Comparative analysis of the 2-year outcomes in the GRIT and TRUFFLE trials

Short title: combined analysis of GRIT and TRUFFLE trials

Wessel Ganzevoort [1], Jim G Thornton [2], Neil Marlow [3], Baskaran Thilaganathan [4], Birgit Arabin [5], Federico Prefumo [6], Christoph Lees [7 8], and Hans Wolf [1] for the GRIT study group* and the TRUFFLE-study group#

1 Department of Obstetrics, Amsterdam University Medical Center, Amsterdam, Netherlands
2 Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham City Hospital, Nottingham, UK
3 Department of Academic Neonatology, UCL Institute for Women's Health, London, UK
4 Fetal Medicine Unit, St George's University Hospitals NHS Foundation Trust and Molecular & Clinical Sciences Research Institute, St George's University of London, UK
5 Center for Mother and Child of the Philipps University, Marburg, Germany
6 Maternal–Fetal Medicine Unit, University of Brescia, Brescia, Italy
7 Department of Obstetrics & Gynaecology, Rosie Hospital, Cambridge, UK
8 Department of Obstetrics and Gynecology, KU Leuven, Belgium

*Listed in Appendix S1
#Listed in Appendix S2

Corresponding author:
Wessel Ganzevoort,
Dept of Obstetrics, Amsterdam UMC
Room H4-278, Meibergdreef 9
1105AZ Amsterdam, The Netherlands
j.w.ganzevoort@amc.uva.nl

Keywords: Fetal growth restriction; monitoring; cardiotocography; short term variation; ductus venosus;

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.20354

Contribution

What does this work add to what is already known?

This work provides comparisons (with reservations) of the effects of several monitoring techniques in the two randomized trials in early FGR.

What are the clinical implications of this work?

This analysis supports that fetal monitoring for early FGR can best be performed by the combination of cCTG and DV Doppler assessment.

Abstract

Objective: To explore the influence on perinatal outcome of different fetal monitoring strategies for preterm fetal growth restriction (FGR).

Design: Cohort analysis of individual participant data from the Growth Restriction Intervention Study (GRIT) and Trial of Umbilical and Fetal Flow in Europe (TRUFFLE) studies.

Setting: European multi-centre trials.

Population: All women from GRIT (n=238) and TRUFFLE (n=503), randomized between 26 and 32 weeks.

Methods: Women were categorized according to their monitoring-intervention method: A. immediate delivery (from GRIT), B. delayed delivery using conventional cardiotocography (CTG, from GRIT), C. delayed delivery using computerized CTG only (cCTG, from GRIT), D delayed delivery using cCTG only (from TRUFFLE) and E. delayed delivery using cCTG and ductus venosus (DV) Doppler (from TRUFFLE).

Primary outcome measure: Survival without impairment at two years.

Results: Gestational age at delivery and birth weight were similar in both studies. Fetal death rate was similar between GRIT and TRUFFLE, but neonatal and late death were more frequent in GRIT (18% vs. 6%; p<0.01). The primary outcome was least common in groups A (70%; 95% confidence interval [CI] 61-78), and B (69%; 95% CI 57-82), and increased with more advanced monitoring in C (80%; 95% CI 68-91) and D (77%; 95% CI 70-84) and was highest in E (84%; 95% CI 80-89); (p trend <0.01).

Conclusions: This analysis supports that fetal monitoring for early FGR can best be performed by the combination of cCTG and DV Doppler assessment.

Trial Registration: GRIT ISRCTN41358726 and TRUFFLE ISRCTN56204499.

Introduction

Fetal growth restriction (FGR) in the early preterm period before 32 weeks is a rare, but serious complication due to its association with adverse perinatal outcome. Treatment of the underlying condition is impossible and the challenge lies in optimal timing of delivery. The risks of prematurity (neonatal complications and impaired neurodevelopment) have to be balanced against the risks of prolonged fetal exposure to hypoxaemia and acidaemia, possibly resulting in stillbirth and brain damage. Obstetricians use a range of tests of fetal wellbeing. However, the sequences of fetal deterioration are difficult to evaluate in humans because observational case series of growth restricted fetuses rarely, if ever, include all tests of wellbeing. More importantly study outcomes are inevitably "censored" because the timing of delivery is subject to the vagaries of parental choice and the managing clinicians.[1] The optimal indication for timing of delivery is still open for debate.[2]

Trials in this patient group are hard to conduct. To date, only two large randomised trials have evaluated how to time delivery in early preterm FGR: GRIT and TRUFFLE.[3, 4] Both have been highly cited, but their influence on practice is difficult to measure. Even though GRIT recruited 548 women (588 babies) and TRUFFLE recruited 503 women, and both trials achieved very high follow-up rates up to two years, the number of participants at any specific gestational age, or with similar clinical risk factors was small. Given how clinical decisions are so specific for gestational age the available analyses are necessarily underpowered.

