The social and cultural shaping of medical evidence: Case studies from pharmaceutical research and obstetric science

Raymond De Vries\textsuperscript{a,}\textsuperscript{*}, Trudo Lemmens\textsuperscript{b}

\textsuperscript{a}University of Michigan, Ann Arbor, MI, USA
\textsuperscript{b}Faculties of Law and Medicine, University of Toronto, Canada

Available online 22 December 2005

Abstract

Most critiques of evidence-based medicine (EBM) focus on the scientific shortcomings of the technique. Social scientists are more likely to criticize EBM for its ideological biases, a criticism that makes sociological sense but is difficult to substantiate. Using data from our studies of (1) the influence of pharmaceutical companies on the conduct and reporting of clinical trials, and (2) obstetric science in the Netherlands (where nearly one-third of births occur at home) we show how the evidence of evidence-based medicine is shaped by forces both structural and cultural. The threats to objective evidence are many, and, if EBM is to be true to its own principles, it must take these threats into account.

Keywords: Evidence-based medicine; Sociology of science; Obstetrics; Pharmaceutical research; Midwifery; The Netherlands

Introduction

Evidence-based medicine (EBM) has an irresistible logic (see Lambert, 2006): who can argue with the rational notion that clinical practice should be based on scientific research? Of course, social scientists, philosophers and historians will tell you that science is not as scientific as its adherents proclaim, but, in general, challenges from the humanities and social sciences are not taken very seriously by medical scientists and caregivers. When EBM is criticized, as it has been, it is for failing to be scientific enough. Much of the criticism of EBM focuses on the quality of the evidence employed. Critics note that EBM, in an effort to do meta-analyses, mixes studies of varying quality—summaries are made using research done with a range of subjects under a number of different conditions (Chikwe, 1996; Egger & Smith, 1998). EBM also is faulted for using evidence only from published studies, creating both a publication bias (Egger & Smith, 1998) and a “feasibility bias” (Zwitter, 2001). Still other critics worry that EBM is driven not by science but by management efforts to improve cost-effectiveness (see Freeman & Sweeney, 2001).

Resistance to EBM comes as no surprise. Medical practitioners and researchers have long been the beneficiaries of widespread public trust in medicine and science, giving them the freedom to make decisions based on their intuition and their past experience in the clinic and the lab. EBM calls this traditional authority into question, and, to add...
insult to injury, EBM hoists doctors and researchers by their own petard, demanding that they take science seriously. A humorous, and telling, example of the frustration associated with the displacement of professional judgment by EBM is found in a satire published recently on the pages of the British Medical Journal (BMJ): “Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized controlled trials.” With tongues firmly planted in cheeks, the authors review the medical evidence on the value of parachute use and conclude: (1) no RCTs of parachute use have been done, and (2) the basis for parachute use is “purely observational,” and could potentially be explained by a “healthy cohort” effect. On a more serious note, they assert: “individuals who insist that all interventions need to be validated by a randomized clinical trial need to come down to earth with a bump” (Smith & Pell, 2003, p. 1460). Advocates of EBM were not amused. The BMJ received 31 “rapid responses” to the piece, many of which expressed displeasure with what one writer called “Smith and Pell’s … attempt to pour ridicule on evidence based health care” (Griffiths, 2003).

As social scientists, we see EBM as a healthy and useful challenge to “medicine-as-traditionally practiced,” but like others, we worry about the quality of the evidence that is mustered by EBM. Our critique, however, does not focus on the inconsistencies and practical problems of data collection. We find EBM to be flawed, not because it fails to be scientific, but because—like all sciences—it imports the biases of researchers and clinicians. In this paper we explore two sources of these biases: the structural arrangement of clinical research and the cultural ideas that shape research questions and research design. Our evidence—drawn from separate studies of (1) the funding of drug research and (2) the use of research data to support government policy on home birth in the Netherlands—shows how bias finds its way into the scientific literature.

---

1See also, Clinicians for the Restoration of Autonomous Practice (CRAP) Writing Group, 2002.

2Each of us has written in greater length on these topics. For more detailed analysis of the influence of research funding arrangements on evidence, see Lemmens (2005); for a closer look at the way culture shapes research and the evidence it produces, see De Vries (2004).

---

Structuring bias: industry funding of drug research

Critics of EBM, which include both academics and investigative reporters (see, e.g., Angell, 2004; Lenzer, 2004), have had some success in identifying financial biases in medical research. Several recent studies challenge the claim of scientists that “fund-generated” bias is eliminated by carefully planned, randomized, and controlled trials; these studies show results of clinical trials supported by industry to be skewed in favor of the sponsor (see Bodenheimer, 2000; Drummond, 1997). Concerns about the impact of industry interests on research and medicine are not new, but it is revealing that they are now officially recognized as a direct threat to the reliability of medical research. Various agencies and organizations, including the Department of Health and Human Services (2004), the Association of American Medical Colleges (Task Force on Financial Conflicts of Interest, 2001, 2002), and the Institute of Medicine (Committee on Assessing The System for Protecting Human Research Participants, 2001) have voiced their concerns and called for better evaluation of the potential impact of such conflicts.

