

Randomised controlled trial of cardiotocography versus Doppler auscultation of fetal heart at admission in labour in low risk obstetric population

Gary Mires, Fiona Williams, Peter Howie

Abstract

Objective To compare the effect of admission cardiotocography and Doppler auscultation of the fetal heart on neonatal outcome and levels of obstetric intervention in a low risk obstetric population.

Design Randomised controlled trial.

Setting Obstetric unit of teaching hospital

Participants Pregnant women who had no obstetric complications that warranted continuous monitoring of fetal heart rate in labour.

Intervention Women were randomised to receive either cardiotocography or Doppler auscultation of the fetal heart when they were admitted in spontaneous uncomplicated labour.

Main outcome measures The primary outcome measure was umbilical arterial metabolic acidosis. Secondary outcome measures included other measures of condition at birth and obstetric intervention.

Results There were no significant differences in the incidence of metabolic acidosis or any other measure of neonatal outcome among women who remained at low risk when they were admitted in labour. However, compared with women who received Doppler auscultation, women who had admission cardiotocography were significantly more likely to have continuous fetal heart rate monitoring in labour (odds ratio 1.49, 95% confidence interval 1.26 to 1.76), augmentation of labour (1.26, 1.02 to 1.56), epidural analgesia (1.33, 1.10 to 1.61), and operative delivery (1.36, 1.12 to 1.65).

Conclusions Compared with Doppler auscultation of the fetal heart, admission cardiotocography does not benefit neonatal outcome in low risk women. Its use results in increased obstetric intervention, including operative delivery.

Introduction

The admission cardiotocogram is a short, usually 20 minute, recording of the fetal heart rate immediately after admission to the labour ward.¹ The main justification for admission cardiotocography is that the uterine contractions of labour put stress on the placental circulation; an abnormal tracing might indicate a deficiency and hence identify potential fetal compromise at an early enough stage to allow intervention. Furthermore, a normal admission cardiotocogram offers reassurance. However, the incidence of intrapartum fetal compromise is low in pregnancies that have been uncomplicated before the onset of labour. Thus, labour admission cardiotocography may represent unnecessary intervention. In such low risk cases, confirmation of a normal fetal heart rate by Doppler auscultation should be sufficient.²

Evidence from randomised trials shows that routine electronic fetal monitoring throughout labour results in increased, and probably unnecessary, intervention for apparent fetal distress.³⁻⁵ Admission cardiotocography in a low risk obstetric population may therefore result in increased obstetric intervention without fetal and neonatal benefit. We compared the effects of labour admission cardiotocography and Doppler auscultation of the fetal heart on neonatal outcome and levels of obstetric intervention in a low risk obstetric population.

Participants and methods

Women were eligible to join the study if they were booked for hospital delivery, attended a hospital or community based consultant led clinic in the third trimester of pregnancy, and had no obstetric complications at that visit that would warrant continuous intrapartum monitoring of fetal heart rate (pre-eclampsia or hypertension in previous or index pregnancy; essential hypertension; diabetes (insulin dependent or gestational); suspected intrauterine growth restriction; placental abruption or praevia or vaginal bleeding of unknown origin; multiple pregnancy; fetal malformation; previous caesarean section; breech presentation; or rhesus isoimmunisation).

Randomisation procedure and study protocol

A researcher obtained informed consent for the study at the third trimester visit. Women were randomised to the cardiotocography or Doppler group with a commercially available computer randomisation program.⁶ The allocation was placed in a sealed envelope and attached to the labour admission page of the woman's case records. The women did not know which group they had been randomised to until their admission in labour. An independent observer checked the randomisation process weekly. The data analysts were blind to the randomisation code.

When women who were recruited to the study were admitted in spontaneous uncomplicated labour, the admitting midwife opened the trial envelope. Women randomised to the cardiotocography group had 20 minutes of cardiotocography before being given any opiate analgesia. Women randomised to the Doppler group had fetal heart auscultation with a hand held Doppler device during and immediately after at least one contraction. Staff on the labour ward interpreted the fetal heart assessments and took action on the basis of protocols.

