76–0.97)	MA not performed	MA not performed	RR 0.91 (0.80–1.05)
Comments on operative vaginal delivery for fetal distress	–	Random-effect model: RR 0.83 (0.67–1.0)	–	–	Detailed IPD cannot be checked
Total operative delivery for fetal distress	MA not performed	MA not performed	MA not performed	639/6478 vs. 675/6426 (9.86 vs. 10.50%) Random-effect (Tau^2 0.01, I^2 52%, chi-square p = 0.10): RR 0.93 (0.80–1.08)	RR 0.94 (0.84–1.05)
Comments on total operative delivery for fetal distress	–	–	–	No errors found	Detailed IPD cannot be checked
Total operative delivery for failure to progress/other reasons	MA not performed	MA not performed	MA not performed	937/6478 vs. 982/6426 (14.46 vs. 15.28%) Random-effect (Tau^2 0.01, I^2 44%, chi-square p = 0.15): RR 0.98 (0.86–1.12)	RR 0.95 (0.86–1.05)
Comments on total operative delivery for failure to progress/other reasons	–	–	–	No errors found	Detailed IPD cannot be checked; errors in addition of cases: sum (n = 1416) doesn't fit with additions of CSFP (n = 763) and OVDFP (n = 652)
Total operative delivery	MA not performed	1920/7697 vs. 2040/7641 (24.94 vs. 26.70%) Fixed-effect ("among-study variance zero"): RR 0.94 (0.89–0.99)	MA not performed	MA not performed	RR 0.94 (0.88–1.01)
Comments on total operative delivery	–	Random-effect model yielded same results	–	–	Detailed IPD cannot be checked

CSFD, cesarean section for fetal distress; CSFP, cesarean section for failure to progress; OVDFD, operative vaginal delivery for fetal distress; OVDFP, operative vaginal delivery for failure to progress.

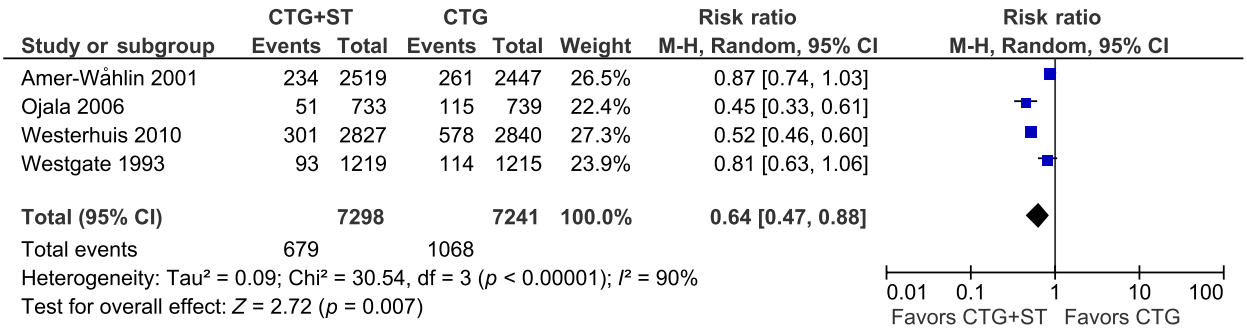

Figure 1. Forest plot and details of an aggregate meta-analysis of the usage of fetal scalp blood sampling in labor.

Table 4. Aggregate meta-analyses comparing CTG+ST vs. CTG alone. The Plymouth, Swedish, Finnish and Dutch RCTs were included in the meta-analyses, calculated with the COCHRANE REVIEW MANAGER statistical computer software version 5.2.7.

Outcome	No. of participants	RR (95% CI) fixed-effect	RR (95% CI) random-effect	I^2	Chi-square *p*-value
Fetal scalp blood sampling	14 539	0.63 (0.58–0.69)	0.64 (0.47–0.88)	90%	<0.00001
Total cesarean section[a]	14 539	1.00 (0.91–1.10)	1.00 (0.91–1.11)	10%	0.34
Fetal distress among all cesarean sections	1546	0.97 (0.77–1.22)	0.84 (0.54–1.32)	66%	0.03
Total operative vaginal delivery	14 539	0.88 (0.81–0.95)	0.88 (0.81–0.95)	0%	0.97
Fetal distress among all operative vaginal deliveries	1977	0.95 (0.85–1.06)	0.90 (0.72–1.12)	73%	0.01
Total operative delivery	14 539	0.93 (0.88–0.99)	0.93 (0.88–0.99)	0%	0.44
Fetal distress among all operative deliveries	3523	0.95 (0.86–1.04)	0.87 (0.68–1.10)	83%	0.0004

[a]Cesarean section data from the Plymouth RCT (7) were obtained from the Cochrane Review (1).

