Building consensus: thresholds for delivery in the TRUFFLE 2 randomized intervention study

B. Mylrea-Foley[1], A. Bhide[2], E. Mullins[1], J. Thornton[3], N. Marlow[4], T. Stampalija[5], R. Napolitano[6], C. Lees[1,7]

[1]Imperial College School of Medicine, Imperial College London, London, UK;
[2]Fetal Medicine Unit, St George's University Hospitals NHS Foundation Trust and Molecular & Clinical Sciences Research Institute, St George's University, London, UK;
[3]School of Clinical Sciences, University of Nottingham, Division of Obstetrics and Gynaecology, Maternity Department, City Hospital, Nottingham, UK;
[4]Academic Neonatology, UCL Elizabeth Garrett Anderson Institute for Women's Health, University College London, London, UK;
[5]Unit of Fetal Medicine and Prenatal Diagnosis, Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy;
[6]Fetal Medicine Unit, University College London Hospitals NHS Foundation Trust, London, UK;
[7]Department of Fetal Medicine, Queen Charlotte's and Chelsea Hospital, Imperial College NHS trust, London, UK

Correspondence to:

Christoph Lees (e-mail: christoph.lees@nhs.net)

Fetal cerebral Doppler changes are associated with adverse neonatal outcome in fetal growth restriction (FGR)[1,2], however it remains unclear as to which Doppler thresholds should trigger delivery. As preparation for a randomised intervention trial in late FGR we describe how we identified Doppler thresholds for cerebral blood flow redistribution (the ration of umbilical and cerebral artery pulsatility index) at which European experts in FGR were willing to randomise women for immediate delivery or expectant management.

We have identified no publications specifically outlining the most robust methods for reaching consensus regarding randomisation thresholds for randomised control trials (RCT). Various methods exist[3] including but not limited to: the Delphi technique[4], the Nominal Group Technique (NGT)[5] and the Consensus Development Conference[6]. In many instances the technique is adapted to fit the needs or composition of the group[7]. The Delphi technique, originally developed in the 1950s, is a structured approach to reaching consensus within a group of experts involving multiple rounds of questionnaires[8]. It can take many months to complete the process which relies solely on opinion without cross referencing evidence. The NGT is a structured meeting in which individuals independently list down their ideas and share these within a group, followed by in-depth discussion and finally a private vote[3].

None of these methodologies were suitable for reaching a consensus amongst our group: we presented preliminary data from the Trial of Umbilical and Fetal Flow in Europe (TRUFFLE) 2 feasibility study hence suggestions were based on preliminary evidence rather than opinion, and neither technique was practical given the large number of TRUFFLE investigators. Our process of reaching consensus was divided into three stages:

Initial discussion of thresholds: Specific cerebral and umbilical Doppler PI ranges were first discussed at a study meeting in September 2018, Turin, at which 47 investigators were present. It became apparent that participants would only be willing to randomise (and potentially deliver) at an earlier gestational age for more extreme Doppler values, hence gestation related graduated thresholds were required.

Analysis of feasibility study: The data from a prospective feasibility study were presented at the Turin meeting[9]. A variety of thresholds derived from previous discussions were considered. For each, we compared the composite adverse outcome incidence between those women who had an abnormal Doppler at least once with those that never had an abnormal Doppler. The Doppler parameter threshold that showed the highest relative risk (RR) for an abnormal composite outcome was suggested as the most appropriate candidate to be used in a future randomised study.

Consensus meeting: We carried out an exercise with participants from a group of 52 European investigators at a face-to-face meeting in March 2019 in Leuven, Belgium. After presentation of the analysis of the TRUFFLE 2 feasibility data, we used an iterative structured survey followed by voting to determine which umbilical and cerebral Doppler PI thresholds were acceptable for use in a prospective intervention study. The participants voted if they were willing to randomise or not at different Doppler PI values in combination with different gestational age epochs between 32^{+0} and 36^{+6} weeks.

Voting took place using a proprietary web based application accessible to all attendees on smart devices (https://www.mentimeter.com/). Using a unique survey code, participants were able to view each question as it was projected on the screen, and anonymously and independently cast

their vote. Choices were made after presentation of the observational study. Doppler scenarios were presented as umbilical artery pulsatility index (PI) and middle cerebral artery (MCA) PI. The derived umbilical: middle cerebral artery ratio (UCR) was not displayed, thereby encouraging decisions to be based on Doppler parameters that were familiar to the investigators.