The problems associated with industry-sponsored research can be divided into two broad categories: (1) problems related to the subtle impact of industry interests on the design of a study, its conduct, and the interpretation of data, and (2) problems of outright manipulation of research, selective publication of data, and ghostwriting. Industry influence on medical research blurs the line between the conduct of science and the marketing of drugs and devices, thereby undercutting the very foundation of EBM. In fact, insistence on the need for EBM—with its over-emphasis on one particular methodology of establishing evidence, i.e. the randomized controlled trial—gives financial interests unprecedented power to shape medical practice (see Elliott, 2004; Lemmens, 2004, 2005).

In the last two decades, both meta-analyses and focused empirical studies have demonstrated that there is a statistically significant association between source of funding and research outcome. Stelfox, Chua, O’Rourke, and Detsky (1998) reviewed articles discussing the use of calcium-channel antagonists for the treatment of angina and hypertension and found that those funded by calcium-channel producers were much more likely to make positive recommendations regarding its use. Two more recent analyses of the scientific
literature dealing with the impact of industry funding on research outcomes show industry-sponsored trials to be three to four times more likely than non-commercially sponsored trials to lead to a conclusion that benefits the sponsor (Bekelman, Li & Gross, 2003; Lexchin, Bero, Djulbegovic, & Clark, 2003).

These studies stop short of accusing industry-funded researchers of deliberate manipulation of results. Favorable findings may be the result of other factors: industry sponsors are likely to select researchers with an existing preference for their products, and they are more inclined to fund trials with a high probability of generating positive outcomes. It is, after all, reasonable to expect that sponsors be inclined to fund a trial only when there is a significant expectation of success. Publication bias also plays a role here: editorial preference for positive findings is a fact of scientific life and not the result of the publication strategies of industry sponsors.

There are, however, indications that intentional manipulation of research by sponsors is occurring (Bodenheimer, 2000). Most recently, New York’s Attorney General sued the pharmaceutical company GlaxoSmithKline for selectively publishing positive data and suppressing negative data—including serious adverse events—of several trials involving its anti-depressant Paxil. The case was recently settled, after GlaxoSmithKline agreed with a court-supervised publication of all of its research data (Lemmens, 2004, 2005).

There are several ways to manipulate trial design and subject selection. The choice of statistical methods can influence the outcome of the study. Comparators in a clinical trial can be selected carefully to make positive results more likely. When a highly effective competitive drug is already on the market, for example, an industry sponsor may prefer a simple placebo control to a trial design including a comparator treatment arm. Studies also suggest that a sponsor’s drug is sometimes compared to an inappropriate dosage of a competitive drug, thus skewing the results in favor of the sponsor’s drug, by making the comparator look less efficient or more toxic (Lexchin et al., 2003). Finally, eligibility criteria can be manipulated so that those who are more likely to respond positively will be recruited into a study.

Concerns about bias are amplified by the industrial context in which medical research increasingly operates. Various authors have detailed the strengthening of industry’s grip on medical research (Angell, 2004; Bodenheimer, 2000; Krimsky, 2003; Rettig, 2000). Evidence of the industrialization of medical research is found in: (1) the proportional increase in research funded or conducted by pharmaceutical sponsors, (2) the growth of contract research organizations, and (3) in the establishment of industry–academia partnerships and the growing links between academic researchers and industry (Krimsky, 2003). The distinction between public institutions and private industry, and between basic and applied research, has faded; pharmaceutical companies are increasingly involved in the funding of basic research and academic institutions participate in bringing to the market publicly sponsored research outcomes.

The move of research out of academic centers has been accelerated by the rise of specialized contract research organizations (CROs). CROs conduct research within specialized private research centers, often contracting with primary care physicians in the community. Market pressures on CROs, together with financial incentives offered to physicians for the recruitment of their patients, are cause for concern. They may impact the safety of trial participants and undermine respect for inclusion criteria, thus affecting the integrity of research.

But there is a much more fundamental concern. Many CROs have developed into “full-service” organizations for drug companies, offering their expertise in conducting trials, doing research analysis and preparing manuscripts for publication. Several authors have described the CRO ghost-authorship process. CROs hire authors to write up trial results and offer the nearly finished manuscript as a publication to established academics in the field (Bodenheimer, 2000; Carpenter, 2002; Healy & Cattell, 2003). Flanagin et al. (1998) report that up to 11% of articles published in six leading academic journals used ghost authors. In the most extreme cases of ghostwriting, academic authors who sign on to the studies have not seen the data, did not write the text themselves, and did not participate in the analysis. Ironically, it is a win–win situation. Researchers lengthen their vitas and academic authorship helps a study get published in the most respected medical journals.

