Taylor & Francis
Taylor & Francis Group

The Journal of Maternal-Fetal & Neonatal Medicine

ISSN: 1476-7058 (Print) 1476-4954 (Online) Journal homepage: http://www.tandfonline.com/loi/ijmf20

Why STAN might not be dead

Serena Xodo, Gabriele Saccone, Ewoud Schuit, Isis Amer-Wåhlin & Vincenzo Berghella

To cite this article: Serena Xodo, Gabriele Saccone, Ewoud Schuit, Isis Amer-Wåhlin & Vincenzo Berghella (2017) Why STAN might not be dead, The Journal of Maternal-Fetal & Neonatal Medicine, 30:19, 2306-2308, DOI: 10.1080/14767058.2016.1247263

To link to this article: https://doi.org/10.1080/14767058.2016.1247263

Accepted author version posted online: 13 Oct 2016.
Published online: 02 Nov 2016.

THE JOURNAL OF MATERNAL-FETAL & NEONATAL MEDICINE

http://informahealthcare.com/jmf
ISSN: 1476-7058 (print), 1476-4954 (electronic)

J Matern Fetal Neonatal Med, 2017; 30(19): 2306–2308
 DOI: 10.1080/14767058.2016.1247263

Taylor & Francis
Taylor & Francis Group

DEBATE ARTICLE

Why STAN might not be dead

Serena Xodo[1], Gabriele Saccone [2], Ewoud Schuit[3], Isis Amer-Wåhlin[4], and Vincenzo Berghella [5]

[1]Department of Gynecology and Obstetrics, School of Medicine, University of Udine, Udine, Italy, [2]Department of Neuroscience, Reproductive Sciences and Dentistry, School of Medicine, University of Naples Federico II, Naples, Italy, [3]Department of Medicine, Stanford Prevention Research Center, Stanford University, Stanford, CA, USA, [4]Department of Learning Informatics Management and Ethics and Department of Women and Child Health, Medical Management Center, Karolinska Institute, Stockholm, Sweden, and [5]Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Sidney Kimmel Medical College of Thomas Jefferson University, Philadelphia, PA, USA

Abstract

Recently, a meta-analysis, including 26 526 laboring vertex singletons at term, summarized all available level-1 data from six high-quality randomized clinical trials (RCTs) on the use of ST analysis (STAN) during labor as an adjunct to conventional intrapartum fetal heart rate monitoring. The meta-analysis showed that STAN did not improve perinatal outcomes or decrease cesarean deliveries. Nonetheless, there are still reasons to believe STAN may have a role in the future research on intrapartum fetal monitoring. Out of six trials included in the meta-analysis, two included all cephalic singletons in labor, and four enrolled only high-risk pregnant women. This combination of both low- and high-risk populations may have distorted the potential impact of STAN. The test for heterogeneity between both subgroups was found to be statistically significant, indicating that the effect of STAN was different in high-risk women compared to a combination of both low- and high-risk women. Furthermore, the classifications of the fetal heart rate patterns used in the included randomized trials were different. Last but not least, despite >26 000 women with singleton gestations were included in the meta-analysis, the evidence still suffers from a lack of power, especially for subgroup analyses. In summary, while the level-1 data so far indicate overall no perinatal benefit of adding STAN to conventional intrapartum fetal heart rate monitoring for the outcomes most of interest, several issues point to the fact that more research is needed before the STAN technology can be deemed of no value for fetal monitoring in labor.

Keywords

Cardiotocography, STAN, labor, delivery, nonstress test

History

Received 20 June 2016
Accepted 8 October 2016
Published online 2 November 2016

Recently, we published a meta-analysis [1], summarizing all available level-1 data from six high-quality randomized clinical trials (RCTs) on the use of ST analysis (STAN) during labor in term singleton gestations as an adjunct to conventional intrapartum fetal heart rate monitoring (i.e. cardiotocography) [2–7]. The meta-analysis included 26 529 laboring vertex singletons at term and showed that STAN did not improve perinatal outcomes or decrease cesarean deliveries compared to cardiotocography (CTG) alone. Nonetheless, there are still reasons to believe STAN may have a role in future research and clinical care.

First: heterogeneity in inclusion criteria among trials

The incidence of cerebral palsy is about 2 in 1000 live births, and today approximately 10% of cerebral palsy cases are believed to be related to intrapartum asphyxia [8]. Given the very low incidence of this adverse outcome some RCTs adopted the strategy to analyze women with higher risk of intrapartum fetal hypoxia. Out of six RCTs on STAN included in the meta-analysis [1], two included all cephalic singletons in labor [4,7], and three enrolled pregnant women with high risk conditions such as maternal disease, prior obstetric complications, hypertensive disorders, intrauterine growth restriction, ruptured membranes for more than 24 h, postdates gestational age, failure to progress, need for pain relief, meconium-stained amniotic fluid or non-reassuring fetal heart rate (NRFHR) [2,3,6]. The French RCT included only women with abnormal CTG in labor with or without meconium-stained amniotic fluid [5]. This combination of both low- and high-risk populations may have distorted the potential impact of STAN found in the meta-analysis. When limiting the analysis to RCTs including only high-risk pregnant women (11 515 singletons included) [3–6], the use of the STAN showed a non-significant reduction by 27% in perinatal composite outcome (1.5% versus 2.0%; RR 0.73, 95% CI 0.53–1.00). When limiting the analysis to RCTs including both low- and high-risk women [4,7], no such effect was found (RR 1.14, 95% CI 0.85–1.53). In fact the test for