In the absence of data from further trials, but while awaiting those, the obvious interim solution is meta-analysis of the existing ones. Ideally this should be an individual patient data meta-analysis (IPD M-A) and both trial databases are available. However, the differences in inclusion criteria and in study period between the two trials precluded such an analysis.

The objective of the current analysis was to explore the influence on perinatal outcome of different fetal monitoring strategies used in these two trials with similar populations regarding gestational age and severity of FGR.

Methods

Because the intention was to perform a meta-analysis on trials of fetal monitoring with long-term infant outcome results we performed a scoping literature search in Pubmed, with the terms "(fetal compromise[Title/Abstract] or fetal monitoring[Title/Abstract]) AND (growth restriction[Title/Abstract] OR fetal growth[Title/Abstract]) AND (long term[Title/Abstract] OR long-term[Title/Abstract] OR wellbeing[Title/Abstract] OR neurodevelopment[title/abstract] OR Griffith[Title/Abstract] OR Bayley[Title/Abstract])" and with a limit on clinical trials. This confirmed that GRIT and TRUFFLE were the only monitoring-intervention studies in early-onset fetal growth restriction.[3, 4] Individual participant data were retrieved from both trial datasets. Methods of both studies were previously described in full and summarized below.

GRIT

In GRIT[3, 5], 548 pregnant women between 24 and 36 completed weeks were recruited in 69 European hospitals between 1993-2001. All fetuses had suspected FGR and the inclusion criterion was clinical uncertainty about whether immediate delivery was indicated. Fetal arterial Doppler and cardiotocography were recorded before inclusion, but interpretation of these findings was left to local standards. Ductus venosus Doppler was not used. Women were randomly allocated to immediate delivery (n=296) or to delivery, which was delayed until the obstetrician was no longer uncertain (n=292). Mode of delivery and monitoring strategies for the deferred group were left up to the attending obstetrician. The main outcome was death or impairment at or beyond 2 years of age. Impairment was a composite outcome comprising any of cerebral palsy, little or no vision, requirement for a hearing aid, or a Griffith's Mental Development Scales General Quotient of 70 or less, assessed by a Griffith-trained assessor. For those babies who were not seen personally after two years the former three diagnoses were accepted from either parental report or the child's family practitioner or paediatrician. The overall rate of death or severe impairment at 2 years was 17.2% in those with known outcome. There was no statistically significant

difference between the groups. A neurological impairment occurred in 6.5% of surviving and assessed children. For the purpose of this analysis the original study abnormal endpoint definition was used, i.e. survival without impairment at or beyond 2 years of age.

TRUFFLE

In TRUFFLE[4, 6] 503 pregnant women between 26 and 32 completed weeks were recruited by 20 European hospitals between 2005-2010. All had fetuses with FGR, defined by elevated umbilical artery pulsatility index (PI) and ultrasound biometry. Women were randomized to one of three groups where delivery was determined by either reduced short-term variation in fetal heart rate using computerized CTG (cCTG), or to one of two criteria based on Doppler ultrasound assessment of the ductus venosus waveforms, namely early (PI >95th centile) or late abnormalities (absent or negative A-wave). Criteria were specified in detail and included specific cut-offs for STV in all groups, although the intensity of monitoring was not prescribed. All ultrasonographers met predetermined criteria for performing ductus venosus Doppler measurements. The main outcome was survival without neurological impairment at two years. Impairment comprised any of cerebral palsy, severe vision or hearing impairment, or a cognitive composite score from the Bayley Scales of Infant and Toddler Development (third edition) of less than 85, assessed by an assessor, trained and accredited specifically for the trial. For those babies who were not seen personally after two years the former three diagnoses were accepted from either parental report or the child's family practitioner or paediatrician. The overall rate of death or severe impairment at 2 years was 18.1% of all infants with a known outcome, with no statistical differences between the groups. The overall rate of neurological impairment at 2 years in surviving assessed children was 9.7%. This was lower in the DV groups than in the cCTG group.

Data strategy

For this analysis, we used data from the subset of participants in GRIT with singleton pregnancy and gestational age between 26 and 32 completed weeks at study entry. From TRUFFLE all women were selected, except one where neonatal data were missing. Baseline variables, process variables and outcomes were compared between groups, with specific emphasis on their relationship with antenatal death and 2-year outcome. We combined 2-year outcome data, defining 'survival free of impairment as survival free of cerebral palsy, severe visual impairment or hearing loss requiring aids, and either a Griffith Quotient >70 (GRIT) or Bayley-III (or adjusted Bayley II) Cognitive composite score >85 (TRUFFLE). The primary outcome was calculated for all infants with known outcome, including all perinatal and late death, but excluding infants lost to two year follow-up.