Healy and Cattell (2003) offer an extended example of ghost authorship, showing how Current Medical Directions (CMD), a company “dedicated to the development of innovative, high-quality health care information” (http://www.cmdconnect.com),
orchestrated the scientific promotion of sertraline, an anti-depressant launched by Pfizer in the 1990s. A comparison of documents from CMD with the results of a search of the database Medline, shows that 55 CMD-organized papers were published in the academic literature between 1998 and 2000. Particularly relevant for the debate about EBM is that ghostwritten articles get published in the most important medical journals, including, in this case, the Journal of the American Medical Association, the American Journal of Psychiatry and the Archives of General Psychiatry. The publication “impact factor”—an increasingly emphasized measure of success for medical publications3—of the ghostwritten articles was also significantly higher than that of the other, non-industry-sponsored articles.

Healy and Cattell indicate that of the 55 articles, 13 did not mention an author associated with the company or did not indicate that they were coordinated through a commercial agency. Contrary to previous “publication strategies”, which included the funding of special supplements to established medical journals, even a careful reader can no longer find an indication that this is part of a targeted marketing campaign. Scientific publications themselves have become a marketing tool and are increasingly controlled by industry interests.

Healy and Cattell highlight another problem associated with the commercialization of research: the large number of publications on one particular drug produced by a CRO and published in the most influential peer-reviewed journals. Even if independent investigators questioned the validity of some of these studies, developed their own investigator-driven research project, and managed to get these results published, they are unlikely to receive the same attention. Bekelman, Li and Gross (2003) concur with this observation, noting that industry-sponsored studies with positive results for the sponsors are often published more than once.

Commercialized research diminishes, if not eliminates, the countervailing force of independent scholarship. Two issues are worth mentioning here. First, the increasing competitive nature of research, characterized by difficulties in gaining access to subjects and the use of financial incentives for researchers and subjects, has put non-industry-funded researchers at a disadvantage (Lemmens & Miller, 2003). The latter do not have the same financial means to recruit patients into clinical trials and to solicit participation among primary care physicians. Second, pharmaceutical sponsors are in an increasingly dominant position when it comes to determining what type of clinical trials will be undertaken and the questions that will be investigated in these trials. In the context of the controversy surrounding the safety and efficacy selective serotonin reuptake inhibitors (SSRIs), industry critics have argued that appropriate clinical trials, investigating the potential risk for suicidal ideation in some patients, have not been developed.

Drug companies respond by pointing to the absence of statistically significant studies that prove risk, thus brushing aside “anecdotal” and “unscientific” clinical case reports of suicidal ideation and reports by individual patients. If this reading of the situation is correct, the insistence on conducting the randomized controlled trial—often seen as the pinnacle of scientific research that fits best the requirements of EBM—has thus become, in and of itself, a tool of industry interests. Arguments based on EBM can be used as a shield to protect financial interests of those who have a dominant position to the trials that create the evidence of EBM.

Finally, recent developments and controversies related to scientific review articles call attention to the way commercialized research directly threatens the public trust in the independence of medical science. Within medicine, review articles are among the core components of the development of good clinical practice. These articles are generally written by the most eminent experts in a particular field, and contain a critical and authoritative analysis of the literature in that field, frequently with explicit recommendations for physicians. Since they reflect the most recent developments in science and are supposed to contain an independent analysis of the best available treatments, they are a primary source for EBM. Because of their importance, review articles have traditionally been treated differently than other articles with respect to conflict of interest issues. Most established scientific journals require review articles to be written by established scholars with no financial interests in any of the products discussed and no financial relationship with the sponsors of such products. Remarkably, and in spite of the fact that financial conflicts of interests are increasingly cited as a serious source for

---

3A journal impact factor is a measure of the frequency with which the “average article” in a journal has been cited in a particular year or period. A journal with a higher impact factor is considered to be more prestigious or of higher quality than those journals with a lesser impact factor.
concern, the NEJM recently softened its rule, suggesting that only people with a “significant” conflict of interest would be excluded and that the editors would evaluate on a case-by-case basis whether financial interests created a problem. In an editorial, the editors invoked the need for realism to support their policy shift. It had become, according to the editors, simply too difficult to find authoritative specialists without a conflict of interest (Drazen & Curfman, 2002).

