Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery

Andrew Kotaska

BMJ 2004;329;1039-1042
doi:10.1136/bmj.329.7473.1039

Updated information and services can be found at:
http://bmj.com/cgi/content/full/329/7473/1039

These include:

References
This article cites 25 articles, 8 of which can be accessed free at:
http://bmj.com/cgi/content/full/329/7473/1039#BIBL

Rapid responses
8 rapid responses have been posted to this article, which you can access for free at:
http://bmj.com/cgi/content/full/329/7473/1039#responses

You can respond to this article at:
http://bmj.com/cgi/eletter-submit/329/7473/1039

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections
Articles on similar topics can be found in the following collections
• Other evidence based practice (354 articles)

Correction
A correction has been published for this article. The contents of the correction have been appended to the original article in this reprint. The correction is also available online at:
http://bmj.com/cgi/content/full/329/7479/1385

Notes

To order reprints of this article go to:
http://bmj.bmjournals.com/cgi/reprintform

To subscribe to BMJ go to:
http://bmj.bmjournals.com/subscriptions/subscribe.shtml
Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery
Andrew Kotaska

As randomised trials continue to ascend in the evolution of evidence based medicine, we must recognise and respect their limitations when examining complex phenomena in heterogeneous populations.

Randomised controlled trials have greatly improved the quality of evidence guiding clinical practice, but when applied to complex phenomena, they have important limitations. Complex patient populations with poorly quantifiable variations between individuals present one area of difficulty; complex procedures requiring skill and clinical judgment present another. A large, well designed, and well executed randomised controlled trial of breech presentation at term, the "term breech trial," by Hannah et al rapidly dictated a new standard of care for the management of breech deliveries around the world. Yet this trial failed to adequately appreciate both the complex nature of vaginal breech delivery and the complex mix of operator variables necessary for its safe conduct. Widespread acceptance of this trial's results has breached the limits of evidence based medicine.

Hannah et al's trial showed a significant increase in perinatal mortality and morbidity in women randomised to a trial of labour compared with elective caesarean section. The trial's methodological flaws have been examined, but the intrinsic limitations of applying large scale randomisation to complex phenomena have received little attention. These limitations are the focus of this paper.

Bias of licence
Many of the term breech trial's 121 centres were in North America, where 13% of breech presentations at term were delivered vaginally. The study achieved a successful vaginal delivery rate of 57% by asking those centres with vaginal birth rates under 40% in the labour group to increase the rate or withdraw from participation. Individual centres rates of vaginal breech delivery at baseline were not reported, but many would have tripled their vaginal delivery rate overnight. The vaginal delivery of a breech baby involves risk. Cord prolapse and trapped fetal parts are unpredict-

Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, V6H 3V5 Canada Andrew Kotaska skotaska@bulkley.net Correspondence to: A Kotaska kotaska@bulkley.net BMJ 2004;329:1039–42 Correspondence to: P Garner pgarner@liv.ac.uk BMJ 2004;329:1055–6
able complications. Every practitioner knows this; and the literature, the courts, and the low baseline rate of such deliveries in North America highlight caution. Maternity units with interest and skill in delivering breech babies vaginally have achieved higher rates: 24% in the United States, 36% in Sweden, 38% in Israel, 39% in Switzerland, 50% in France, and 53% in Norway. These retrospective studies carefully selected women for a trial of labour using various safety criteria and showed lower mortality and morbidity associated with vaginal breech delivery than in the term breech trial. Few obstetrical units other than Lew's have published vaginal delivery rates as high as the term breech trial.

Statistical power required a high vaginal delivery rate to enhance the trial's ability to detect small differences in outcome. With this aim, the researchers encouraged practitioners to increase their vaginal breech delivery rate beyond their previous comfort level. Despite being difficult to quantify, comfort level (or “practitioner comfort level”) plays a pivotal part in the safety of complex procedures. Protected from medicolegal liability by the equipoise of a randomised trial, some practitioners must have pushed their comfort levels with vaginal breech delivery. This constitutes a significant bias: one of licence. The trial protocol's liberal labour guidelines allowed 0.5 cm dilation/h and 3.5 hours for the second stage. This is considerable licence, and few obstetricians from centres with proved safety in vaginal breech delivery would find them acceptable.

A discriminating procedure

Human skill varies. This is particularly evident when tasks are complex, require careful judgment, and have narrow margins for error. Increasingly complex tasks are discriminating, with more effort and skill required to master them. The safe vaginal delivery of a breech baby requires considerable skill and is a discriminating obstetrical procedure. Skill is required in multiple areas: not just in the delivery technique, but in ultrasound assessment, the selection of cases, intrapartum fetal surveillance, the conduct of labour, and paediatric support. A coordinated, well functioning unit is difficult to quantify.