Some women developed complications after randomisation that warranted continuous intrapartum monitoring of the fetal heart rate. These women had their admission assessment and labour managed according to the obstetric complication.

At delivery, we obtained umbilical cord blood samples to measure pH and base deficit. When possible

Editorial by
Goddard

Department of
Obstetrics and
Gynaecology,
Ninewells Hospital
and Medical School,
Dundee, Tayside
DD1 9SY

Gary Mires
senior lecturer
Peter Howie
professor

Department of
Epidemiology and
Public Health,
Ninewells Hospital
and Medical School
Fiona Williams
lecturer

Correspondence to:
G Mires
g.j.mires@dundee.
ac.uk

BMJ 2001;322:1457-62

both arterial and venous samples were taken. Arterial blood results were included in the analysis only if both an arterial and venous sample had been obtained and the results suggested no errors due to a poor quality sample.

Outcome measures

The primary outcome measure was metabolic acidosis at delivery, defined as an umbilical cord pH < 7.20 with a base deficit of > 8.0 mmol/l. Secondary outcome measures were other assessments of neonatal outcome (Apgar scores, need for intermittent positive pressure ventilation at resuscitation, admission to neonatal intensive care) and obstetric intervention (use of continuous fetal heart rate monitoring in labour, artificial rupture of membranes, augmentation of labour, monitoring of scalp pH, epidural analgesia, operative delivery).

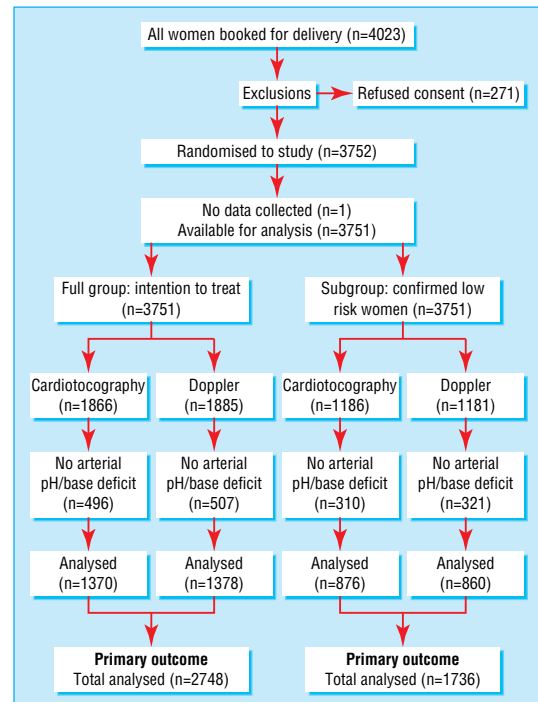
Sample size and analysis

A final target sample size of 1704 confirmed low risk women was based on an excess of umbilical cord blood metabolic acidosis of 4% in the Doppler group. This gave an α of 0.05 with 80% power. This effect size was chosen as a clinically important difference. In a pilot study in which all women had admission cardiocotography, the incidence of metabolic acidosis was 7%.

During the study the sample size and power calculation were reviewed by the project group and modified. Our original grant application was awarded on the basis of recruiting 3370 low risk women and based on an arbitrary 3% excess of metabolic acidosis in the Doppler group, which we considered clinically important and defined the target power of 90%. As the study progressed, it became clear that recruitment was lower than expected for various justifiable but unalterable reasons and that more women were developing complications that required continuous monitoring than had been predicted. These problems were highlighted in the interim report to the funding body, and we were given a supplementary award to extend the recruitment period. In this application, we stated that the power would be reduced to 80% as this was the conventional level. At that time we believed that we would be able to achieve the reduced sample size of 2552.

Just after receiving the supplementary award, we audited the interobserver and intraobserver error in abstracting data from clinical notes. This showed that data on arterial blood gas concentrations were not available in all women, which would create a further shortfall in the number of women available for analysis of the primary outcome variable. Thirteen months before recruitment was due to end the steering group met with two research midwives to discuss the recruitment and viability of the study. At this stage we calculated that at the end of recruitment the number of low risk women on admission in labour with data on arterial blood gas concentrations would be about 1800, a shortfall of 752.