Other journals, reflecting a continuing belief in the independence of science and the integrity of its practitioners, insist that financial interests should not be invoked to exclude respected specialists from writing review articles or editorials. The idea behind this approach is that science is always verifiable and that peer review and possible exposure of errors are sufficient as intra-professional sanctions. A recent controversy surrounding the journal Nature Neuroscience highlights the questionable wisdom in this approach (Brownlee, 2004). The controversy erupted after the journal refused to publish a letter by two readers, who pointed out that one of two authors of a recent review article (Nemeroff & Owens, 2002) on new neurological treatments held a patent in one of the recommended products and was a prominent figure on the advisory board of two other companies who were likely to benefit directly from the recommendations in the article. Only after the controversy erupted in the press did the journal revise its disclosure policy to request disclosure of such interests in its article (Editorial, 2003). Notably, the only problem admitted was the lack of disclosure of significant conflicts of interest, not the corruption of evidence by the conflict.4

This close look at industry-funded research shows that the “scientific process”—often heralded as a fundamental part of EBM—is capable of being used as part of a marketing campaign (Angell, 2004). When commercial interests influence the design of the research protocol, the selection of research subjects, the conduct of the trial, the collection of data, the interpretation of results, and the publication of the outcome, there is good reason to worry about the integrity of the process that produces medical evidence.

The cultural biases of medical research

A second source of bias in medical research is culture: the collection of ideas, values, norms, and beliefs that characterize different societies. Unlike the structured biases of medical research discussed above, cultural bias is difficult to see because it is built into one’s taken for granted understanding of the world. Notice, for example, that EBM is most often used to evaluate the wisdom of supplanting one allopathic therapy with another: when EBM is simply an effort to determine a (statistically) significant difference between the outcomes under allopathic treatment “A” and “B”, cultural bias is nearly impossible to see. Testing “A” against “B” leaves unnoticed and unchallenged assumptions about the interchangeability of bodies, about the efficacy of allopathic intervention, about the randomness of error.5

In order to see cultural bias we must find situations where EBM is called on to compare the effectiveness of two practices based on differing assumptions about the body and the advisability of medical intervention (see Landsman, 2006; Narain-das, 2006). Only in these cases—and they are relatively rare—can we demonstrate that prior cultural understandings infect the collection of evidence, and thereby call attention to the cultural bias of all EBM.

One of those rare places where the influence of cultural ideas on the collection of medical evidence is visible is Dutch maternity care. Because it is a glaring exception in the world of modern obstetrics, the “Dutch way of birth” is well known among maternity care providers. Measured in terms of health outcomes or by its use of sophisticated technology, the medical system of the Netherlands rivals that of any nation, and yet nearly one-third of the births there take place at home. This number stands in stark contrast to the rest of the developed world: in no other country with a modern health system do more than 3% of births occur at home.

---

4Interestingly, there is a widespread assumption that disclosing a conflict of interest will somehow eliminate or discount the influence of that conflict. No evidence is adduced to support this assumption. Research suggests that disclosure may, in fact, have harmful results. Research from Carnegie Mellon University (Cain, Loewenstein, & Moore, 2003) suggests that disclosure gives the discloser license to be more biased while having no effect on the how information is perceived (i.e., disclosure does not lead to the discounting of information as biased).

5In recent years EBM has been used to test the efficacy of alternative and complementary therapies, but in these cases alternative therapies are judged by allopathic standards, leading advocates of alternative medicine to complain that EBM is missing the point (see Tonnelli & Callahan, 2001, and a rebuttal by Ernst, 2002).
Midwives are primary attendants at 71% of Dutch home births; they attend 48% of all of that nation’s births. It is a system that works quite well in terms of cost-efficiency and quality. The high midwife-attended home birth rate is coupled with the world’s lowest rates of surgical intervention in birth and very low rates of infant mortality.

Why has this unique system of care prevailed in the Netherlands and not elsewhere? The easiest way to answer to this question is to point to unique features of the Dutch health care system including:

- a health insurance system that directs childbearing women into primary care and encourages birth at home;
- excellent education for midwives;
- well-organized postpartum care.

But these systemic features cannot do all of the explaining. Why did the Dutch create this system when every other developed nation abandoned birth at home? Here we must look to culture. Several features of Dutch culture have contributed to the persistence of midwife-assisted birth at home. The Dutch were the first to build homes specially for the nuclear family, and these small, tidy, and well-lit homes remain the center of family life, have unique ideas about the role of women, who are seen as strong and independent and yet tied to the family, value thriftiness, which manifests itself in a rational approach to social policy, shun the heroic both in the military and in medicine, and have a unique sense of solidarity rooted in collaborative efforts to prevent flooding and in their history of pillarization.6

As a result of these cultural ideas, the Dutch believe that the home is the proper place for birth and that midwives are the preferred caregiver at birth. These beliefs create a special problem for Dutch gynecologists and obstetrical researchers. In light of the uniformity of obstetrics outside the Netherlands, and in light of the need to scientifically support health care practices, how do the Dutch defend their “anachronistic” way of birth? In struggling with this problem, the practitioners, researchers, supporters, and detractors of Dutch maternity care give us a detailed example of how the assumptions one brings into the gathering of evidence influence the shape and use of that evidence.7

This case is especially useful for our analysis because there is a clear and recognized division in Dutch obstetrics: some researchers—committed to the Dutch view of birth—are compelled to gather evidence to justify their exceptional maternity care policy; others—whose allegiance to the larger culture of obstetric medicine makes them skeptical of (and slightly embarrassed by) “old-fashioned” Dutch birth practices—have done research intended to expose the Dutch way of birth as dangerous. In examining this divide we see the operation of cultural bias in the collection, analysis, and interpretation of data.