Fifty-two participants voted on each of the scenarios. For each question, voting was left open until over 85% of participants had cast their vote (45 votes or more). The proportion of participants who voted on each scenario ranged from 88% to 96% (46/52 to 50/52). Eleven scenarios with varying Doppler thresholds were presented at gestational ages ranging from 32 to 36^{+6} weeks. For all but one scenario there was over seventy percent agreement on willingness to randomise women with example Doppler measurements at different gestations (Figure 1). Figure 2 shows a sample scenario.

The response to Question 2 (Figure 2), seen in Figure 3 as a blue/red divided circle, demonstrates that experts were not willing to use a single UCR threshold for all gestations. The small majority of participants voted for "no" in Question 2, with a UCR of above 0.8, ruling out the use of 0.8 as an absolute threshold for delivery for all gestations. An adjustment of UCR was required at 34 weeks to ensure there was consensus regarding willingness to randomise patients to either immediate or delayed delivery. For all scenarios from 34 weeks of gestation onwards, there was consensus regarding the willingness to randomise patients with Doppler results corresponding with an UCR above 0.8. For scenarios below 34 weeks there was willingness to randomise only with a more extreme UCR above 1.0.

Hence we show how knowledge from an observational study can be used to inform a consensus process for thresholds for randomization in an RCT. The TRUFFLE 2 investigators demonstrated

a spontaneous method of reaching consensus which included aspects of the Nominal Group Technique and Consensus Development Conference, where participants were given suggestions from analysis of the feasibility study, an in depth discussion was facilitated and an anonymous vote was held. We believe this to have been a practical and robust method for defining randomisation thresholds. The strengths of this approach are a multimodal approach to reaching consensus and deriving Doppler thresholds based on evidence, not solely on expert opinion. Furthermore, by making these deliberations public we aim to disseminate the results of this consensus building process.

A potential weakness is that the consensus method was developed prospectively, hence not clearly defined in advance[3]. Although not specified a priori, 70% agreement has generally been defined as achieving consensus in published literature[4,10,11]. Furthermore, the process was carried out by and on European experts in FGR, the TRUFFLE 2 Study Group. It might therefore be argued that the experts were 'self-defined'. There is discussion over who is regarded an expert in the Delphi process. Fink et al. suggested that 'An expert should be a representative of their professional group, with either sufficient expertise not to be disputed or the power required to instigate the findings'[12]. The group voting on Doppler thresholds comprised independent Fetal Medicine specialists with a particular interest in FGR and its management attending from centres across 11 European countries.

Final randomisation thresholds chosen for the TRUFFLE 2 RCT are a UCR above 1.0 from 32^{+0} to 33^{+6} weeks and a UCR above 0.8 from 34^{+0} to 36^{+6} weeks. The confirmation at the end of this process that all participants were satisfied with the result not only confirms consensus, but also strengthens the legitimacy of the protocol for the planned multicentre interventional trial. By

involving expert investigators in the process of defining randomisation thresholds, hence giving all ownership of the protocol, we hope to improve recruitment and adherence to the protocol for the future randomized study. Simultaneously, it reassures participants and those external to the study that the choices made for Doppler thresholds are based both on the best available evidence to date and on those that expert clinicians would be willing to make in their routine clinical practice.

References

1.   Meher S, Hernandez-Andrade E, Basheer SN, Lees C. Impact of cerebral redistribution on neurodevelopmental outcome in small-for-gestational-age or growth-restricted babies: A systematic review. Ultrasound Obstet Gynecol. 2015 Oct;46(4):398–404.

2.   Stampalija T, Arabin B, Wolf H, Bilardo CM, Lees C, Brezinka C, Derks JB, Diemert A, Duvekot JJ, Ferrazzi E, Frusca T. Is middle cerebral artery Doppler related to neonatal and 2-year infant outcome in early fetal growth restriction?. American journal of obstetrics and gynecology. 2017 May 1;216(5):521-e1.

3.   Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: Characteristics and guidelines for use. Am J Public Health. 1984;74(9):979–83.

4.   Hsu CC, Sandford BA. The Delphi technique: Making sense of consensus. Pract Assessment, Res Eval. 2007;12(10):1–8.

5.   Harvey N, Holmes CA. Nominal group technique: An effective method for obtaining group consensus. Int J Nurs Pract. 2012 Apr;18(2):188–94.

