

American Journal of Evaluation

<http://aje.sagepub.com>

Blowback: Consequences of Evaluation for Evaluation

Ernest R. House

American Journal of Evaluation 2008; 29; 416 originally published online Sep 5, 2008;
DOI: 10.1177/1098214008322640

The online version of this article can be found at:
<http://aje.sagepub.com/cgi/content/abstract/29/4/416>

Published by:



<http://www.sagepublications.com>

On behalf of:

[American Evaluation Association](#)

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: <http://aje.sagepub.com/cgi/alerts>

Subscriptions: <http://aje.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://aje.sagepub.com/cgi/content/refs/29/4/416>



Blowback

Consequences of Evaluation for Evaluation

Ernest R. House

University of Colorado Boulder

Drug studies are often cited as the best exemplars of evaluation design. However, many of these studies are seriously biased in favor of positive findings for the drugs evaluated, even to the point where dangerous effects are hidden. In spite of using randomized designs and double blinding, drug companies have found ways of producing the results they want, including manipulation of treatment, selection of sample, control of data, and calculated publication. Regulatory agencies have been neutralized. We have entered an era when evaluations are controlled by sponsors to produce the findings they want. Evaluations have become too important to be left to the evaluators. Such deceptive practices threaten the integrity of the evaluation field, perhaps its existence. There is no doubt these practices will spread into educational and social evaluation.

Keywords: *drug evaluation; bias; politics; randomized trials*

In July 2001, totally unexpectedly, I found myself having bypass surgery. The heart surgery was surreal; could this be happening? A few days after surgery, I was in intensive care when the head resident arrived with three assistants. He read numbers aloud off the chart—cell counts, blood pressures, pulses, six things I didn't recognize. He said to one young assistant, "He's anemic, low pulse, low blood pressure, etc. What should we do?" The assistant said, "Give him a unit of blood, and increase the heart beat to 70." The head resident said, "No, no! This man's an athlete. He's been swimming for 30 years. That's why his pulse is so low. Haven't you read his history? He doesn't even need this pacemaker!" With that he walked over to me and ripped off the external pacemaker attached to wires protruding from my chest. I thought to myself, "I hope to hell this guy knows what he's talking about."

He did. I recovered. What impressed me throughout the ordeal was that the quantitative indicators the MDs based decisions on actually meant something—unlike the world of educational and social research. The experience was fascinating, though I can't say I recommend it for entertainment.

Since then I have been taking five prescription drugs, plus aspirin. I hope these drugs have been rigorously evaluated. Unfortunately, there is reason to be concerned. Although some regard randomized drug trials as the ideal for all evaluations, drug studies are among the most biased evaluations being conducted. What's happening could be a precursor for what happens elsewhere in evaluation.

Drug Evaluations as Symptomatic

Let me declare my vested interests. I take two drugs for cholesterol and three for blood pressure. I am dependent on these drugs. Furthermore, I have a significant financial interest

Author's Note: Presented at the American Evaluation Association, Portland, OR, November 4, 2006. The author thanks Gene Glass, Steve Lapan, Bill Madura, Bob Stake, Kevin Wellner, and Bob Williams for useful comments on drafts of this article. Ernest R. House, University of Colorado, Campus Box 249, Boulder, CA 80309; e-mail: ernie.house@colorado.edu.

in the pharmaceutical firms that profit from selling drugs. I have an interest in how these studies are done both as a patient and an investor.

Until the 1980s, drug researchers were pretty much independent of drug companies, but now the companies are involved in every aspect of drug evaluation. The former editor of the *New England Journal of Medicine* writes, “Researchers don’t control clinical trials anymore; sponsors do” (Angell, 2004, p. 100). Drug evaluations have become increasingly biased, though they are randomized studies. Here’s how bias enters.

Choice of placebo as comparator. Drug companies insist on using placebos against which to compare new drugs. In other words, the comparison of a new drug is against no treatment, not against other treatment drugs in use. Because 77% of new drugs are slight variations of those already on the market, many new drugs would not look more effective than the old drugs in head-to-head trials.