All those who would do research on the place of birth face the same inherent difficulties. First, it is impossible to do randomized clinical trials on this topic. Not only would it be unethical to assign a women to give birth in a setting she would not ordinarily choose, it would also create a confounding variable: the emotional state of a women birthing in an environment she finds unfriendly would influence the outcome of birth. Second, extremely large samples are required to find significant differences between the outcome of home and hospital birth. The Dutch have developed a list of indications (the “VIL”—in English, the “Obstetrics Indications List”) that is used to separate “physiological” and “pathological” births; only women in the first category may give birth at home. This careful screening makes it difficult to discern the effects of caregiver and place of birth on morbidity and mortality: women defined as healthy by the VIL have very few poor outcomes. Given this reality, researchers have three choices: they can (1) use existing statistics, (2) do “prospective studies” that analyze outcomes based on an “intention to treat” design,8 or (3) devise new outcome measures capable of discovering small differences in outcomes. Review of the Dutch literature on place of birth shows

---

6This is an admittedly brief sketch of the way cultural ideas have shaped maternity care in the Netherlands. For a more complete explanation, see De Vries (2005, pp. 139–179).

7This review of the scientific debate over Dutch maternity care is based on several years of research in the Netherlands by De Vries (2005). Although the research reviewed here—published from the mid-1970s through the early 1990s—antedates the development and institutionalization of EBM, the influence of culture illustrated by this body of work remains a challenge for the collection of scientific evidence.

8This design—where analyses are based on planned, rather than the actual, place of birth—is necessary because of the simple fact that most complicated births end up in the hospital; to simply compare home and hospital births builds in a negative bias toward the hospital and a positive bias toward home birth.
interesting variation in methods used to gather evidence and the meanings assigned to that evidence.

Existing statistics

Both supporters and skeptics have used existing statistics to make their case about the Dutch way of birth. Huygen (1976) looked at changes in perinatal mortality rates within the Netherlands between 1953 and 1970 and discovered (p. 245): “the perinatal mortality of hospital deliveries went down from 65.0 to 33.8 (almost halved), but for home deliveries it went down from 21.6 to 6.9 (being less than one third the figure for 1953).” He concluded (p. 248): “Studies have proved that by good case selection and good prenatal care it is seldom necessary to rush emergencies to the hospital and it is possible to obtain exceedingly low perinatal mortality figures…in home deliveries.”

On the other side, Hoogendoorn (1978, 1986) used existing statistics to generate a strong critique of Dutch maternity care. In 1978 he published an analysis that showed provinces with high rates of hospitalized births also had the lowest rates of perinatal death. Eight years later he used existing statistics to compare the rates of decline in perinatal mortality in several European countries between 1970 and 1984. He concluded: “After 1940 and especially after 1950 the perinatal mortality rate in the Netherlands has shown a remarkable decrease, to the extent that the rate for 1982 was only 25% of the 1940 figure. Since 1982, however, this rate has stagnated. The proportion of deliveries at home has also decreased progressively until approximately 1980, but since has remained constant. In virtually all European countries, the perinatal mortality has decreased more than in the Netherlands, which has lost its relatively high favorable position. Reconsideration of the problems of obstetrical care and particularly also the desirability of home vs. clinical delivery appears necessary” (1986, p. 1439).

Hoogendoorn’s work generated a number of responses, mostly from supporters of the Dutch way of birth, who, not surprisingly, found his analyses wanting. His critics pointed out that correlation does not imply causation (a challenge that, interestingly, was not leveled against Huygen’s more favorable analysis). For example, Kloosterman (1978)—a well-known gynecologist and defender of the Dutch way of birth—noted the problem of spurious correlation in Hoogendoorn’s research: he agreed that there is a correlation between increased hospitalization and lower perinatal mortality, but pointed out that the monetary inflation is also strongly correlated with decreasing infant death. Kloosterman went on to look at the correlation between degree of hospitalization and perinatal mortality in the 13 largest cities in the Netherlands and found inconsistent results: perinatal mortality decreased in cities where the percent of hospital births declined and in cities where it increased.

Prospective studies

Studies that compare outcomes for women choosing home and hospital birth—all of whom are classified as “physiological” according to the VIL—have been used exclusively by supporters of the Dutch way of birth (see Damstra-Wijmenga, 1982; Eskes, 1989; Eskes, Van Alten & Treflers, 1993; Van Alten, Eskes, & Treflers, 1989).

The best known, and most often cited prospective study of birth outcome in the Netherlands is the “Wormerveer research.” This study followed the 7980 women who booked at one of the practices of independent midwives in the Dutch town of Wormerveer between 1969 and 1983. The research showed: “that the selection of pregnant women into groups at high and low risk is possible using the relatively modest means available to the midwife...the data available on perinatal mortality and infant morbidity warrant the conclusion that within the scope of the Dutch system of obstetric care it is possible to achieve very good results with midwifery care for selected women” (Van Alten, et al., 1989, pp. 660, 662).