Analysis strategy

Data were classified for different monitoring and intervention strategies: A. immediate delivery when fetal condition was uncertain (from GRIT), B. delayed delivery using conventional CTG (from GRIT), C. delayed delivery using cCTG with STV calculation (from GRIT), D. using cCTG with STV calculation (from TRUFFLE CTG group) and E. delayed delivery using cCTG with STV calculation and ductus venosus (DV) Doppler (both DV groups from TRUFFLE combined, as results of these groups were not statistically different). A second analysis aimed at differences in infant outcome over time. The years of the studies were grouped as 1994 to 1997, 1998 to 2001, 2005 to 2007 and 2008 to 2010, dividing both studies in a first and second half.

Parameters for analysis included monitoring strategy classification, year of randomisation grouping, gestational age at randomisation in weeks, umbilical artery absent or reversed (ARED) flow, and birth weight Z-score. Birth weight Z scores were calculated using an in-utero fetal weight model developed by Hadlock et al.[7] Because in GRIT an estimated fetal weight at inclusion was not recorded we included birth weight in the analysis.

Statistics

Baseline characteristics, process variables and outcomes were compared two-sided for statistical significance by ANOVA, the Mann–Whitney U-test or Pearson's chi-square test, as appropriate.

The association of demographic, clinical and diagnostic parameters at study inclusion with the endpoints was first explored by univariable analysis. Those parameters that were significantly different between infants with normal 2-year outcome and infants with death or neurological impairment in univariable analysis were entered in a multivariable logistic regression analysis to adjust for association between parameters and to calculate odds ratios (OR).[7, 8] A regression analysis was planned first without gestational age or birth weight and secondly with these parameters, to determine if the odds ratios of intervention strategy classification or study period was affected by these parameters by possible collider bias.

Logistic regression analysis was started with the intervention strategies or with study period groups and other parameters were added stepwise, based on probability for the primary endpoint (healthy survival at two years). The probability for entry in the model was set at 0.05 and removal at 0.10. Statistical calculations were performed using Statistical Package for the Social Sciences (SPSS program, version 25; IBM Corp., New York, NY, USA).

Institutional review

This study was exempted from review by the institutional ethics review boards.

Results

We included 238 women from GRIT and 502 women from TRUFFLE. Gestational age at inclusion or at delivery, and birth weight were similar between all subgroups (Table 1). Because of the randomization sequence, allocation to different groups was evenly spread across the years. More women had absent or reversed umbilical artery (ARED) flow at study entry in GRIT than in TRUFFLE (70% vs. 41%; p<0.01). However, if later measurements were included rates of umbilical ARED flow were comparable between the groups (on average 64%). The interval to delivery was shorter in the GRIT delayed delivery groups than in the TRUFFLE groups (median 3 days versus 8 days, p<0.01). Fetal death rate was comparable between GRIT and TRUFFLE, but neonatal and late death was more frequent in GRIT (18% vs. 6%; p<0.01). Fetal death rate was similar in women monitored by conventional CTG without STV compared to monitoring by cCTG (both delayed GRIT; 5 vs. 6%), but perinatal mortality was lower in the cCTG GRIT group (13% vs 23%, not statistically significant). The primary outcome (survival without neurological impairment at 2 years) was worst in the immediate delivery and CTG without STV groups, intermediate in the cCTG groups and best in the cCTG + DV strategy group (Pearson chi-square <0.01). This is also graphically presented in figure 1 showing a decreasing perinatal death rate across groups, most apparent for those included at the lowest gestational age group (26-27 weeks). A similar trend was observed when data were grouped for year of randomization (Pearson chi-square <0.01) (figure 2). Other parameters that were significantly associated with the primary outcome were gestational age at randomization, umbilical ARED flow, birth weight and birth weight Z-score.

Because the study periods of GRIT and TRUFFLE differed (1993-2001 versus 2005-2010) and study period and intervention group classification were highly associated (Pearson correlation 0.82) it was not justified to combine these in one regression analysis. We therefore performed a separate regression analysis for monitoring groups and for year of inclusion with the other parameters that were significantly associated with the primary

outcome, except birth weight and birth weight Z-score. Birth weight and birth weight Z-score were omitted because they were correlated with both gestational age and ARED flow. Odds ratios of monitoring strategy and study period were not affected by addition of gestational age, but the area under the curve (AUC) of a receiver operating characteristic (ROC) curve of the regression model increased from 0.63 to 0.69. Odds ratios for the model are shown in figure 3a and b. Gestational age at inclusion and umbilical ARED flow, which both are measures for the severity of FGR, were highly associated with the primary outcome. Year of inclusion and monitoring strategy had a similar, moderate effect.

Discussion

Main findings

We observed a trend for improved long-term infant outcome with less neonatal death and less neurological impairment using more advanced fetal monitoring strategies after adjustment for severity of FGR using gestational age at delivery and umbilical ARED flow.