Other recognised discriminating procedures include carotid endarterectomy, where surgical skill is one of the strongest predictors of the operation's utility; surgery for cancer, where additional surgical training improves outcome; and vaginal hysterectomy. Rates of vaginal hysterectomy vary greatly among surgeons, and the learning curve to increase an individual surgeon's rate takes years. It has been suggested that "encouraging more surgeons to perform more vaginal hysterectomies may result initially in an increasing complication rate because it is technically more difficult."

In the United Kingdom and in North America, the baseline vaginal hysterectomy rate for non-malignant disease is 20%. Encouraging a group of surgeons to suddenly increase their rate from 20% to 60% would not be a meaningful way to evaluate the safety of the procedure compared with abdominal hysterectomy. Nor should all women have an abdominal hysterectomy because some are poor candidates for vaginal surgery and because some surgeons lack the skill or experience to support a safe, high vaginal hysterectomy rate. Case selection depends on diagnosis, parity, size and mobility of the uterus, and operator skill, which together determine a safe baseline rate. Increasing the rate arbitrarily and randomising such a complex mix of patient and operator characteristics would compromise safety, yet this is what happened in the term breech trial.

Homogenising populations and clinicians

Randomisation improves the internal validity of trials by homogenising study and control populations thereby avoiding bias from differences between the two. Clinically important factors that are variable within populations are, however, homogenised as well. Their importance to the outcome is lost, and the trial loses external validity for individuals within subsets of the population. Results represent the mean outcome for all participants and are not applicable to subgroups at lower risk. Although subgroup analysis, found in the term breech trial, can potentially identify subpopulations in which a procedure may be safer, it is statistically weak. When phenomena are complex, many patient and practitioner characteristics influence outcome, rendering individual subgroups small and meaningful analysis of them difficult.

A major limit of evidence based medicine is the difficulty in applying the results of randomised trials to individual patients. For example, most would agree that a multiparous woman in advanced labour at 38 weeks with a 5000 g fetus in a frank breech presentation with flexed head and no nuchal cord represents a low risk subgroup of all breech presentations. By studying a heterogeneous group of women, the term breech trial lacks the external validity needed to guide us with the management of such women. Yet experienced obstetricians will now press for an emergency caesarean, not trusting their clinical judgment to discern low risk situations, because all women with a breech presentation have been assigned a similar risk status by a randomised controlled trial.

Multicentre trials can also lead to homogenisation of the intervention. Previous randomised trials of breech presentations were too small to detect clinically meaningful differences in outcome, so a large multicentre trial was required to improve statistical power. Yet despite the interest, altruism, and self referential experience of the practitioners, the involvement of 121 centres resulted in an average level of care. Encouraging practitioners to exceed their comfort level with vaginal breech delivery lowered that standard.

If generic levels of care had always been accepted as the ideal, none of the surgical subspecialties would have arisen. The standard of care shown by the term breech trial is not the best we can offer. Although breech deliveries commonly occur under average conditions, it does not mean that committed centres are unable to offer better than average care. Collectively we have been improving our “mean” level of care for years, and the perinatal risk associated with breech delivery has continued to drop despite stabilisation of the caesarean section rate.
Simplified risk reduction

Historically, the greatest decrease in perinatal mortality from vaginal breech deliveries was reported by Bracht in 1938. Other techniques recommended to enhance the safety of vaginal breech delivery include routine determination of fetal weight, head attitude, and nuchal or presenting cord using ultrasonography; continuous fetal monitoring; radiological pelvimetry; cautious attention to the progress of labour; and preparing for emergency symphysiotomy should fetal parts become trapped. Although poorly amenable to scientific analysis, some of these techniques are likely to be important for safe vaginal breech delivery. None were included in the term breech trial. It is impractical for a large, multicentre trial to use complex risk reducing strategies. Meaningful quality control in 121 centres is impossible, and more caution would have meant fewer vaginal deliveries, increasing the number of participants needed to achieve similar statistical power. Therefore, the researchers chose a simple labour protocol with few risk avoidance strategies. The lack of proved effectiveness of other strategies ostensibly justified their exclusion; yet our current inability to analyse safe vaginal breech delivery does not preclude its existence. The resulting standard of care, arguably reasonable for a large, multicentre trial, falls short of its designation as the definitive study of vaginal breech delivery.

Since publication of the term breech trial, the onus has been placed on individual obstetrical units to retrospectively examine their experience with vaginal breech delivery and to show safety. Several have done so and continue to offer vaginal breech delivery. Several have done so and continue to offer vaginal breech delivery.21 22 Safety in these specific centres is due to heterogeneity of human skill, not to statistical anomaly, and vaginal breech delivery in those units should be studied and emulated. For complex phenomena, a large, randomised, multicentre trial does not overrule demonstrated safety.

In the case of carotid endarterectomy, it should ideally be performed at a centre and by a surgeon with a perioperative stroke rate of 3%, not 6%. If unavailable, a patient might elect medical treatment, as the risks could outweigh the benefits. Similarly, a woman with an average breech presentation and access to average care may decide that a caesarean section is safer than a trial of labour; yet even that conclusion is potentially flawed: without a bias of licence, the maternity unit caring for her may well have a low, safe, baseline vaginal breech delivery rate.