The research of Wiegers, Kierse, Van der Zee and Berghs (1996) looked at the outcomes of 1836 births accompanied by midwives at home and in the polyclinic. To measure birth outcome, the researchers constructed a “perinatal outcome index” consisting of 22 items on childbirth, 9 on the condition of the newborn, and 5 on the condition of the mother after birth. They discovered that for women using midwife care, location of birth made no difference in outcome for primiparous women, when controlling for social and medical background. For
multiparous women, perinatal outcome was significantly better for planned home births than for planned hospital births, with or without control for background variables.

These prospective studies provide the evidence used by the government to create the policies that promote home birth in the Netherlands, evidence that is necessary to defend the peculiar Dutch approach to birth against the more universal culture of obstetrics. As Tew and Damstra-Wijmenga (1991, p. 55) point out, these studies contradict “the claims on which the organization of maternity services in most developed nations is now based, namely, that childbirth is made so much safer by the application of high technology that only this option should be provided.” The loyal opposition to the Dutch way of birth finds these studies insufficiently scientific; members of the opposition have instead looked for new, more sensitive measures to compare home and hospital birth.

Using new measures

The desire to find a better, more scientific way of comparing the outcomes of home and hospital birth led several researchers—from the Vrije Universiteit and the Katholieke Universiteit, Nijmegen, two schools known for their critical stance towards home birth and midwifery—to develop a research design based on objective measures of umbilical cord blood pH and neonatal behavior. The suspicion of these researchers—that in spite of favorable outcomes shown in prospective studies, midwife-assisted birth at home was unsafe—was not unreasonable. As noted above, in the vast majority of cases, women defined as healthy by the VIL, have excellent outcomes, regardless of where they give birth. In order to get around this problem, this team devised a research strategy that would allow comparisons to be made even when there was no overt morbidity and mortality. How? Researchers would look for small but significant differences in the pH of blood taken from the umbilical cord. Lower pH values are suggestive of oxygen deprivation (acidosis) and hence less than optimal outcomes for the neonatal brain. In conjunction with measures of cord blood pH, researchers also assessed outcomes using a scale that measured the neurological condition of the newborn. Developed by Prechtl (1977), the scale involves observation of the body and reflexes of the newborn. These close observations of the health status of the neonate introduced some variability into otherwise similar birth outcomes.

In a series of articles and papers these researchers proposed, tested, and defended their use of measures of umbilical cord blood pH and neurological scores as a fitting way to look more closely at the outcomes of home and hospital births (see De Jong, 1975; Stolte et al., 1979; Van den Berg-Helder, 1980). It was their suspicion that the management of birth at home would result in more acidotic children. If proved true, this would be bad news for supporters of the Dutch system because acidosis in fetuses is associated with growth retardation and damage to the central nervous system. In a paper presented at a conference in 1976, Stolte et al. (1979)—researchers advocating this new and more scientific approach to the study of the outcomes of Dutch maternity care—demonstrated a correlation between low pH values of umbilical cord blood and compromised neuromotor skills. Throughout the 1980s several studies were done using these outcome measures, most of which showed specialist care in hospitals to be superior to birth at home with midwives: on average, babies born at home and/or under midwife care were more acidic and had poorer neurological scores (see Eskes, Jongsma, & Houx, 1981; Lievaart & De Jong, 1982).

Taking a closer look at this body of work and the responses it generated allows us to see how cultural ideas (“of course birth at home is safe” vs. “of course birth at home is dangerous”) shape how evidence is gathered, interpreted, and reported.

One of the best-known publications from this research was a report authored by Lievaart and De Jong that appeared in The American Journal of Obstetrics and Gynecology (1982). In this research the 85 first births accompanied by midwives (65 at home and 20 in the hospital) were compared to 27 first births accompanied by gynecologists in the hospital. Both groups were considered “normal” births according to the obstetric indications list. To evaluate outcomes, the pH of cord blood was tested and the newborn was assessed using Prechtl’s method of neurological examination. Babies born at home did not fair as well as their hospital-born counterparts. Not only were the neurological scores worse—10 of the infants born at home were classified as “non-optimal,” while none of the hospital infants were—but so too the pH values. The authors concluded (p. 385):

The outcome advances more or less definite evidence that the obstetric system prevailing in
the Netherlands, although concomitant with satisfying neonatal mortality figures...is not adequate from the point of view of neonatal morbidity. The morbidity of the new born infants delivered under the care of midwives of pregnancies deemed by them as normal is without any doubt much higher than expected on the basis of the philosophy of the underlying system of obstetric care. The fallacy of the system is not rooted in the place to be born, e.g., home or hospital. It is also not preponderantly related—at least not in the present study—to the lack of capability of the midwives to select abnormal pregnancies among those pregnancies originally thought to be normal. The better outcome of infants born in the hospital under the care of a gynecologist is most probably (also) due to the tools of surveillance used in the supervision of deliveries, i.e., electronic monitoring and determination of fetal scalp blood pH and the capability of performing a cesarean section.