Dutch RCT data (384/2827 vs. 431/2840) becomes RR 0.89 (95% CI 0.83–0.96) (1044/7697 vs. 1162/7641); thus adding Dutch trial data slightly narrowed the CI.

The European consortium performed a "sensitivity analysis" that excluded the Plymouth RCT, based on its different ST analysis methodology. The sensitivity analysis resulted in a change of result from a total operative delivery RR of 0.94 (95% CI 0.88–0.99) (Table 3) to 0.95 (95% CI 0.89–1.00). While this change in RR is insignificant, it does result in a CI that includes unity.

The most detailed trial data were presented in the IPD MA by Schuit et al. (5). We found addition errors in this MA, as pointed out in Tables 3 and 6. For example, when the numbers of interventions for "fetal distress" and "failure to progress" are added, which, if not otherwise stated, are expected to include the total number of cesarean sections (CS) and instrumental vaginal deliveries, respectively, we found summary discrepancies in all figures of the individual RCTs [for details, see Table 3 in Schuit et al. (5)]. To illustrate, in the Swedish RCT the number of CSs for fetal distress was 194 and for failure to progress 217, resulting in 411 CSs. The number reported is 447, i.e. an excess of 36 cases. Similar discrepancies are noted for instrumental vaginal delivery (ventouse or forceps) and operative delivery (CS plus instrumental vaginal). It is unclear what the excess cases represent if they are unclassified operative deliveries or errors.

New meta-analyses of operative delivery

As shown in Table 3, the MAs varied in their analysis of operative delivery rates. All MAs included the French RCT, but for previously stated reasons we excluded the French trial and performed new MAs according to the following hierarchy of analyses and sub-analyses:

- total CS rate, with sub-analysis of CS for fetal distress (FD) among all CSs.
- total instrumental (operative) vaginal delivery rate, with sub-analysis of instrumental delivery for FD among all instrumental vaginal deliveries.
- total operative delivery rate (including CS and operative vaginal deliveries), with sub-analysis of operative delivery for fetal distress (ODFD) among total operative deliveries.

Details of the hierarchy of cases included in these MAs are shown in Table 5 and the results of the MAs are summarized in Table 4. The forest plot in Figure 2

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) 571–586

Table 5. Details and hierarchy of cases included in the meta-analyses of operative delivery (n = 14 539).

Meta-analysis	CTG+ST analysis n = 7298	... of whom had operation for fetal distress	CTG alone n = 7241	... of whom had operation for fetal distress
Cesarean section	777 (10.6%)	208 (26.8%)	769 (10.6%)	212 (27.6%)
Operative vaginal delivery	927 (12.7%)	358 (38.6%)	1050 (14.5%)	426 (40.6%)
Total operative delivery	1704 (23.3%)	566 (33.2%)	1819 (25.1%)	638 (35.1%)

demonstrates a significant 7% reduction in total operative delivery rate in the CTG+ST group (RR 0.93, 95% CI 0.88–0.99), mainly as a result of a significant 12% decrease in instrumental vaginal delivery rate (RR 0.88, 95% CI 0.81–0.95) (Table 4). The total CS rate was not affected. A minority of operative deliveries were performed for FD, 27–39% in the CTG+ST group and 28–41% in the CTG group (Table 5); sub-analyses showed no significant differences in ODFD among either CSs or instrumental vaginal deliveries (Table 4).

Metabolic acidosis: discrepancies in the meta-analyses

As mentioned above, in the Cochrane, American and Stockholm MAs, metabolic acidosis rates were a mixture of calculations using BD_{ecf} and BD_{blood} (Tables 2 and 6). As in the Finnish RCT, it appears that the difference between BD_{ecf} and BD_{blood} metabolic acidosis calculations was not considered, despite important differences in methodology. According to a personal communication between Welin and colleagues (25) and the principal Finnish author, Dr. Ojala, the figures of metabolic acidosis in extracellular fluid were 6/714 (0.8%) in the CTG+ST group and 4/722 (0.6%) in the CTG group (25,26). To the best of our knowledge, these data have not been published by the Finnish RCT authors.