We discussed whether to ask for further funding, review the power calculations, or abandon the study. We dismissed asking for further funding as unrealistic and then reviewed the power calculation. We were not prepared to drop the target power of study below 80%. However, we discussed at length the relevance of the clinical effect size of 3%. Our predicted sample size was sufficient to detect a 4% difference, and our clinical decision at the time was, and remains, that a 4% differ-



Flow chart of trial

ence was as relevant as a 3% difference. We therefore agreed to set the effect size at 4% with an 80% power.

Apart from the small audit, no analysis was done until the study was complete and all data had been entered and double checked. All analysis was done blind to the randomisation group.

The whole group analysis was by intention to treat at the time of randomisation. We also did a subgroup analysis on women who were still low risk when they were admitted in labour.

Results

In all, 4023 women met the entry requirements; 271 (7%) did not wish to participate, which left 3752 to be randomised in the third trimester (figure).

Whole group analysis

Comparison between the two groups showed no significant differences in the incidence of metabolic acidosis at delivery. Women in the cardiotocography group were significantly more likely than women in the Doppler group to have continuous monitoring of fetal heart rate during labour, have epidural analgesia, and require an operative delivery. There were no other significant differences (table 1).

Subgroup analysis

Between randomisation during the third trimester of pregnancy and admission in labour, 1384 women (37%) developed an obstetric complication that warranted continuous fetal heart rate monitoring in labour (table 2). We did a subgroup analysis after these women were excluded.

Comparison between groups showed no significant differences in the incidence of metabolic acidosis at delivery. Women who had admission cardiotocography were significantly more likely than women in the Doppler group to have continuous fetal heart rate

monitoring in labour, augmentation of labour, epidural analgesia, and require an operative delivery. There were no other significant differences (table 3).

In this group of confirmed low risk women, 21.5% (255/1186) randomised to have admission cardiotocography were considered to have an abnormal fetal heart trace at the onset of labour compared with 3.6% (42/1181) of women in the Doppler group ($P < 0.0001$). In the cardiotocography group the commonest abnormalities were decelerations (147, 58%) and reduced variability (111, 43%) of cases. In the Doppler group, the commonest abnormalities were a bradycardia and decelerations (both 18 (43%) cases).

Discussion

In our clinical environment, admission cardiotocography had no neonatal benefit, as assessed by metabolic acidosis at delivery, but resulted in increased obstetric intervention. We obtained the same result whether the analysis was for the whole group or just women who remained at low risk when admitted in labour. As the aim of the study was to investigate the effect of admission cardiotocography or Doppler auscultation on the incidence of metabolic acidosis in low risk pregnancies, we felt justified in performing a subgroup analysis in which we excluded women who developed complications between randomisation and the onset of labour. It would have been preferable to obtain women's consent before labour and then wait until they were admitted in labour before randomising them. We could not do this because of ethical and logistical constraints.

Effect on fetus

Previous descriptive studies have suggested that admission cardiotocography may help identify a compromised fetus when the uterine contractions of early labour act as a functional stress on the placental circulation.^{1-7,8} These uncontrolled studies, however, do not allow conclusions to be drawn about the clinical usefulness or indeed clinical risks of admission cardiotocography. The assumed benefits are not confirmed by our trial.

We found no neonatal benefit from women having admission cardiotocography, as assessed by the presence of metabolic acidosis at delivery. There were no significant differences in the secondary neonatal outcome measures between the two methods of fetal heart rate assessment. However, our sample size is insufficient to provide definitive information about these outcomes. Many of the confidence intervals were wide and included unity.