This study, published in the best-known American journal of obstetrics, presented a clear challenge to the Dutch way of birth. Pieter Treffers—a champion of home birth and then Chair of obstetrics at the University of Amsterdam—and his colleagues responded in a letter to the editor. Published nearly a year later, they criticized Lievaart and De Jong for:

- making conclusions about “the system prevailing in the Netherlands” based on so few cases from one region;
- misreporting the results of studies available only in Dutch;
- sloppiness in their research: pH values were given for only 81 of the 85 cases in the midwife group and 26 of 27 cases in the gynecologist group; there was no control over when the cord was clamped or how long the blood was stored before pH analysis was done—midwives are more likely to clamp late and late clamping lowers the cord blood pH as does prolonged storage of the blood;
- biased selection of cases.

The authors close their letter with this statement (Treffers, Van Alten, & Pel, 1983, p. 872):

We conclude that the evidence produced by [Lievaart and De Jong] is insufficient to support their pretentious statements and that the system they are propagating implies a very high level of active intervention, which, in itself, could have undesirable consequences.

Lievaart and De Jong (1983) defended their research in a reply published in the same issue. In most cases they responded adequately to the critique, but their response to the criticism of how they collected and handled the cord blood contains an important non sequitur (p. 873):

The laboratory housing the Corning 175 automatic pH and blood gas system were (sic) alongside the delivery rooms used by the midwives and the gynecologists. The acid-base measurements were performed by the same technicians immediately after the arrival of the blood. Consequently, the time intervals between the sampling of the cord blood and the assay did not differ between the gynecologist group and the midwife ambulatory group. Since the acid-values in the cord blood of the neonates delivered by the midwives in the hospital ambulatory unit did not differ from the values determined in the cord blood of the neonates delivered by the midwives at the patients’ homes we can safely assume that the influence of different transport and storage times was of minor importance.

Simply having the laboratory at the same distance from the delivery room does not insure that the time intervals did not differ between the two groups. In fact, the similarities found in pH values for home and hospital deliveries of midwives are likely to be the result of the fact that midwives practice the same way at home and in the hospital. In defending themselves in this way—by not reporting the actual time intervals for midwives and gynecologists—Lievaart and De Jong seem to acknowledge that they did not keep adequate records of clamping and storage times.

Unhappy with this study, Treffers and his colleagues replicated the research, paying careful attention to the collection and storage of cord blood. In one study the researchers measured the effects of various techniques of collecting and storing cord blood: they discovered slight variations in the time and temperature associated with storage had significant effects on pH levels, and they concluded that the only reliable way to measure the pH of cord blood is to puncture the cord immediately after birth, store the samples on ice and test them within 30 min. In a second study they used these findings to repeat the work of Lievaart and De
When researchers took pains to assure that cord blood samples from the clients of midwives and gynecologists were treated in an identical manner, the results were opposite to those reported by Lievaart and De Jong: women attended by midwives had significantly higher values for their cord blood pH. The researchers concluded (Knuist, Eskes, & Van Alten, 1987, p. 364):

This study shows with respect to umbilical pH values, that there is no cause for concern about the Dutch obstetric system in which midwives take care of pregnant women and deliveries.

Eager to get these results to the readers of American Journal of Obstetrics and Gynecology, Treffers and his colleagues attempted to get this research published there, but neither study was accepted for publication. Instead, the first was published in the Journal of Perinatal Medicine (Pel & Treffers, 1983), an English language journal published in Germany, and the second was accepted by the NTvG, the Dutch Journal of Medicine (Knuist et al., 1987).

An interesting footnote to this story comes from the work of two graduate students, Geert Berghs and Esmeralda Spaanjards, who were part of the project using cord blood pH and Prechtl’s scale to assess the safety of home birth. In their research, they followed 1034 “normal pregnancies” (as defined by the VL) attended by midwives (N = 638), general practitioners (N = 128), and gynecologists (N = 268). Using Prechtl’s scale they discovered that 84.6% of newborns in their study were “Normal,” 12.3% were “Suspect” (i.e., showed signs suggestive of neurological problems) and 3.1% were “Abnormal.” When the researchers analyzed neonatal outcome by caregiver, however, they found no differences between the three groups. They concluded that there were no significant differences in the outcomes of normal births managed by midwives, general practitioners and gynecologists. They did discover that the 268 women in the study who chose the care of a gynecologist were more likely to have experienced a medical intervention during birth and that their infants had higher rates of morbidity, measured by Apgar scores and eventual hospitalization (Berghs & Spaanjards, 1988).