The Cochrane Review aimed to analyse BD_{blood} metabolic acidosis and included the revised Dutch RCT BD_{blood} data (14) in the MA (Tables 2 and 6). The Dutch RCT rates for metabolic acidosis in the CTG+ST group vs. the CTG group, based on those calculated for blood, were 41/2827 (1.45%) vs. 66/2840 (2.32%), respectively; when calculations for metabolic acidosis in extracellular fluid were applied, the subsequent rates were much smaller, 19/2827 (0.67%) vs. 27/2840 (0.95%) (14). This is a crucial point in MAs because the incidence of BD >12.0 mmol/L may differ by a factor of 4 between BD_{blood} and BD_{ecf} calculations (27,28). This difference in definition of an essential RCT outcome variable would be considered a high risk bias according to the Cochrane Handbook for Systematic Reviews of Interventions (19).

The data in the Cochrane Review stated to represent rates of metabolic acidosis in the Swedish RCT, 12/2159 (0.56%, CTG+ST) vs. 24/2079 (1.15%, CTG alone), and in the Stockholm MA, 12/2519 (0.48%, CTG+ST) vs. 24/2447 (0.98%, CTG alone), are not those published by the Swedish RCT authors. The actual rates of metabolic acidosis in the Swedish RCT were 15/2159 (CTG+ST) vs. 31/2079 (CTG alone) in the original article (8) and 18/2565 (0.70%, CTG+ST) vs. 35/2484 (1.41%, CTG alone) in the revised article including imputed data (12).

As discussed above, there were six cases of metabolic acidosis in the CTG+ST analysis group and four in the CTG alone group in the Finnish RCT. Altogether 1472 cases were randomized in the RCT, with blood gas data available in 1436. However, the European consortium MA included all 1472 randomized cases as denominator

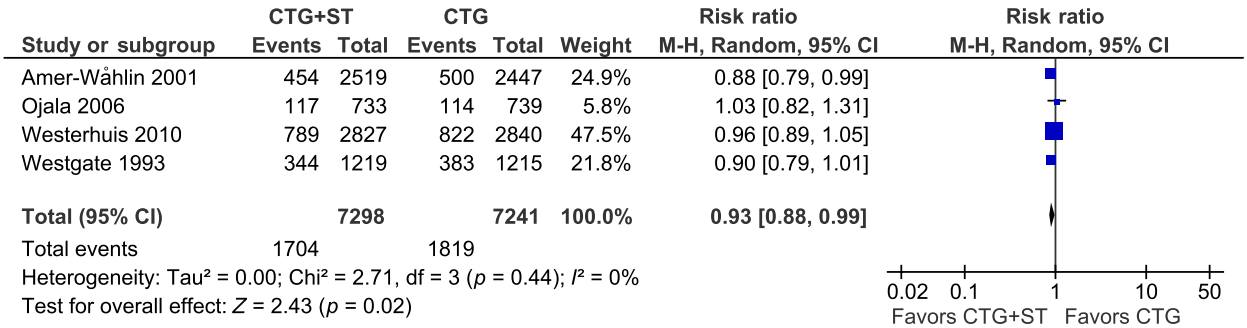

Figure 2. Meta-analysis of total operative delivery (sum of cesarean sections and instrumental vaginal deliveries). Data on total cesarean section from the Plymouth randomized controlled trial (Westgate et al., 1993) (7) were obtained from the Cochrane Review (1).

Table 6. Perinatal outcome. Calculations are CTG+ST analysis vs. CTG alone.