6.   Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, Marteau T. Consensus development methods: A review of best practice in creating clinical guidelines. J Heal Serv Res Policy. 1999;4(4):236–48.

7.   Schneider P, Evaniew N, Rendon JS, McKay P, Randall RL, Turcotte R, Vélez R, Bhandari M, Ghert M. Moving forward through consensus: Protocol for a modified Delphi approach to determine the top research priorities in the field of orthopaedic oncology. Vol. 6, BMJ Open. 2016. p. e011780.

8. Delbecq A, H. Van de Ven A, Gustafson DH. Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes. Vol. 1, Group & Organization Studies. SAGE Publications; 1975. 180 p.

9. Stampalija T and TRUFFLE 2 feasibility authors. Fetal cerebral Doppler changes and outcome in late preterm fetal growth restriction: a prospective observational cohort study. UOG. 2020 (in review).

10. Slade SC, Dionne CE, Underwood M, Buchbinder R. Standardised method for reporting exercise programmes: Protocol for a modified Delphi study. BMJ Open. 2014;4(12).

11. Kleynen M, Braun SM, Bleijlevens MH, Lexis MA, Rasquin SM, Halfens J, Wilson MR, Beurskens AJ, Masters RSW. Using a Delphi Technique to Seek Consensus Regarding Definitions, Descriptions and Classification of Terms Related to Implicit and Explicit Forms of Motor Learning. de Lussanet MHE, editor. PLoS One. 2014 Jun 26;9(6):e100227.

12. Fink A. The Survey Handbook - The Survey Kit. 2nd ed. Laughton D, Novak V, Selhorst J, Foster D, editors. Thousand Oaks: SAGE; 2003. 184 p.

Figure legends

Figure 1: Results of randomisation threshold vote held at Leuven TRUFFLE meeting in March 2019. (Blue = "yes" willing to randomise. Red = "no" not willing to randomise.)

Figure 2: The randomisation threshold vote held in Leuven in March 2019. Question 2 asked whether participants would be willing to randomise a patient at 32 weeks with an Umbilical PI of 1.5 and Middle Cerebral Artery PI of 1.6. The response to this question did not reach a consensus. With a corresponding UCR of 0.94 at 32 weeks, the votes were divided 55% "no" (26/47) and 45% "yes" (21/47).

Figure 3: Results of the randomisation threshold vote shown as UCR and gestation. Blue closed circles = majority "yes" willing to randomise. Red closed circles = majority "no" not willing to randomise. Blue/Red divided circle = "divided" no majority agreement reached.

Figure 1: Results of randomisation threshold vote held at Leuven TRUFFLE meeting in March 2019.