Address for correspondence: Vincenzo Berghella, Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, Thomas Jefferson University, 833 Chestnut Street, First Floor, Philadelphia, PA 19107, USA. E-mail: vincenzo.berghella@jefferson.edu

heterogeneity between both subgroups was found to be statistically significant ($p = 0.04$), indicating that the effect of STAN was different in high-risk women compared to a combination of both low- and high-risk women. Future meta-analysis using individual participant data from all studies may clarify whether STAN may be beneficial in a subgroup of high-risk women in labor of a singleton in cephalic presentation.

Second: learning curve

The STAN methodology depends on visual interpretation of the CTG trace [9]. Many observational studies investigating the effects of long-term use of STAN have confirmed that training and experience do improve the interpretation of the fetal heart rate patterns, leading to a decrease in the metabolic acidosis rate [10–13]. Evidence for an existing learning curve differed between trials included in the meta-analysis. The Swedish trial did find a difference between the first and second part of the trial [3]. The American RCT found no training effect which might be due to the low inclusion rate [7]. With most centers including only a few patients per month learning and experience with STAN seems hard to achieve. The remaining included trials have not investigated their learning curve. Using individual participant data from all trials, a subgroup analysis of the centers with highest inclusion rates would be of interest to further elucidate on this point.

Third: Hawthorne effect

Another possible explanation for the negative findings of some trials and consecutive meta-analyses [1,2,4–7], may be the change in behavior due to a subjects' awareness of being observed, called the Hawthorne effect [14]. The Hawthorne effect is actually a threat to the accuracy of any study collecting data via direct observation [15]. This phenomenon has four characteristics: 1. It is time dependent, which means that the magnitude and direction of the behavioral change depends on the total time practitioners using STAN are aware of being observed; 2. It has a performance ceiling, so the influence of the effect grows over time until reaching an upper limit; 3. Its impact decreases over time after peak performance; 4. Last but not least, behavior change occurs always after practitioners become aware of being watched [16]. It is reasonable to hypothesize that at least in the first part of some of the trials, practitioners' performance and previous experience with CTG but not with STAN might have improved the outcomes in the CTG arm, thereby limiting the effect of additional ST information in comparison to CTG alone. In the second part of the trials the Hawthorne effect, which is time dependent and decreases after the peak performance, might have reduced its influence on practitioners, allowing a more realistic comparison between CTG and STAN + CTG. In the Swedish RCT, a very obvious change occurred in the CTG arm over time and it was hypothesized that with increasing trust in the addition of STAN, the trust with CTG decreased, leading to an increase in operative labor in the CTG arm compared to the CTG + ST arm [3]. Again a subgroup analysis in centers where the inclusion rate permitted learning, would possibly share some light on this perspective.

Fourth: CTG three versus four tier classification

European RCTs used a four-tier classification of the fetal heart rate pattern. Belfort et al. [7], however, did not follow this system since the FDA had approved only the three-tier system. This is a crucial difference among trials which directly influences the clinical behavior, and thus the effect on outcome. In the four tier system classification, the NRFHR pattern has been subdivided in two CTG categories (intermediate and abnormal), thus modulating the cut offs of the different ST events, in order to discriminate NRFHR patterns needing intervention from those not needing intervention [17]. From intermediate to abnormal category the intervention cutoffs of ST events decrease, which means that the worse the CTG trace, the lower the value of ST event (if it occurs) suggesting clinical intervention. With only one category of NRFHR CTG, as in the three-tier classification, these gradations are not present, and intervention may be therefore advocated in situations where it could be avoided. A prospective observational study has demonstrated that the inter-observer agreement on classification of the intrapartum CTG is poor, especially in case of intermediary or abnormal CTG traces, and that the addition of fetal STAN led clinicians to decide whether to intervene or not in a more standardized way [18]. In fact, in the meta-analysis a statistically non-significant 21% decrease in perinatal composite outcome was found when analysis was limited to RCTs using only four-tier system ($n = 12\,987$) for the STAN (1.5% versus 1.9%; RR 0.79, 95% CI 0.61–1.03). This effect was not statistically different ($p = 0.09$) from the effect of STAN found in the study of Belfort et al. that used a tree-tier system (RR 1.14, 95% CI 0.83–1.57), however, given the effect estimates in both subgroups a differential effect could not be excluded [19].