Meta-analysis	Cochrane review Neilson (2012)	European MA Becker et al. (2012)	American MA Potti & Berghella (2012)	Stockholm MA Salmelin et al. (2013)	IPD MA Schuit et al. (2013)
Apgar score <7 at 5 min	103/7678 vs. 108/7624 (1.34 vs. 1.42%) Fixed-effect (I^2 0.0%, chi-square p = 0.44): RR 0.95 (0.73–1.24)	103/7697 vs. 108/7641 (1.34 vs. 1.41%) Fixed-effect (Tau^2 0, I^2 0%): RR 0.95 (0.73–1.2)	103/7678 vs. 108/7624 (1.34 vs. 1.42%) Unclear MA type: RR 0.95 (0.73–1.24)	MA not performed	89/6524 vs. 78/6463 (1.36 vs. 1.21%) RR 1.14 (0.84–1.54)
Comments on Apgar score	–	Random-effect model yielded same results	Detailed RCT data not provided	–	–
Metabolic acidosis (pH <7.05 plus BD_{ecf} >12.0 mmol/L)	MA not performed	50/7697 vs. 73/7641 (0.65 vs. 0.96%) Random-effect (Tau^2 0.13, I^2 33%): RR 0.72 (0.43–1.19) Fixed-effect: RR 0.68 (0.48-0.97)	59/7318 vs. 81/7256 (0.81 vs. 1.12%) Random-effect (Tau^2 0.28, I^2 62%, p = 0.03): RR 0.80 (0.44–1.47)	51/6459 vs. 61/6409 (0.79 vs. 0.95%) Random-effect (Tau^2 0.28, I^2 63%, chi-square p = 0.04): RR 0.96 (0.49–1.88)	57/6524 vs. 73/6463 (0.87 vs. 1.13%) (Tau^2 0.09, I^2 42%): RR 0.76 (0.53–1.10)
Comments metabolic acidosis BD_{ecf} analysis	–	Finnish RCT BD_{ecf} data obtained via principal investigator, yet incorrect data in MA; incorrect data from Swedish RCT used	Original, not revised Swedish RCT data included; Finnish RCT metabolic acidosis from BD_{blood} data, not BD_{ecf}	Incorrect data from Swedish RCT used; Finnish RCT metabolic acidosis from BD_{blood} data, not BD_{ecf}	Additional cases in comparison with original or revised articles: Swedish +1, French +6; error in addition of cases
Metabolic acidosis (pH <7.05 plus BD_{blood} >12.0 mmol/L)	78/7318 vs. 113/7259 (1.06 vs. 1.56%) Random-effect (Tau^2 0.24, I^2 62%): RR 0.78 (0.44–1.37)	MA not performed	MA not performed	MA not performed	RR 0.82 (0.58–1.16)
Comments metabolic acidosis BD_{blood} analysis	MA is a mixture of BD_{blood} metabolic acidosis (Finnish, Dutch RCTs) and BD_{ecf} metabolic acidosis (Plymouth, Swedish, French RCTs)	–	–	–	Detailed IPD cannot be checked; Swedish and French RCTs excluded for unclear reasons
Cord artery pH <7.15	MA not performed	MA not performed	MA not performed	MA not performed	RR 0.99 (0.91–1.08)
Comments cord artery pH <7.15	–	–	–	–	Detailed IPD cannot be checked
Cord artery pH <7.05	MA not performed	Included 5 RCTs, detailed data not provided Random-effects (statistics not provided): RR 0.97 (0.64–1.5)	MA not performed	MA not performed	RR 0.87 (0.70–1.09)

Table 6. Continued

Meta-analysis	Cochrane review Neilson (2012)	European MA Becker et al. (2012)	American MA Potti & Berghella (2012)	Stockholm MA Salmelin et al. (2013)	IPD MA Schuit et al. (2013)
Comments cord artery pH <7.05	–	Data not reported in original or revised Swedish RCT, unclear how Swedish data were retrieved	–	–	Detailed IPD cannot be checked
Cord artery pH <7.00	MA not performed	MA not performed	MA not performed	MA not performed	RR 0.89 (0.62–1.26)
Comments on cord artery pH <7.00	–	–	–	–	Detailed IPD cannot be checked; error in addition of cases
Cord artery BD_{ecf} >12.0 mmol/L	MA not performed	MA not performed	MA not performed	MA not performed	RR 1.07 (0.90–1.29)
Comments on cord artery BD_{ecf} >12.0 mmol/L	–	–	–	–	Detailed IPD cannot be checked; error in addition of cases; low quality of French RCT data [see (6)]
Cord artery BD_{blood} >12.0 mmol/L	MA not performed	MA not performed	MA not performed	MA not performed	RR 0.98 (0.82–1.16)
Comments on cord artery BD_{blood} >12.0 mmol/L	–	–	–	–	Detailed IPD cannot be checked; error in addition of cases
Admission neonatal intensive care unit	615/7678 vs. 685/7624 (8.00 vs. 8.98%) Fixed-effect (I^2 0.0%, chi-square p = 0.97): RR 0.89 (0.81–0.99)	Included 5 RCTs, detailed data not provided Random-effect (statistics not provided): RR 0.90 (0.76–1.1)	264/7678 vs. 289/7624 (3.44 vs. 3.79%) Unknown MA type: RR 0.90 (0.76–1.06)	MA not performed	RR 0.92 (0.76–1.09)
Comments neonatal intensive care unit admission	–	Fixed-effect model yielded same results	Error in numbers of index cases, cannot be checked due to lack of detailed information in MA article	–	Detailed IPD cannot be checked; fewer cases included from Finnish and French RCTs than in original articles
Neonatal encephalopathy	8/7678 vs. 15/7624 (0.10 vs. 0.20%) Fixed-effect (I^2 0%, chi-square p = 0.51): RR 0.54 (0.24–1.25)	Sarnat & Sarnat grade ≥2 data only: included 3 RCTs, detailed data not provided Random-effect (statistics not provided): RR 0.66 (0.19–2.3)	8/7678 vs. 15/7624 (0.10 vs. 0.20%) Fixed-effect (I^2 0%, chi-square p = 0.51): RR 0.54 (0.24–1.25)	7/6478 vs. 11/6426 (0.11 vs. 0.17%) Random-effect (statistics not provided): RR 0.63 (0.24–1.63)	RR 0.42 (0.11–1.64)