Short term combined end points

Randomised trials often utilise short term end points because they are easier to measure than longer term outcomes. It is also easier to show a statistical difference in a combined end point rather than a single end point. Yet combined end points can be misleading. In the term breech trial, the end point included perinatal mortality and various short term morbidities, including hypotonia, transient brachial plexus injury, and isolated low arterial cord pH or base excess, whose lasting significance is unclear. In countries with low perinatal mortality, this combined end point occurred in 5.7% of planned vaginal deliveries and 0.4% of women undergoing elective caesareans. Mortality was not significantly different (5 of 511 or 0.6%) in the planned vaginal delivery group compared with zero in the planned caesarean group. One of these deaths, included in the intention to treat analysis, occurred before the onset of labour in a cephalic twin weighing 1150 g, highlighting concerns about the adequacy of case selection and raising questions about the appropriateness of intention to treat analysis at all cost. Regardless, the impact of the trial's results was due primarily to an excess of short term morbidity in the planned vaginal delivery group.

Long term outcomes in breech babies are hard to assess epidemiologically, but were retrospectively shown to be equivalent in 1645 children, irrespective of the planned mode of delivery. Researchers from the term breech trial have published details on death or abnormal neurodevelopment over two years in a subgroup of 923 children from the term breech trial. They found a similar incidence: 2.8% in the planned vaginal delivery group and 3.1% in the elective caesarean section group. The use of a short term combined end point seems to have been misleading.

The limits of evidence based medicine

Delivering a breech presentation vaginally is a skill: guided by science, its safety relies on the experience of practitioners and caution. In the term breech trial, large scale randomisation, which homogenised both the study population and clinical intervention, resulted in an average level of care in an average population, limiting the trial’s external validity in centres showing above average skill and in women of below average risk. Encouraging practitioners to exceed their comfort level ensured a high vaginal delivery rate and adequate statistical power, but introduced a bias of licence and compromised safety. Using a combined short term end point overstated any true effect on long term neurodevelopment.

The philosophical limits of evidence based medicine include failing to appreciate and cultivate the complex nature of sound clinical judgment, failing to
appreciate the relevance of poorly quantifiable clinical phenomena that are obscured by randomisation, and devaluing the intangible differences between individuals, thus potentially devaluing them (patients and care providers). The condensation of vaginal breech delivery by one randomised controlled trial and its sweeping effect on clinical practice show how each of these philosophical limits can be exceeded.

The author thanks Ewart Woolley, whose skill in vaginal breech deliveries inspired him, and Robert Liston and Michael Klein for their support and direction.

Funding: None.

Competing interests: None declared.

Ethical approval: Not required.


6 University of Toronto Maternal, Infant and Reproductive Health Research Unit. Term breech trial study protocol. 1998-8.


(Q&A 5 October 2004)
Corrections and clarifications

Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery

In the final stage of production of this Education and Debate article by Andrew Kotaska, human error conspired with an electronic glitch to produce two correspondence addresses—only one of which was correct (30 October, pp 1039-42). As well as publishing Kotaska's email address, we wrongly inserted the email address for the corresponding author (P Garner) of the preceding article (although that email address also appeared in the correct place, attached to Garner's article).

Needy diagnosed hypothyroidism

We failed to spot a spelling mistake and an outdated drug name in this "10-minute consultation" article by H U Rehman and T A Bajwa (27 November, p 1271). It should be carpal (not carpel) tunnel syndrome, and the recommended international non-proprietary name (rINN) of the treatment for hypothyroidism is levothyroxine (not thyroxine, which is the former British approved name, BAN).

A memorable teacher

Serendipity

In retrospect, my belief in serendipity began in the summer of 1951. I was between my first and second years as a clinical medical student at the Welsh National School of Medicine when I was attracted to a medical secretary working in the MRC unit at nearby Llandough Hospital.

She told me that the unit was looking for two medical students who, for a modest stipend, would assist in an MRC survey of coal miners' chest diseases in one of the South Wales mining valleys. Apart from the attractions of the secretary, no stipend was too modest for me at that time, and I applied. Having been interviewed by Archie Cochrane, who was running the survey, I was the third choice of the four who applied. One of the first two dropped out, so I was in.

During that long distant summer, I cannot recall having been given any important duties beyond arranging x ray envelopes in alphabetical order. But I do remember many hours in the company of Archie Cochrane. It was a small team, and we spent long hours classifying chest x rays and poring over the data, which were well beyond my comprehension. We also spent time in his Jaguar car, of which he was quite proud, visiting individual miners in their homes who for many diverse and interesting reasons did not wish to be x rayed. Considerable energy was expended in ensuring that the survey should be 100%.

Once or twice he had a migraine, which meant a longer session the next day. To me, he was kind and quickwitted. I doubt that I ever gave him any of the answers he wanted, but he tried hard to make me think.