Strengths and Limitations

This analysis was on the outset intended to perform a meta-analysis with individual patient data. Conventional meta-analysis proved impossible, because the inclusion selection and the type of interventions were entirely different.

Although inclusion criteria differed between GRIT and TRUFFLE, the study populations are very similar with regard to gestational age, birthweight and birth weight Z-score. Absent/reversed umbilical Doppler flow at inclusion was more frequent in GRIT. GRIT participants were included when there was uncertainty regarding the necessity of delivering the baby and were at inclusion probably in a later stage of the pathological process of FGR than TRUFFLE participants. As a result of this, in TRUFFLE the interval of inclusion to delivery was longer and with later Doppler measurements ARED flow became as frequent as in GRIT.

A difficult problem was that the trials were performed during subsequent time periods and that it was impossible to separate outcome improvement due to advances in obstetric and neonatal care from the monitoring strategy differences between the trials. On statistical analysis both study period and trial allocation were highly correlated. Both seemed to have an effect, but the proportional contribution of these parameters to the primary outcome could not be determined exactly.

Birth weight and gestational age have a potential for collider stratification bias: they are not independent parameters in a population with early preterm FGR. They are influenced by severity of FGR, monitoring methods and the decision to deliver the baby at a certain moment. Of course, it would have been better if more data on underlying pathology of FGR

were known and available to improve outcome prediction. Unfortunately, except from Doppler data, there are no appropriate markers for underlying pathology of FGR. Therefore, gestational age in this study should be regarded as a proxy for the severity of FGR, and not as an independent variable. The study population is rather uniform in its selection through functional Doppler markers of FGR and we expect that this resulted in a similarity of underlying pathologies that caused FGR in the intervention groups. A further argument that collider bias by gestational age is unlikely is the observation that a regression analysis with or without gestational age resulted in similar odds ratios for monitoring strategy or study period, while the AUC of the ROC curve for the model increased by addition of gestational age.

The two studies used two different tests to assess neurodevelopmental outcome, Griffiths General Quotients in GRIT and Bayley Cognitive Composite scores in TRUFFLE, and used different cut offs commensurate with prevailing practice. By using a cut-off <85 in TRUFFLE, and a stricter less prevalent cut-off of <70 in GRIT it would be unlikely that differences in developmental outcome were inflated in advantage of TRUFFLE.[9] This supports our conclusion that outcome was really better in TRUFFLE DV groups.

GRIT was a study on delivery in women with FGR when the clinician was uncertain whether to deliver the baby or not. The attending obstetrician was free to decide on fetal monitoring methods and the mode of delivery. TRUFFLE was designed to compare different monitoring strategies and delivery criteria were specifically defined for each strategy. Cut-offs for STV had been defined based on the studies of Dawes et al., based on associations with fetal acidaemia.[10] DV Doppler cut-offs were substantiated by longitudinal studies into monitoring parameters and adverse perinatal outcomes.[1, 11] However, the minimal frequency of monitoring was specified liberally (DV Doppler once weekly, cCTG twice a week). Lack of a fixed, frequent schedule for monitoring in TRUFFLE and the complete absence of such data from GRIT precludes a secondary longitudinal analysis of monitoring data, which could have given more information of their relative impact on decision making.

In both studies deliveries were indicated by other criteria than fetal distress. In GRIT indication for delivery was not included in the database, while we know that in TRUFFLE 30% were delivered outside the pre-specified monitoring criteria.[12] Also, there were insufficient data regarding co-interventions such as corticosteroids (probably in nearly all patients) and magnesium (probably only in a few patients for eclampsia prevention, fetal neuroprotection was not used then). These restrictions hamper the combined analysis of the data targeted at the effect of the monitoring strategies.

Our intention to test the superiority of cCTG over conventional CTG was only partly successful. Only few articles have compared the predictive value of cCTG with visual inspection of the CTG, all use only short-term outcomes and are underpowered for perinatal outcomes. Turan et al. described monitoring parameters in a cohort of 56 growth-restricted fetuses and showed a slightly better correlation of low STV with low umbilical cord pH as compared to traditional CTG.[15] Another study showed no significant relationship between both methods and fetal or neonatal survival.[16] To our knowledge, there are no published data comparing the efficacy of cCTG versus standard CTG using visual inspection. Currently, probably due to the limited association with fetal hypoxemia and acidaemia of cCTG,[17, 18] there is significant practice variation in the use of cCTG.

Interpretation

We observed an improvement of infant outcome over time and by advancement of fetal monitoring. This correlation is, of course, not only statistical. It is logical that advancement in obstetric and neonatal care over the past 20 years have an effect and that advancements in fetal monitoring methods form a part of this.

In the delayed GRIT group with cCTG infant outcome was better than in the group with visual assessment, although this difference was not statistically significant and might have been caused by other differences between centres that used cCTG and those that did not. The main advantage of cCTG is that it provides a numerical result, which allows for a strictly protocolled intervention decision, as applied in TRUFFLE strategy. Its use is therefore

essential for future intervention trials. However, it has not been proven to be superior in a head to head comparison.