Table 6. Continued

Meta-analysis	Cochrane review Neilson (2012)	European MA Becker et al. (2012)	American MA Potti & Berghella (2012)	Stockholm MA Salmelin et al. (2013)	IPD MA Schuit et al. (2013)
Comments neonatal encephalopathy	No uniform definition of encephalopathy in RCTs; no data in Plymouth RCT – data in MA provided by RCT authors	Fixed-effects model yielded same results; 3 RCTs included but Sarnat & Sarnat stage ≥2 data reported only in Swedish and Dutch RCTs	No uniform definition of encephalopathy in RCTs; no data in Plymouth RCT – unclear how Plymouth data were obtained	No uniform definition of encephalopathy in RCTs; figures for Sarnat & Sarnat stage ≥1 from Swedish RCT used but ≥2 from Dutch RCT	Detailed IPD cannot be checked; numbers fewer than in original RCT articles, reported hypoxic-ischemic encephalopathy
Neonatal intubation	7/714 vs. 9/722 (0.98 vs. 1.24%)	MA not performed because only 1 RCT	MA not performed	MA not performed	RR 0.64 (0.35–1.20)
Comments on neonatal intubation	Data available only from Finnish RCT	–	–	–	Detailed IPD cannot be checked; data from French and Dutch RCTs supplemented; incorrect data from Finnish RCT
Perinatal death	8/7697 vs. 5/7641 (0.10 vs. 0.065%) Fixed-effect (I^2 0.0%, chi-square p = 0.69): RR 1.49 (0.53–4.18)	Included 3 RCTs, detailed data not provided Random-effect (statistics not provided): RR 1.17 (0.38–3.6)	8/7697 vs. 5/7641 (0.10 vs. 0.065%) Fixed-effect (I^2 0%, chi-square p = 0.69): RR 1.49 (0.53–4.18)	MA not performed	RR 1.24 (0.33–4.61)
Comments on perinatal death	No data on perinatal death in Plymouth RCT – data in MA provided by RCT authors	Fixed-effects model yielded same results; data from 4 RCTs available, Finnish RCT data inexplicably excluded; unclear if data were corrected for lethal malformations	Perinatal death data included from 5 RCTs, but no data on how perinatal mortality data in the Plymouth RCT were obtained	–	Swedish RCT represented by mortality corrected for lethal malformations, but Dutch RCT by uncorrected mortality

when calculating metabolic acidosis rate rather than only those 1436 in which cord blood gas data were available. A similar error in data extraction was made from the Swedish RCT, in which the cases with missing cord blood gas data were included in the denominator. Further, the number of cases with metabolic acidosis (12 in the CTG+ST group and 24 in the CTG group) are not those published in the Swedish RCT.

The American MA authors have, like those of the Stockholm MA, also used the Finnish RCT BD_{blood} values rather than the BD_{ecf} values (Tables 2 and 6). Metabolic acidosis data from the original (8) but not from the revised (12) Swedish RCT article were included in the American MA, although the revised data from the Dutch RCT (14) were used. After imputation of missing data in the IPD MA, one additional case of metabolic acidosis using BD_{ecf} occurred in the Swedish RCT data file and six in the French RCT data file, whereas in the Finnish and Dutch RCT files there were no additions (Table 6). The Swedish authors themselves performed an imputation data analysis that resulted in 53 cases (12), compared with the 54 cases in the IPD MA. This discrepancy remains unexplained.

The French RCT reported the number of cases with BD_{ecf} >12.0 mmol/L. As discussed in the accompanying Part I review (6), as many as 15% of the cases in the RCT had a BD_{ecf} value >12.0 mmol/L and of these, only 15% fulfilled the cord blood sample validation criterion of an arteriovenous pCO_2 gradient >0.5 kPa defined in the study protocol. This confirms that the French RCT included cases at extraordinary high risk for fetal compromise, and indicates a poor quality of cord blood samples.