Interpreting cardiotocograms

We found much higher levels of concern about the fetal heart rate after admission cardiotocography than after Doppler auscultation. There is wide intraobserver and interobserver variation in the interpretation of cardiotocograms even among experts.⁹⁻¹⁵ Fetal heart variability is difficult to interpret visually, and there is a tendency to over report abnormalities.¹²⁻¹⁴⁻¹⁷ We found a high percentage of admission cardiotocograms were reported as abnormal, with reduced variability and variable decelerations the most commonly reported abnormalities. This high rate of abnormal admission traces is in keeping with findings in other studies.¹⁸

Table 1 Neonatal and obstetric outcomes in whole group analysis after cardiotocography or Doppler auscultation at admission

Outcome	No in cardiotocography group	No in Doppler group	Odds ratio (95% CI)
Cord arterial blood metabolic acidosis	252/1370	262/1378	0.96 (0.79 to 1.17)
Apgar score <7 at 5 min	36/1858	34/1868	1.07 (0.65 to 1.75)
Need for IPPV at resuscitation	11/1865	14/1878	0.79 (0.33 to 1.85)
Admission to neonatal intensive care	89/1864	105/1878	0.85 (0.63 to 1.14)
Hypoxic ischaemic encephalopathy	8/81	15/99	0.61 (0.22 to 1.65)
Continuous fetal heart rate monitoring in labour	1246/1865	1128/1882	1.35 (1.17 to 1.54)
Artificial rupture of membranes	1065/1864	1031/1879	1.10 (0.96 to 1.25)
Augmentation of labour	714/1862	692/1878	1.07 (0.93 to 1.22)
Monitoring of fetal scalp pH	197/1866	177/1885	1.14 (0.91 to 1.42)
Epidural analgesia in labour	617/1866	565/1885	1.15 (1.00 to 1.33)
Operative delivery*	602/1866	551/1885	1.15 (1.00 to 1.32)
Caesarean section	193/1866	165/1885	1.20 (0.96 to 1.50)

IPPV=intermittent positive pressure ventilation.

*Caesarean section, forceps, or ventouse.

Table 2 Reasons for exclusion after randomisation (some women had more than one reason for exclusion)

Complication	No (%) of cases
Antepartum haemorrhage	159 (4.2)
Raised blood pressure	271 (7.2)
Suspected small for dates	56 (1.5)
Preterm labour	48 (1.3)
Gestational diabetes	2 (0.1)
Fetal anomaly	2 (0.1)
Reduced fetal movements and suspected fetal compromise	63 (1.7)
Meconium stained liquor	99 (2.6)
Intrauterine death	3 (0.1)
Persistent breech	67 (1.8)
Membranes ruptured before labour	164 (4.4)
Induction of labour	833 (22.2)
Baby born before arrival at hospital	19 (0.5)
Elective caesarean section	61 (1.6)
Woman withdrew from trial	31 (0.8)
Other	44 (1.2)
Total	1384 (36.9)

Table 3 Comparison of neonatal and obstetric outcomes in women who remained low risk at labour

Outcome	No in cardiotocography group	No in Doppler group	Odds ratio (95% CI)
Cord arterial blood metabolic acidosis	159/876	154/860	1.02 (0.79 to 1.31)
Apgar score <7 at 5 minutes	25/1181	18/1171	1.39 (0.72 to 2.66)
Need for IPPV at resuscitation	5/1185	4/1178	1.24 (0.29 to 5.51)
Admission to neonatal intensive care	46/1185	45/1175	1.01 (0.65 to 1.57)
Hypoxic ischaemic encephalopathy	6/42	5/43	1.27 (0.31 to 5.34)
Continuous fetal heart rate monitoring in labour	672/1186	551/1178	1.49 (1.26 to 1.76)
Artificial rupture of membranes	640/1185	614/1175	1.07 (0.91 to 1.27)
Augmentation of labour	246/1183	202/1175	1.26 (1.02 to 1.56)
Use of fetal scalp pH	96/1186	76/1181	1.28 (0.93 to 1.77)
Epidural analgesia in labour	325/1186	261/1181	1.33 (1.10 to 1.61)
Operative delivery*	313/1186	247/1181	1.36 (1.12 to 1.65)
Caesarean section	61/1186	43/1181	1.43 (0.95 to 2.18)

IPPV=intermittent positive pressure ventilation.

*Caesarean section, forceps, or ventouse.

Variability in fetal heart rate cannot be assessed in the Doppler group.