Conclusion

This comparative analysis supports the hypothesis that fetal monitoring for early preterm FGR can best be performed by the combination of cCTG and DV Doppler assessment.

Acknowledgements: We gratefully acknowledge the support of Prof Andy Vail, Centre for Biostatistics, University of Manchester for the provision of the relevant GRIT study data.

The authors in the TRUFFLE study group in Appendix S2 meet the requirements for pubmed-indexation for their role in the conducting of the study and approve the content of the manuscript.

Funding: This analysis was performed without funding. The original TRUFFLE-study was partly funded by Grant Number 94506556, ZonMw, POBox 93245, 2509 AE Den Haag, The Netherlands. The GRIT study was funded by the United Kingdom Medical Research Council, a European Union Concerted Action and the Dutch Princess Beatrix Foundation.

References

1. Bilardo CM, Wolf H, Stigter RH, Ville Y, Baez E, Visser GH, Hecher K. Relationship between monitoring parameters and perinatal outcome in severe, early intrauterine growth restriction. *Ultrasound Obstet Gynecol* 2004; **23**: 119-125.
2. Brodszki J, Morsing E, Malcus P, Thuring A, Ley D, Marsal K. Early intervention in management of very preterm growth-restricted fetuses: 2-year outcome of infants delivered on fetal indication before 30 gestational weeks. *Ultrasound Obstet Gynecol* 2009; **34**: 288-296.
3. TheGRITStudyGroup. A randomised trial of timed delivery for the compromised preterm fetus: short term outcomes and Bayesian interpretation. *Br J Obstet Gynaecol* 2003; **110**: 27-32.
4. Lees CC, Marlow N, van Wassenaer-Leemhuis A, Arabin B, Bilardo CM, Brezinka C, Calvert S, Derks JB, Diemert A, Duvekot JJ, Ferrazzi E, Frusca T, Ganzevoort W, Hecher K, Martinelli P, Ostermayer E, Papageorghiou AT, Schlembach D, Schneider KT, Thilaganathan B, Todros T, Valcamonico A, Visser GH, Wolf H, group Ts. 2 year neurodevelopmental and intermediate perinatal outcomes in infants with very preterm fetal growth restriction (TRUFFLE): a randomised trial. *Lancet* 2015; **385**: 2162-2172.
5. Thornton JG, Hornbuckle J, Vail A, Spiegelhalter DJ, Levene M. Infant wellbeing at 2 years of age in the Growth Restriction Intervention Trial (GRIT): multicentred randomised controlled trial. *Lancet* 2004; **364**: 513-520.
6. Lees C, Marlow N, Arabin B, Bilardo CM, Brezinka C, Derks JB, Duvekot J, Frusca T, Diemert A, Ferrazzi E, Ganzevoort W, Hecher K, Martinelli P, Ostermayer E, Papageorghiou AT, Schlembach D, Schneider KT, Thilaganathan B, Todros T, van Wassenaer-Leemhuis A, Valcamonico A, Visser GH, Wolf H, Group T. Perinatal morbidity and mortality in early-onset fetal growth restriction: cohort outcomes of the trial of randomized umbilical and fetal flow in Europe (TRUFFLE). *Ultrasound Obstet Gynecol* 2013; **42**: 400-408.
7. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991; **181**: 129-133.
8. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav Res Methods Instrum Comput* 2004; **36**: 717-731.
9. Cirelli I, Bickle Graz M, Tolsa JF. Comparison of Griffiths-II and Bayley-II tests for the developmental assessment of high-risk infants. *Infant Behav Dev* 2015; **41**: 17-25.
10. Dawes GS, Moulden M, Redman CW. Short-term fetal heart rate variation, decelerations, and umbilical flow velocity waveforms before labor. *Obstet Gynecol* 1992; **80**: 673-678.
11. Hecher K, Bilardo CM, Stigter RH, Ville Y, Hackeloer BJ, Kok HJ, Senat MV, Visser GH. Monitoring of fetuses with intrauterine growth restriction: a longitudinal study. *Ultrasound Obstet Gynecol* 2001; **18**: 564-570.
12. Visser GH, Bilardo CM, Derks JB, Ferrazzi E, Fratelli N, Frusca T, Ganzevoort W, Lees C, Napolitano R, Todros T, Wolf H, Hecher K, investigators Tg, Marlow N, Arabin B, Brezinka C, Diemert A, Duvekot JJ, Martinelli P, Ostermayer E, Papageorghiou AT, Schlembach D, Schneider K, Thilaganathan B, Valcamonico A. The TRUFFLE study; fetal monitoring indications for delivery in 310 IUGR infants with 2 year's outcome delivered before 32 weeks of gestation. *Ultrasound Obstet Gynecol* 2016. DOI:10.1002/uog.17361.
13. Turan S, Turan OM, Berg C, Moyano D, Bhide A, Bower S, Thilaganathan B, Gembruch U, Nicolaides K, Harman C, Baschat AA. Computerized fetal heart rate