The IPD MA was the only one to evaluate metabolic acidosis with both BD_{ecf} and BD_{blood} >12.0 mmol/L (Table 6). BD_{blood} >12.0 mmol/L data were presented from the Finnish and Dutch RCTs but not from the Swedish and French RCTs. It is unclear why this approach was undertaken. The BD_{blood} calculation algorithm was not reported in the IPD MA article (or in the Finnish and Dutch RCTs), but with access to original individual participant pH and pCO_2 data from each trial there is no risk of discrepant *post hoc* calculations. The concept of an IPD MA is that all individual RCT data shoould be handled as if the MA was a single large multicenter RCT. This is important, since comparing different BD algorithms, the incidence of BD_{blood} >12.0 mmol/L may differ by more than 150% (2.5 times) (28). Different algorithms in the Finnish RCT and the IPD MA to calculate BD_{blood} might explain why the original 17 cases of BD_{blood} >12.0 mmol/L in the Finnish RCT increased to 23 when included in the IPD MA. Such large discrepancies between blood gas analyzers are clinically important: at low and moderately high BD_{blood} values the inter-analyzer difference might be 3–4 mmol/L and at high values up to 8–9 mmol/L (28).

New meta-analysis of neonatal metabolic acidosis

Our judgement is that the relevant rates of metabolic acidosis in extracellular fluid should be represented by data published in the original Plymouth and French RCT articles (7,10), the Swedish and Dutch revised data articles (12,14), and data presented by Welin and coworkers after communication with the principal Finnish RCT author (25). Figure 3 shows a forest plot with inclusion of these data from the Plymouth, Swedish, Finnish and Dutch RCTs: fetal surveillance with CTG+ST analysis resulted in a significant 39% reduction in metabolic acidosis compared with surveillance with CTG alone (RR 0.61, 95% CI 0.41–0.91).

Admission to the neonatal intensive care unit: discrepancies in the meta-analyses

Admissions to the neonatal intensive care unit (Table 6) were reported in all five RCTs. Fewer cases were included

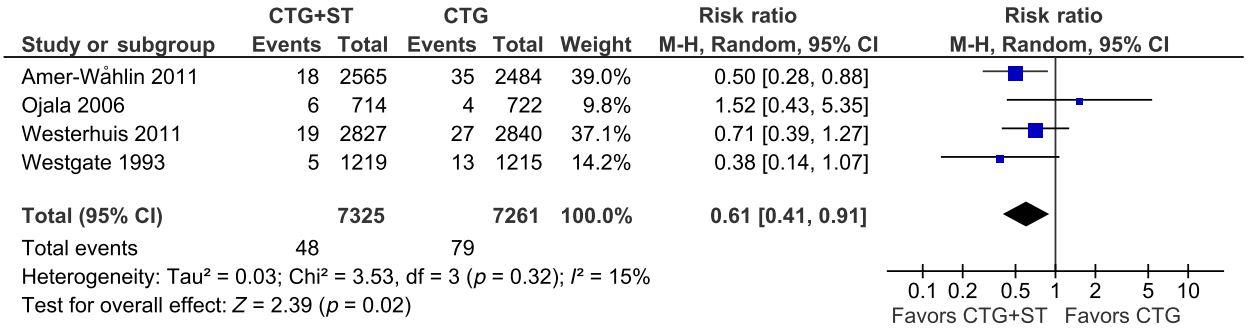

Figure 3. Meta-analysis of neonatal metabolic acidosis. Data from the Finnish randomized controlled trial (9) are from Dr. Ojala's personal communication with Welin et al. (25), the Swedish (Amer-Wåhlin et al., 2011) and Dutch (Westerhuis et al., 2011) data are from the revised articles (12,14), while the Plymouth data (Westgate et al., 1993) are from the original article (7).

in the IPD MA in the original Finnish RCT (n = 49 vs. 52) and French RCT (n = 10 vs. 11). Since retrieval of missing data cannot create fewer cases, these differences raise concern about bias or incorrect summations.

Neonatal encephalopathy: discrepancies in the meta-analyses

The Stockholm MA authors stated that encephalopathy was reported in all four of the RCTs that were included. This is incorrect, since the occurrence of neonatal seizures but not encephalopathy was reported in the French RCT (Table 6). Furthermore, they stated that hypoxic ischemic encephalopathy (HIE) was explicitly reported only in the Dutch RCT, but the Swedish RCT reported figures for HIE stage ≥1 and ≥2 separately. Both RCTs classified neonatal encephalopathy according to the criteria of Sarnat & Sarnat (29). The other RCTs either failed to report this outcome or failed to define its stage.