Our study was a pragmatic trial in which the midwives and junior obstetric staff interpreted the admission fetal heart assessments. Our findings suggest that when an admission cardiotocogram shows

What is already known on this topic

The admission cardiotocogram is a short recording of the fetal heart rate immediately after admission to the labour ward

Opinion varies about its value in identifying a potentially compromised fetus

In low risk women, the incidence of intrapartum fetal compromise is low

What this study adds

Compared with Doppler auscultation of the fetal heart, admission cardiotocography has no benefit on neonatal outcome in low risk women

Admission cardiotocography results in increased obstetric intervention, including operative delivery

possible reduced baseline variability or mild decelerations, midwives and obstetricians will take defensive action. This starts with continuous monitoring of fetal heart rate, which leads to increased obstetric intervention in the form of augmentation of labour, epidural analgesia, and, ultimately, increased rates of operative delivery. Other randomised trials have also found that routine electronic fetal monitoring in labour results in increased unnecessary intervention for fetal distress.³⁻⁵

Maternal outcomes

Perhaps the most important finding is the increased rate of operative delivery in women who had admission cardiotocography. Among women who were low risk at admission, there was an absolute increase of 5.5% in operative delivery and 1.5% increase in caesarean sections. The rising caesarean section rate in the United Kingdom continues to generate much debate and concern.¹⁹⁻²³ The increased use of continuous monitoring of fetal heart rate in labour in women who had admission cardiotocography in this study is likely to be a contributing factor.

This study has confirmed that among women with low risk features at the onset of labour, the admission cardiotocogram is no better than Doppler auscultation of the fetal heart in identifying a potentially compromised fetus. Admission cardiotocography was associated with increased obstetric intervention including higher rates of operative delivery. Although caution is needed in generalising conclusions to the whole population, our results point to potential problems with admission cardiotocography. These problems are

likely to persist while difficulties remain in interpreting cardiotocograms.

We thank the research midwives Maureen McLeod and Suzanne Lucas for recruiting women and collecting data.

Contributors: GM and PH had the idea for the study. GM and FW coordinated the study and analysed the data. The paper was written jointly between all three authors. GM will act as guarantor.

Funding: Chief Scientists Office of the Scottish Executive, Edinburgh.