analysis, Doppler ultrasound and biophysical profile score in the prediction of acid-base status of growth-restricted fetuses. *Ultrasound Obstet Gynecol* 2007; **30**: 750-756.
14. Madazli R. Prognostic factors for survival of growth-restricted fetuses with absent end-diastolic velocity in the umbilical artery. *J Perinatol* 2002; **22**: 286-290.
15. Kapaya H, Jacques R, Rahaim N, Anumba D. "Does short-term variation in fetal heart rate predict fetal acidaemia?" A systematic review and meta-analysis. *J Matern Fetal Neonatal Med* 2016; **29**: 4070-4077.
16. Pels A, Mensing van Charante N, Vollgraff Heidweiller-Schreurs CA, Limpens J, Wolf H, de Boer M, Ganzevoort W. The prognostic accuracy of short term variation of fetal heart rate in early-onset fetal growth restriction: a systematic review. *Eur J Obstet Gynecol Reprod Biol, accepted for publication* 2019.

Figure legends

Figure 1: Perinatal death rate and neurodevelopmental impairment at 2 years of age specified for group classification and gestational age at randomisation of women participating in GRIT and TRUFFLE.

Figure 2: Perinatal death rate and neurodevelopmental impairment at 2 years of age specified for period and gestational age at randomisation of women participating in GRIT and TRUFFLE.

Figure 3: Odds ratios with 95% confidence interval for survival without neurological impairment at 2 years, calculated by multivariable regression analysis, with:

A – Monitoring group classification: AUC 0.69 (95% CI 0.64 to 0.74)

B – Year of inclusion: AUC 0.68 (95% CI 0.63 to 0.73)

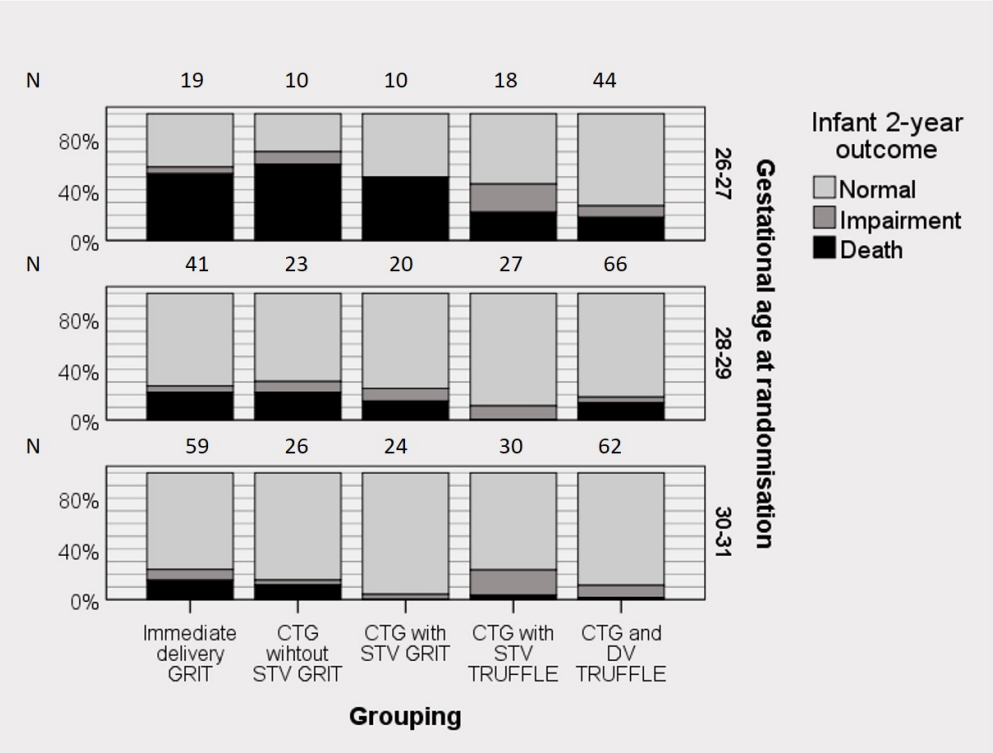
Table 1: Perinatal and outcome characteristics specified for fetal monitoring strategy of women participating in GRIT and TRUFFLE

A. immediate delivery when fetal condition was uncertain (from GRIT), B. delayed delivery using CTG without STV calculation (from GRIT), C. delayed delivery using cCTG-STV calculation (from GRIT)), D. delayed delivery using cCTG-STV calculation (from TRUFFLE CTG group) and E. delayed delivery using cCTG-STV and DV assessment (both DV groups from TRUFFLE combined).