In the Cochrane, American and Stockholm MAs, neonatal encephalopathy represents a mixture of Sarnat & Sarnat encephalopathy stage ≥1 (Swedish RCT), stage ≥2 (Swedish, Dutch RCTs), unknown stage (Finnish RCT) and seizures (French RCT) (Table 6). The European MA reported data on neonatal encephalopathy stage ≥2 from three RCTs but, as mentioned above, such data were provided in only the Swedish and Dutch RCT articles. We determined that the third study included in the European MA was the Finnish RCT, in which encephalopathy was not defined. In this trial one neonate in the CTG group was diagnosed with encephalopathy but two had seizures, which might be in conflict with the Sarnat & Sarnat classification, since seizures usually represent Sarnat & Sarnat stage ≥2 encephalopathy. The Sarnat & Sarnat encephalopathy classification cannot be performed retrospectively by MA authors and for this reason we could not evaluate the correctness of the results reported in the European MA.

The IPD MA included seven cases of (undefined) HIE from the Swedish RCT. The Swedish authors themselves reported 11 cases of HIE stage ≥1, among whom three cases were stage 2. Similar discrepancies could be applied to the calculations in the Finnish and Dutch RCTs: in the former, the RCT authors reported one case of encephalopathy and two cases of seizures and the IPD MA authors included one case of HIE; in the latter, the RCT authors reported four cases of HIE stage ≥2 and the IPD MA authors included two cases. Such differences between the original RCTs and the IPD MA remain unexplained.

A new meta-analysis of neonatal encephalopathy

A comparison of Sarnat & Sarnat classified neonatal encephalopathy should only include data from the Swedish and Dutch RCTs. Our MA including stage ≥2 showed no effect of CTG+ST monitoring (RR 0.75, 95% CI 0.04–15.69) on the prevalence of this complication (Figure 4).

Neonatal intubation: discrepancies in the meta-analyses

The need for neonatal intubation was reported only in the Finnish RCT (n = 16), but data from the French and Dutch RCTs were supplemented in the IPD MA (Table 6). However, in the IPD MA only 12 cases from the Finnish RCT were included. Such a difference between the original trial and the IPD MA remains unexplained.

Perinatal mortality: discrepancies in the meta-analyses

Data on perinatal death (Table 6) were reported in all RCT articles except in the Plymouth trial. The Cochrane Review author contacted the authors of the original reports for missing data and the two deaths in the Plymouth trial (both in the CTG+ST group) were included in the MA. These deaths were also reported by Jennifer Westgate in her thesis (30). The American MA included the two Plymouth cases, while the Stockholm authors did not perform a MA.

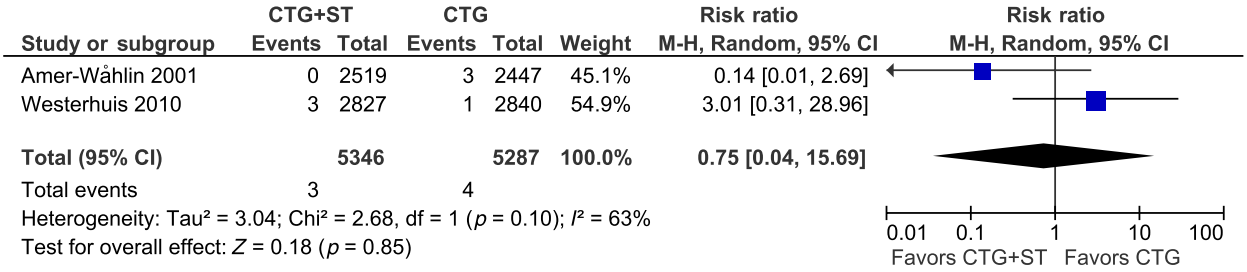

Figure 4. Meta-analysis of neonatal encephalopathy stage ≥2 according to Sarnat & Sarnat (29). The Sarnat & Sarnat classification was used only in the Swedish (Amer-Wåhlin et al., 2001) and Dutch (Westerhuis et al., 2010) trials (8,11).

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) 571–586

The European consortium MA authors inexplicably excluded the Finnish RCT mortality data. However, in the IPD MA, also performed by the European consortium, the Finnish data were included. In the IPD MA the Swedish RCT was represented by mortality corrected for lethal malformations but the Dutch RCT was represented by uncorrected mortality figures. These discrepancies raise concern about the interpretation of data on perinatal mortality.