Competing interests: None declared

- 1 Ingemarsson I. Electronic fetal monitoring as a screening test. In: Spencer JAD, Ward RHT, ed. *Intrapartum fetal surveillance*. London: Royal College of Obstetricians and Gynaecologists, 1993:45-52.
- 2 Prentice A, Lind T. Fetal heart rate monitoring in labour—too frequent intervention, too little benefit? *Lancet* 1997;ii:1375-7.
- 3 Grant AM. Electronic fetal monitoring alone versus intermittent auscultation in labour. In: Enkin MW, Kierse MJNC, Renfrew MJ, Neilson JP, eds. *Cochrane pregnancy and childbirth database*. Oxford: Update Software, 1993.
- 4 Thacker SB, Stroup DF, Peterson HB. Efficacy and safety of intrapartum electronic fetal monitoring: an update. *Obstet Gynecol* 1995;86:613-20.
- 5 Thacker SB, Stroup DF. Continuous electronic heart rate monitoring for fetal assessment during labor *Cochrane Database Syst Rev* 2000;(2):CD000063.
- 6 Florey C du V. *Randomiser*. Dundee: University of Dundee, 1995.
- 7 Pello LC, Dawes GS, Smith J, Redman CW. Screening of the fetal heart in early labour. *Br J Obstet Gynecol* 1988;95:1128-36.
- 8 Phelan JP. Labor admission test. *Clin Perinatol* 1994;21:879-85.
- 9 Helland M, Marton K, Ueland K. Factors involved in the interpretation of fetal monitor tracings. *Am J Obstet Gynecol* 1981;151:737-42.
- 10 Beaulieu MD, Fabia J, Leduc B, Brisson J, Bastide A, Bloudin D, et al. The reproducibility of intrapartum cardiotocogram assessments. *Can Med J* 1982;127:214-6.
- 11 Nielson PV, Stigsby B, Nickelsoen C, Nim J. Intra and inter observer variability in the assessments of intrapartum cardiotocograms. *Acta Obstet Gynecol Scand* 1987;66:421-4.
- 12 Borgotta L, Shrouf PE, Divon MY. Reliability and reproducibility of non stress test readings. *Am J Obstet Gynecol* 1988;159:554-8.
- 13 Lidagaard O, Bottcher LM, Weber T. Description, evaluation and clinical decision making according to various fetal heart rate patterns—interobserver and regional variability. *Acta Obstet Gynecol Scand* 1992;71:48-53.
- 14 Flynn AM, Kelly J, Matthews R. Predictive value of, and observer variability in, several ways of reporting antenatal cardiotocograms. *Br J Obstet Gynecol* 1982;89:434-40.
- 15 Lotgering FK, Wallenberg HCS, Schouten HJA. Interobserver and intraobserver variation in the assessment of antenatal cardiotocograms. *Am J Obstet Gynecol* 1982;144:701-5.
- 16 Trimpos JB, Keirse MJNC. Observer variability in the assessment of antenatal cardiotocograms. *Br J Obstet Gynecol* 1978;85:900-6.
- 17 Dawes GS, Lobb M, Moulden M, Redman CWG, Wheeler T. Antenatal cardiotocograms and interpretation using computers. *Br J Obstet Gynecol* 1992;99:791-7.
- 18 Samo AP, Phelan JP, Ahn MO. Relationship of early intrapartum fetal heart rate patterns to subsequent patterns and fetal outcome. *J Reprod Med* 1990;35:239-42.
- 19 McIlwaine GM, Cole SK, Macnaughton MC. The rising caesarean section rate—a matter of concern. *Health Bull* 1985;43:301-5.
- 20 Macfarlane A, Chamberlain G. What is happening to caesarean section rates? *Lancet* 1993;342:1005-6.
- 21 Chamberlain G. What is the correct caesarean section rate? *Br J Obstet Gynecol* 1993;100:403-4.
- 22 Treffers PE, Pel M. The rising trend in caesarean birth. *BMJ* 1993;307:1017-8.
- 23 Wilkinson C, McIlwaine G, Boulton-Jones C, Cole S. Is the rising caesarean section rate inevitable? *Br J Obstet Gynecol* 1998;105:45-52.

(Accepted 19 February 2001)

Commentary: changes between protocol and manuscript should be declared at submission

Sandy Goldbeck-Wood

BMJ, London
WC1H 9JR
Sandy
Goldbeck-Wood
assistant editor
sgoldbeck-wood@
bmj.com

Mires et al's article addresses a wide readership on an important topic and uses robust methods. It has the potential to change clinical practice, which is one of the yardsticks by which journals measure the influence of papers they publish. It is therefore just the kind of article we are keen to publish in the *BMJ*.

We were worried, therefore, to discover by chance that the power calculation, the pilot incidence, and the

degree of clinically important difference declared in the submitted manuscript differed from those declared in the original study protocol. We then faced a question of publication ethics: should we continue with publishing a paper we believed would interest *BMJ* readers, despite irregularities in its presentation, or should we reject it because of poor publication practice?

The key question seemed to be whether the paper's scientific validity, given the departure from the prespecified power calculation, was now too attenuated to deliver a clear and valid message for general readers. With statistical advice, we concluded that the study, although no longer able unequivocally to exclude a difference between the cardiocography and intermittent Doppler auscultation arms, had sufficient power to make a new and useful contribution to the debate over monitoring in delivery units. We therefore chose to publish it. In doing so, we hope to open a debate on the ethical issues it raises.

Need for openness

What view should we take of Mires et al's decisions to conduct an interim audit and modify their targets accordingly? Murray, a statistician, believes that is wrong to modify power calculations. The fact that the choice of a clinically relevant difference is arbitrary, he argues, is all the more reason for choosing it in a prespecified, rather than data driven, way. Nesheim, on the other hand, an obstetrician and trialist with experience in ethics, argues for greater leniency when hindsight reveals inaccurate baseline

assumptions about recruitment rates or rates of outcomes.