	GRIT			TRUFFLE		Total
	Immediate delivery (A)	Delayed delivery				
		vCTG no STV (B)	cCTG-STV (C)	cCTG-STV (D)	cCTG-STV and DV (E)	
N	121	62	55	165	337	740
Nulliparous	74 (61%)	33 (53%)	33 (60%)	100 (61%)	218 (65%)	458 (62%)
Gestational age at inclusion	29.5 (28.5 to 31.0)	29,5 (28.5 to 31.5)	29.5 (28.5 to 30.5)	29.2 (27.9 to 30.1)	29.2 (27.9 to 30.4)	29.5 (28.1 to 30.5)
Interval to delivery (days) ‡	0 (0 to 1)	4 (2 to 8)	2 (1 to 8)	7 (2 to 17)	8 (3 to 17)	5 (1 to 14)
Umbilical ARED flow inclusion ‡	85 (70%)	44 (71%)	37 (67%)	61 (37%)	147 (44%)	374 (51%)
Umbilical ARED flow any time	85 (70%)	44 (71%)	37 (67%)	94 (57%)	210 (62%)	470 (64%)
Fetal death	2 (2%)	3 (5%)	3 (6%)	2 (1%)	10 (3%)	20 (3%)
Live born	119 (98%)	59 (95%)	52 (94%)	163 (99%)	327 (97%)	720 (97%)
Gestational age at delivery	30.5 (28.6 to 31.5)	30.6 (29.0 to 32.4)	30.7 (29.6 to 31.6)	30.6 (29.0 to 32.0)	30.7 (29.3 to 32.3)	30.6 (29,1 to 31.9)
Birth weight	880 (740 to 1100)	920 (745 to 1085)	928 (733 to 1068)	965 (800 to 1115)	990 (806 to 1200)	953 (780 to 1160)
Birthweight Z-score	-3.3 (-3.9 to -2.4)	-3.3 (-4.3 to -2.7)	-3.5 (-4.1 to -2.8)	-3.3 (-3.8 to -2.7)	-3.3 (-3.8 to -2.8)	-3.3 (-3.8 to -2.8)
Neonatal death‡	23 (19%)	11 (19%)	4 (8%)	10 (6%)	17 (5%)	65 (9%)
Late death < 2 years[&]	3 (3%)	0 (---)	1 (2%)	1 (1%)	1 (0%)	6 (1%)
Evaluated at 2 years	91 (98%)	45 (94%)	46 (98%)	131 (86%)	271 (88%)	584 (90%)
Severe impairment†*	4 (4%)	2 (4%)	1 (2%)	6 (4%)	3 (1%)	16 (3%)
Abnormal development *	8 (9%)	4 (9%)	3 (7%)	20 (15%)	19 (7%)	54 (9%)
Composite primary outcome Alive with normal development at two years ‡	83	41	43	111	252	530
% of known outcome; (95% CI) #	70%; (61% to 78%)	69%; (57% to 82%)	80%; (68% to 91%)	77%; (70% to 84%)	84%; (80% to 89%)	79%; (75% to 82%)

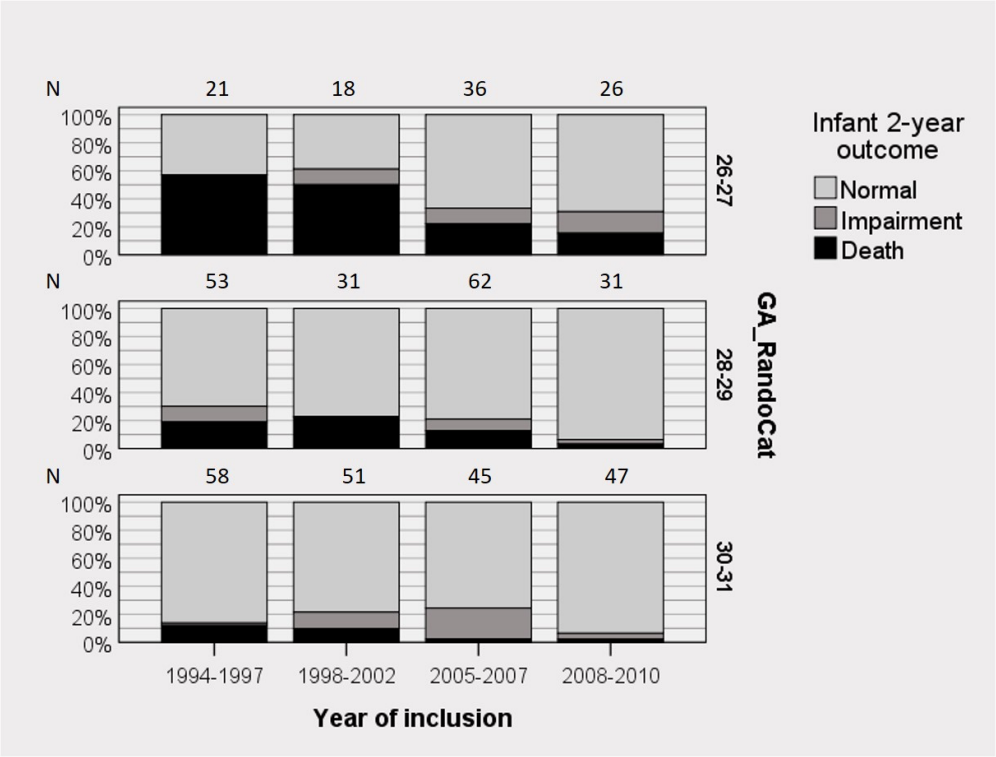
‡ Distribution differs across the categories (Kruskal-Wallis test or Pearson Chi-Square, $p < 0.05$)
& Late deaths = death after first discharge home, and before the age of two years
* Percentage of evaluated infants