Conclusions

To perform an MA, the included RCTs should address the same research question, be of comparable quality regarding selection bias, attrition rates and confounding variables, and include comparable populations (16). As discussed in the accompanying Part I review (6), there were considerable discrepancies in these aspects among the five RCTs. Furthermore, numerous errors appear to have occurred in the MAs, either created *de novo* in handling of the original or imputed data or through a failure to recognize some critical differences in data presentations in the RCTs. Metabolic acidosis, an essential perinatal outcome parameter, was presented as a mixture of BD_{blood} and BD_{ecf} data in the Cochrane, American and Stockholm MAs.

None of the five MAs contained complete and relevant data from all of the five RCTs. The decisions of the authors of the various MAs to include some or all of the RCT data in their analyses differed considerably. While the RCTs included in the MAs clearly differed in inclusion criteria, a question central to any MA is whether the system being studied was used as intended and was labelled by its manufacturer. With this in mind, we are of the opinion that the French RCT should have been excluded, since initiating ST monitoring in fetuses with clearly abnormal CTGs is contrary to existing guidelines. Conversely, the exclusion of the Plymouth RCT on the basis of its older technology would appear unwarranted, as this RCT, using manual rather than automated ST analysis, would to an even greater extent have challenged the ability of the CTG+ST analysis system to improve perinatal outcomes.

It is unfortunate that the IPD MA, with its potential clinical and statistical advantages over the aggregate MAs, was found to have several errors. This could have led to unintended bias in both the experimental and control groups. For the outcomes of FBS, operative delivery, ODFD, neonatal metabolic acidosis and neonatal Sarnat & Sarnat encephalopathy stage ≥2, we have performed new MAs. These showed not only, like the previous MAs, a significant reduction in FBS usage in the CTG+ST group (reduction of 36%), but also significant reductions in total operative delivery rate (reduction of 7%) and in neonatal metabolic acidosis rate (reduction of 39%). The results of the ongoing multicenter RCT in the United States (http://clinicaltrials.gov/ct2/show/NCT01131260) are some months away. Certainly the contribution of the USA data will help to determine whether the addition of ST analysis to conventional CTG results in improved perinatal outcomes.

Funding

No special funding for this review.

References

1. Neilson JP. Fetal electrocardiogram (ECG) for fetal monitoring during labour. Cochrane Database Syst Rev. 2012;(4):CD000116.
2. 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.
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 ST analysis 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. 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;93:556–69.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.
16. Ried K. Interpreting and understanding meta-analysis graphs – a practical guide. Aust Fam Physician. 2006;35:635–8.
17. Devane D. Systematic reviews. Statistical tests for heterogeneity. 27/02/2012. Available online at: www.iresearch4birth.eu/iResearch4Birth/resources/cms/documents/Statistical_tests_for_heterogeneity_Declan_Devane.pdf (accessed February 10, 2014).
18. Huedo-Medina T, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? 2006. *CHIP documents*. Paper 19. Available online at: http://digitalcommons.uconn.edu/chip_docs/19 (accessed February 10, 2014).
19. Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Available online at: http://handbook.cochrane.org/front_page.htm (accessed February 10, 2014).
20. StatsDirect Ltd. Available online at: http://www.statsdirect.com/webhelp/#contents.htm%3FTocPath%3D_____1 (accessed February 10, 2014).
21. Rosén KG. ST analysis reviewed. Am J Obstet Gynecol. 2013;209:394.
22. Schuit E. Reply. Am J Obstet Gynecol. 2013;209:394–5.
23. Sundström A-K, Rosén D, Rosén KG. Fetal surveillance. Gothenburg: Neoventa Medical AB, 2000.
24. Amer-Wåhlin I, Arulkumaran S, Hagberg H, Marsál K, Visser GHA. Fetal electrocardiogram: ST waveform analysis in intrapartum surveillance. BJOG. 2007;114:1191–3.
25. 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.
26. Norén H, Rosén KG. Intrapartum ST analysis. Fetal Maternal Med Rev. 2008;19:325–58.
27. 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.
28. 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.
29. Sarnat HB, Sarnat MS. Neonatal encephalopathy following fetal distress. A clinical and electroencephalographic study. Arch Neurol. 1976;33:696–705.
30. Westgate JA. An evaluation of electronic fetal monitoring with clinical validation of ST waveform analysis during labour. Thesis, University of Plymouth, 1993.

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) 571–586