Whatever view statisticians and trialists take of "evolving" power calculations, however, the editor's view on submission of manuscripts is clear. Good publication ethics require that all important changes that occur between the protocol and submitted manuscript must be declared at the time of submission.

How would others deal with the problem Mires et al faced? Only hindsight allows confident "prediction" of the likely rates of outcomes or of numbers of participants recruited, and similar difficulties must have been faced by many researchers. How has it been managed elsewhere? How ought it to be managed, and how far does the continuing responsibility for ensuring effective use of research funds extend? Do funding bodies have a commitment to seeing an approved trial through to a scientifically meaningful conclusion when unforeseen recruitment difficulties arise? Should editors require all trialists to submit protocols, along with details of any amendments that have been made? We welcome your views. Please send a rapid response to bmj.com.

Competing interests: None declared.

Commentary: research governance must focus on research training

Gordon D Murray

Cases of research fraud are regularly reported in the *BMJ*, usually in the context of a doctor being disciplined by the General Medical Council. Such cases are inexcusable and undermine public confidence in science and the medical profession. However, I have long argued that in terms of the contamination of the medical literature, the effects of blatant fraud are modest compared with the huge number of published papers that are seriously misleading because they ignore the basics of good research practice.^{1,3} Data driven hypotheses are put forward as if they were prospective, or multiple analyses are done on accumulating data in a game of "chase the P value." The importance of prespecifying a carefully formulated question, adhering to the protocol, and interpreting the results in the light of the original question, does not seem to be widely appreciated.

The paper by Mires et al is a case in point. I was asked to referee the manuscript for the *BMJ*, and the power calculation described in the manuscript was based on having 80% power to detect at the 5% significance level a clinically relevant difference of 4% in the incidence of metabolic acidosis, assuming a background rate of 7% (derived from pilot data). By coincidence, I was also asked to referee the authors' final report to the trial's funding body. This gave me access to the original grant application, where the power calculation was based on having 90% power to detect a clinically relevant difference of 3% in the incidence of metabolic acidosis with 5% significance, assuming a background rate of 6% (derived from the same pilot data).

The final report to the funding body explained some but not all of these midcourse corrections, but they were not mentioned in the *BMJ* manuscript. This is, of course, of particular concern in an open study such as this, where there can be no robust proof that the changes were not data driven.

Effect of changes

This point might be seen as academic, but the fine detail of the power calculation is crucial in interpreting the results. The study had negative findings, with the 95% confidence interval for the change in the rate of metabolic acidosis being -2.3% to 3.5% . In the original power calculation a difference of 3% was regarded as clinically relevant, and the confidence interval does not exclude the possibility of such a difference. Thus the study is inconclusive. However, with the modified power calculation 4% was regarded as the smallest clinically relevant difference, and the confidence interval does exclude such a difference. Thus the study, which had been inconclusive and rather inadequate, now gives a strong negative finding that establishes the equivalence of the two interventions in terms of the primary outcome measure.

I would be the first to acknowledge that the value judgment of whether 3% or 4% should be regarded as the smallest clinically relevant difference is rather arbitrary, but it is precisely because such judgments are difficult that they must be discussed and agreed beforehand rather than in the light of the data. I am not suggesting that Mires et al intended to deceive. But the manuscript as originally submitted to the *BMJ* was mis-

Department of
Community Health
Sciences, University
of Edinburgh
Medical School,
Edinburgh
EH8 9AG

Gordon D Murray
head of public health
sciences

[gordon.murray@
ed.ac.uk](mailto:gordon.murray@ed.ac.uk)

leading because it did not set out the original power calculation and the reasons for the subsequent changes.

Detection and prevention

In terms of detection, this case raises the important question of whether authors should be required to submit original study protocols, and protocol amendments, along with their manuscripts. The increase in workload for reviewers would be substantial, but I believe it could be justified for important pragmatic studies that have the potential to modify clinical practice. Maybe this could be the selection criterion. Authors who believe that their results ought to affect clinical practice would be required to submit their protocol.

The issue of prevention is more important. With crucial guidelines on research governance being drafted by the Department of Health, the Scottish Executive, and the GMC, great emphasis must be placed on research training. With so much evidence that even experienced investigators do not fully appreciate the importance of “the scientific method,” there is much work left to be done.