Fifth: power

Despite >26 000 women with singleton gestations were included in the six RCTs, the evidence still suffers from a lack of power, especially for subgroup analyses related to study population and the system for STAN. To the future mother and her obstetrician or midwife, probably the most important outcome would be an improvement in neonatal outcome, e.g. the primary outcome of adverse composite perinatal outcome (intrapartum fetal death, neonatal death, Apgar score ≤ 3 at 5 min, neonatal seizure, metabolic acidosis, intubation for ventilation at delivery, or neonatal encephalopathy), and in particular in neonatal metabolic acidosis, which predicts later childhood function [20]. Our meta-analysis found an overall non-significant decrease by 26% in neonatal metabolic acidosis (0.5% versus 0.7%; RR 0.74, 95% CI 0.54–1.02). With an α of 0.05 and 80% power for all these following analyses, a sample size of about 29 316 patient is required, pointing to a type II error. The number of unselected women needed to screen by any STAN system to prevent a case of neonatal metabolic acidosis is about 545. When analysis is restricted to only high-risk women, a sample size of about 11 079 high-risk women is required, pointing to the fact that a few more of these women enrolled in the RCTs would have made results significant. The number of high-risk women needed to screen by STAN to prevent a case of

composite perinatal outcome is about 205. When analysis is restricted to only four-tier STAN system, a sample size of about 16 613 women is required, pointing again to a type II error. The number of unselected women needed to screen by four-tier system STAN to prevent a case of composite perinatal outcome is about 250.

In summary, while the level 1 data so far indicate no perinatal benefit of adding STAN to conventional intrapartum fetal heart rate monitoring for the outcomes most of interest, several issues point to the fact that more research (e.g. trials focusing on high-risk pregnancies) is needed before the STAN technology can be deemed of no value for fetal monitoring in labor.

Declaration of interest

The authors report no conflict of interest.

ORCiD

Gabriele Saccone http://orcid.org/0000-0003-0078-2113
Vincenzo Berghella http://orcid.org/0000-0003-2854-0239

References

1. Saccone G, Schuit E, Amer-Wåhlin I, et al. Electrocardiogram ST analysis during labor: a systematic review and meta-analysis of randomized controlled trials. Obstet Gynecol 2016;127:127–35.
2. Westgate J, Harris M, Curnow JS, Greene KR. Randomised trial of cardiotocography alone or with ST waveform analysis for intrapartum monitoring. Lancet 1992;340:194–8.
3. Amer-Wåhlin I, Hellsten C, Norén H, 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.
4. Ojala K, Vääräsmäki M, Mäkikallio K, et al. A comparison of intrapartum automated fetal electrocardiography and conventional cardiotocography–a randomised controlled study. BJOG 2006;113:419–23.
5. Vayssière C, David E, Meyer N, 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.
6. Westerhuis ME, Visser GH, Moons KG, 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.
7. Belfort MA, Saade GR, Thom E, et al. A randomized trial of intrapartum fetal ECG ST-segment analysis. N Engl J Med 2015;373:632–41.
8. Rei M, Ayres-de-Campos D, Bernardes J. Neurological damage arising from intrapartum hypoxia/acidosis. Best Pract Res Clin Obstet Gynaecol 2016;30:79–86.
9. Luce BR, Drummond M, Jönsson B, et al. EBM, HTA, and CER: clearing the confusion. Milbank Q 2010;88:256–76.
10. Kwee A, van der Hoorn-van den Beld CW, Veerman J, et al. STAN S21 fetal heart monitor for fetal surveillance during labor: an observational study in 637 patients. J Matern Fetal Neonatal Med 2004;15:400–7.
11. Norén H, Blad S, Carlsson A, et al. STAN in clinical practice—the outcome of 2 years of regular use in the city of Gothenburg. Am J Obstet Gynecol 2006;195:7–15.
12. 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;202:546.e1–7. doi: 10.1016/j.ajog.2009.11.033.
13. 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.
14. Gittelsohn J, Shankar AV, West KP, et al. Estimating reactivity in direct observation studies of health behaviors. Hum Organ 1997;56:182–9.
15. McCarney R, Warner J, Iliffe S, et al. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007;7:30–1.
16. Chen LF, Vander Weg MW, Hofmann DA, Reisinger HS. The Hawthorne effect in infection prevention and epidemiology. Infect Control Hosp Epidemiol 2015;36:1444–50.
17. Devoe LD, Ross M, Wilde C, et al. United States multicenter clinical usage study of the STAN 21 electronic fetal monitoring system. Am J Obstet Gynecol 2006;195:729–34.
18. Westerhuis ME, van Horen E, Kwee A, et al. Inter- and intra-observer agreement of intrapartum ST analysis of the fetal electrocardiogram in women monitored by STAN. BJOG 2009;116:545–51.
19. Schuit E, Amer-Wahlin I, Ojala K, et al. Effectiveness of electronic fetal monitoring with additional ST analysis in vertex singleton pregnancies at >36 weeks of gestation: an individual participant data metaanalysis. Am J Obstet Gynecol 2013;208:187.e1–.e13.
20. Randolph DA, Nole TL, Ambalavanan N. Generic Database and follow-up subcommittees for the Eunice Kennedy Shriver National Institute of Child-Health and Human Development Neonatal Research Network. Outcomes of extremely low birthweight infants with acidosis at birth. Arch Dis Child Fetal Neonatal Ed 2014;99:263–8.