(Blue = "yes" willing to randomise. Red = "no" not willing to randomise.)
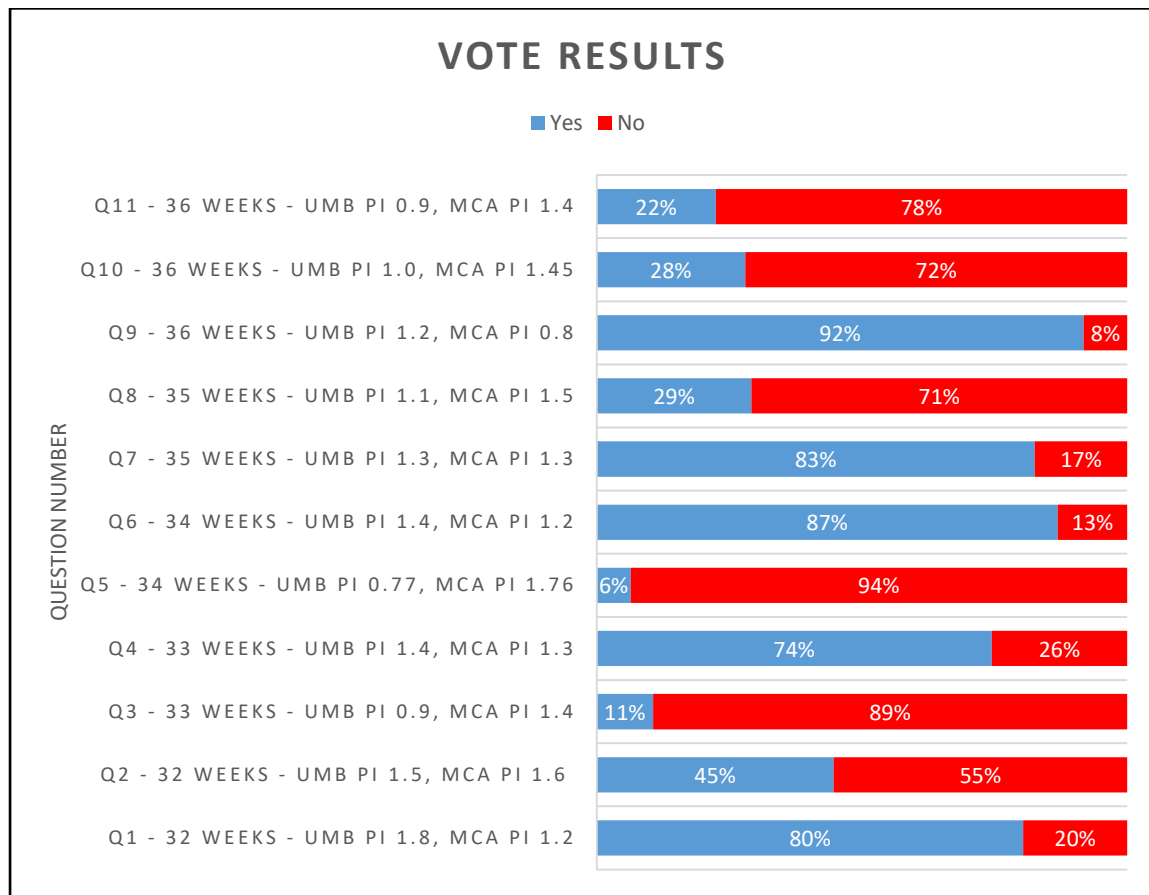
Figure 2: The randomisation threshold vote held in Leuven in March 2019. Question 2 asked whether participants would be willing to randomise a patient at 32 weeks with an Umbilical PI of 1.5 and Middle Cerebral Artery PI of 1.6. The response to this question did not reach a consensus. With a corresponding UCR of 0.94 at 32 weeks, the votes were divided 55% "no" (26/47) and 45% "yes" (21/47).
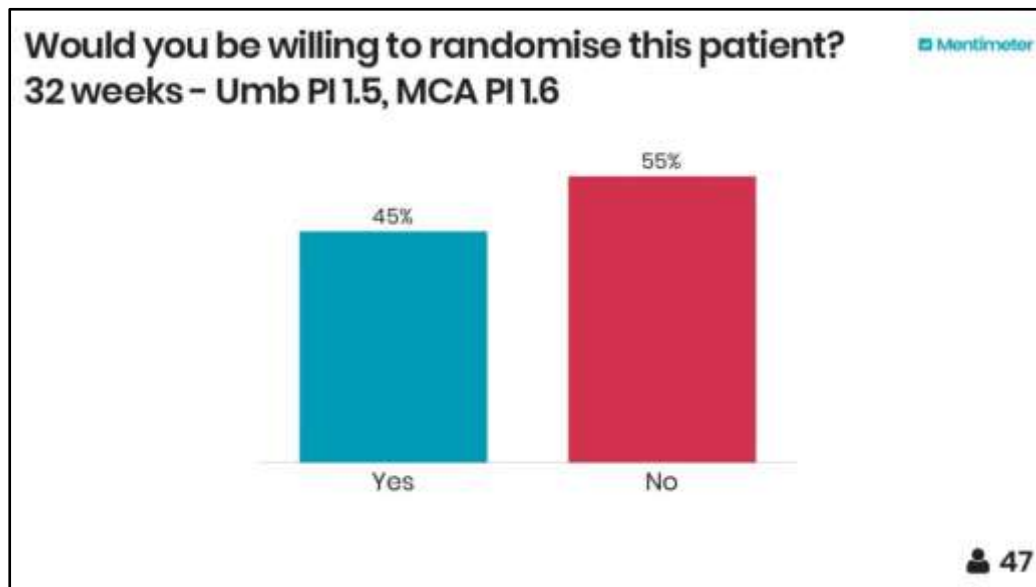
Figure 3: Results of the randomisation threshold vote shown as UCR and gestation. Blue closed circles = majority "yes" willing to randomise. Red closed circles = majority "no" not willing to randomise. Blue/Red divided circle = "divided" no majority agreement reached.