Competing interests: The University of Edinburgh could benefit financially through running courses on research methodology.

- 1 Murray GD. The task of a statistical referee. *Br J Surg* 1988;75:664-7.
- 2 Murray GD. Statistical aspects of research methodology. *Br J Surg* 1991;78:777-81.
- 3 Murray GD. Promoting good research practice. *Stat Methods Med Res* 2000;9:17-24.

Commentary: Approach to power calculations has to be realistic

Britt-Ingjerd Nesheim

Department of
Obstetrics, Ullevål
University Hospital,
Oslo, Norway
Britt-Ingjerd
Nesheim
professor
b.i.nesheim@
ioks.uio.no

The best time to plan a controlled clinical trial is after the trial is finished. Then, you have the answer to all the questions you need to ask before starting, such as:

- What is the best way to recruit your patients?
- How to choose the exclusion criteria? Could the exclusion criteria bias the recruitment—and the results?
- What will the exclusion rate and the dropout rate be? How many will decline to participate?
- What is the incidence of the outcome measure in your control group?
- Which difference in outcome measure between the control group and the experimental group is clinically relevant?
- Which level of statistical power should you go for?

In many instances, the investigator has to make an informed guess, which may be wrong. Researchers are commonly overoptimistic about recruitment. In 1979, Lasagna commented on a trial where out of 8027 possible candidates 100 people participated.¹ This led to what is now popularly called Lasagna’s law: in any trial, the incidence of the disease studied will be reduced to 10% of the original estimate.

What are the options for the investigator when the recruitment to a study is ebbing? Funding agencies are usually not happy to put much money into a study that

turns out to be more expensive than was originally thought. Should the whole study be thrown away and forgotten? That would be a waste of time, money, and effort.

Can redoing the power calculation be defended? In an ideal and purist world, it cannot. In the real world of clinical trials, I think it can. Often, the size of the difference in outcome measures is chosen rather arbitrarily. In the optimistic planning phase, a small difference may be chosen, while a larger one could be just as clinically appropriate. The same applies to the level of statistical power: it should not matter much what the original calculation was, as long as it is stated in the paper what the power of this study is.

Traditionally, too little emphasis has been placed on methods when publishing clinical trials. The CONSORT statement should be helpful in creating new attitudes.² The transparency in reporting must also incorporate recruitment problems, and, as here, the necessity of redoing the power calculations.

Competing interests: None declared.

- 1 Lasagna L. Problems in publication of clinical trial methodology. *Clin Pharmacol Ther* 1979; 25:751-3.
- 2 Moher D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.

People make mistakes: what then?

I was eating my supper when the ward servant knocked on the door (the wires of the intercom had been nibbled by rats, and it was not working): “Nurse say, ‘Doctor please come quick.’” So I came and found a distraught nursing aide standing by a diabetic patient who had just come out of a ketoacidotic coma. “Doctor I have given him 10 CCs of insulin, and now I remember that Mr Gabriel in the school say ‘Insulin is always measured in units, not in CCs.’”

I looked at the man’s arm, and there was a large swelling where the insulin had been injected subcutaneously. Putting the drip to dextrose, I sent for sister laboratory worker, the theatre sister, and the theatre orderly. We excised the swelling under local anaesthetic, and I arranged that sister would recheck the blood sugar—normal at the time, at midnight, and 5 am.

In the morning I discussed with matron what to do. I pointed out that if the nurse had said nothing the man would have died, and quite likely no one would have known why. The nurse was to be congratulated on saving the patient. If matron sacked her or made her pay for the extra treatment the next nurse would not tell us. The patient got better, the hospital paid for the extra treatment, and everyone was happy. We told the man that a mistake had been made but that, because it was noticed in time, he had recovered. He was happy too.

We discussed it at the next class for trained staff (held weekly), and I was told that no one was likely to make that mistake again. I wonder if some fatal but treatable errors are not known about because of our reactions to mistakes?

Anne Seymour *retired mission doctor, South Shields*