† Severe impairment of hearing or vision, or cerebral palsy
# Percentage of all infants with known outcome, including perinatal death



Perinatal death rate and neurodevelopmental impairment at 2 years of age specified for group classification and gestational age at randomisation of women participating in GRIT and TRUFFLE.

206x156mm (150 x 150 DPI)



Perinatal death rate and neurodevelopmental impairment at 2 years of age specified for period and gestational age at randomisation of women participating in GRIT and TRUFFLE.

206x156mm (150 x 150 DPI)

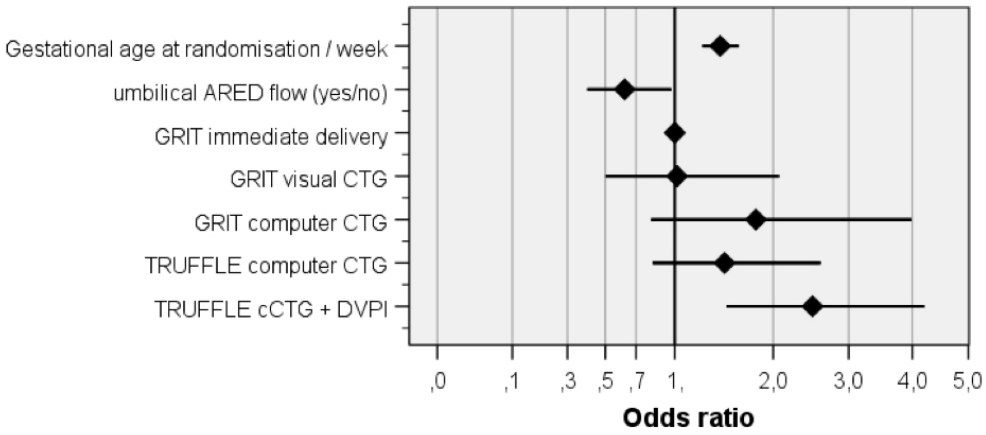


Figure 3A: Odds ratios with 95% confidence interval for survival without neurological impairment at 2 years, calculated by multivariable regression analysis, with monitoring group classification: AUC 0.69 (95% CI 0.64 to 0.74)

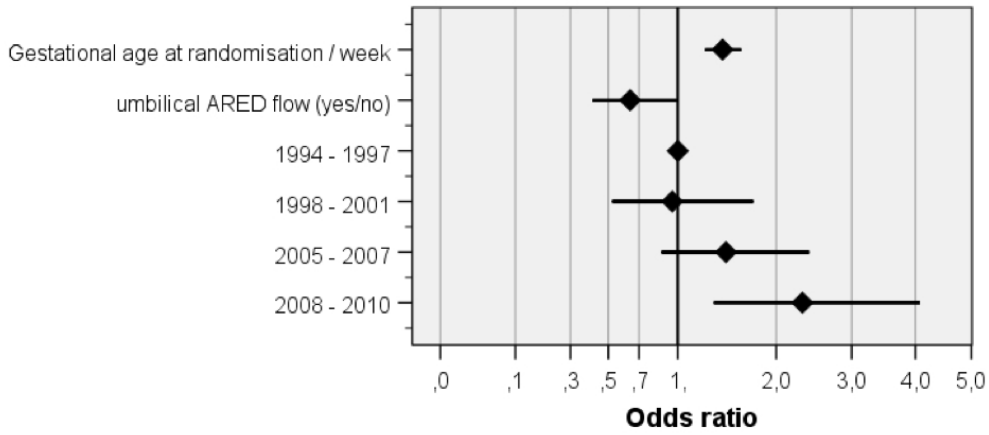


Figure 3B: Odds ratios with 95% confidence interval for survival without neurological impairment at 2 years, calculated by multivariable regression analysis, with year of inclusion: AUC 0.68 (95% CI 0.63 to 0.73)