In response to these seemingly contrary findings, Stolte and Eskes, directors of the project, claimed that a finding of “no difference” proved the superiority of specialist care because the population of women who choose care from gynecologists were more likely to smoke, were less educated and required more interventions. Eskes (1992) published an article formally making this argument, noting that there were no differences in neurological scores between groups “despite the lower socio-epidemiologic profile” of women under the care of gynecologists.

The cultural shape of evidence: supporting two sciences of birth

Idealized accounts of science suggest that science stands outside of custom, that it is not influenced by “they way things have always been done”. But evidence from the Netherlands reveals that accepted definitions of what is “normal” in pregnancy determine what is accepted by journal editors as good science. Editors of the American Journal of Obstetrics and Gynecology, most of whom do not share the cultural assumptions of the Dutch, accept the work of Lievaart and De Jong and reject the work of Treffers because of what they assume to be true about birth. They are unwilling to let research evidence influence their belief that “birth is normal only in retrospect.” Van Teijlingen points out that in the Netherlands the same process is at work, but it brings opposite results. He quotes a Dutch physician (Van Teijlingen, 1994, p. 180):

Eskes argued that Lievaart and De Jong’s article was rejected [for publication] in the Netherlands for ideological reasons and accepted in the US for scientific reasons. Whilst Kloosterman maintained that it was rejected in the Netherlands for scientific reasons and accepted in the US for ideological reasons.

The co-existence of two sciences of obstetrics in the Netherlands demonstrates how ideology about the “best” way to give birth affects both the generation and interpretation of the evidence that

---

10 One interviewee reported that the American Journal of Obstetrics and Gynecology refused to publish an article by Berghs and Spaanjards based on their research that showed extremely low inter-observer agreement about the interpretation of electronic fetal monitoring recordings taken during the second stage of labor (see Berghs & Spaanjards, 1988, pp. 129–140). The letter of refusal stated that it would be “immoral” to publish these results.
is so vital to EBM (see Gordon, 2006; Landsman, 2006; Naraimdas, 2006). Research by advocates of home birth supports the safety of birth at home while research done by those opposed to domiciliary deliveries provides evidence that the practice is unsafe. The contrast between the Dutch system and obstetric systems in other developed countries also suggests the source of the ideology that drives EBM. Our evidence suggests that mainstream obstetric science follows mainstream obstetric practice. A patient and expectant approach to birth in the clinic, where all is considered normal until proved otherwise, produces a science that proves intervention to be unnecessary. Alternatively, an aggressive approach to birth in the clinic, where birth is regarded in normal only in retrospect, generates a science that demonstrates the need for monitoring and intervention.

In his work on “decision support techniques” and EBM, Berg (1997) has shown that guidelines interact with clinical practices to create new and varied forms of care (see also Gordon, Landsman, 2006; this volume). Obstetric science in the Netherlands shows that clinical practice also plays an important part of in the creation of guidelines. The assumed relation between science and practice is turned on its head: practice is not based on science; rather science is based on practice. In order to explain why mainstream obstetric science in the Netherlands is the opposite of mainstream obstetric science in other developed countries, we must explain why the practice of obstetrics is so different there: we must look to the social structures that give rise to its maternity care system and to the cultural values that generated and sustain those structures. In seeking to understand the rejection of Dutch obstetric science in mainstream medical journals published outside the Netherlands, we must look to the cultural ideas that shape the practices in those societies.12

Conclusion

Our research suggests that evidence of EBM suffers from biases generated by funding and from biases built into the way researchers perceive the world. While EBM offers a much-needed critique of medical practice based on hunches and anecdotal experience, it is susceptible to manipulation by those who fund research and to distortion by cultural assumptions that determine outcomes.

We have shown how a commendable desire to demand the evidence of randomized clinical trials can work to the advantage of industry interests: the alchemy of the clinical trial transforms EBM from a challenger to a protector of corporate agendas. Our data from the Netherlands demonstrate that the cultural assumptions of researchers, visible in existing in clinical practice, shape the gathering and interpretation of evidence. Awareness of these threats to the integrity of the evidence that drives EBM is a necessary step toward making EBM an effective tool for clinicians.

Acknowledgements

This paper was written when De Vries and Lemmens were colleagues at the Institute for Advanced Study, Princeton, NJ, USA, and De Vries held a faculty appointment at St. Olaf College, Northfield, MN, USA. Funding for this research was supplied by the Fogarty Center (National Institutes of Health, NIH) Grant No. F06-TW01954 (De Vries) and National Center for Complementary and Alternative Medicine (NIH) Grant No. K01AT000054-01 (De Vries), the Netherlands Institute for Health Services Research (NIVEL), Utrecht (De Vries), and by Genome Canada through the Ontario Genomics Institute (Lemmens).

References


Ernst, E. (2002). What’s the point of rigorous research on alternative/complementary medicine? Journal of the Royal Society of Medicine, 95, 211–313.