Selection of subjects. One way of preventing side effects from showing up in clinical trials is to use younger rather than older subjects. Young people suffer fewer side effects generally. For example, in trials of nonsteroidal, anti-inflammatory drugs, only 2% of the subjects were aged more than 65 years, though these drugs are targeted explicitly at people aged 65 years or more (Bodenheimer, 2000).

Manipulation of dosages. If a company does wish to show that its drug is more effective than a competitor, it may compare the two by using lower dosages of the competitor drug and higher dosages of the new drug. For example, when AstraZeneca’s heartburn drug Prilosec went off patent, the company took a new patent out on the active ingredient and called it Nexium (Angell, 2004). In four clinical trials, the company compared higher doses of Nexium to Prilosec. Nexium looked like an improvement though it was the same drug. Or in a review of clinical trials of nonsteroidal, anti-inflammatory drugs, 48% of the trials used higher doses of the sponsoring company’s drug. Or in trials comparing two statins, Pfizer compared 80 mg of Lipitor to 40 mg of Pravachol, a competitor drug.

Method of drug administration. Sometimes the competitor drug will be administered in such a way as to make it less effective, such as by giving it orally rather than by injection (Bodenheimer, 2000).

Manipulation of timescales. Drugs are tested for a short time, even when used for much longer periods. Blood pressure drugs are tested for a few months though taken for a lifetime. Timescale lies at the heart of the Vioxx controversy (Pollack & Abelson, 2006). Merck did not count those subjects who dropped out of the treatment group because Vioxx was making them ill if these people didn’t have heart attacks or strokes within 14 days after dropping out, even when they went on to have coronary events soon after.

Selective publishing. Pfizer published favorable findings for 6-month trials of Celebrex, an arthritis drug, even though the company had findings in hand for 12 months showing no effect. The Food and Drug Administration (FDA) requires only two positive trials for drug approval, no matter how many trials are conducted, and the agency has no control over what the company publishes. Companies must submit all clinical results to the FDA, but they can publish only favorable ones if they choose (Mathews, 2005).

Suppression of negative results. One review of 122 articles in the *Journal of the American Medical Association* found that 65% of harmful effects were not completely reported. Also, in 62% of clinical trials, the primary outcome was changed. Authors cherry-picked results (Mathews, 2005). Researchers are subject to binding contracts that allow companies to determine what can be revealed from the study, and companies have repeatedly denied permission for reporting of negative information (Armstrong, 2006; Harris, 2006; Mathews, 2005; Zimmerman & Tomsho, 2005).

Opportunistic data analysis. Companies retain control over data analyses by collecting data from several sites and providing only partial data to cooperating researchers. As one drug executive said, "We are reluctant to provide the data tape because some investigators want to take the data beyond where the data should go" (Bodenheimer, 2000, p. 1541).

Control of authorship. Drug companies employ ghost writers to write reports and circulate drafts to investigators who will be listed as authors. Reports are written to show the product in the most favorable light. How much listed authors can change the draft depends on how much they want to argue with the company (Bodenheimer, 2000).

The high-profile Vioxx study displays several sources of bias: timescale manipulation, concealment of data, suspect statistical analysis, and deceptive publication (Berenson, 2006a, 2006b; Pollack & Abelson, 2006). Furthermore, 20 million people took Vioxx over a period of 5 years before it was withdrawn. It may have caused 100,000 heart attacks. There are 11,500 lawsuits pending.

You may think such evaluations are the exception, but they are not. Marcia Angell, former editor of the *New England Journal of Medicine*, writes, "Bias is now rampant in drug trials" (Angell, 2004, p. 106). Reviews of clinical trials substantiate her judgment (Als-Nielsen, Chen, Gluud, & Kjaergard, 2003; Lexchin, Bero, Djulbegovic, & Clark, 2003). Danish researchers analyzed 370 randomized drug trials chosen from a sample of 167 meta-analyses in the Cochrane Library (Als-Nielsen et al., 2003). They found that the studies recommended the experimental drug as the "treatment of choice" in 51% of trials sponsored by for-profit organizations compared to 16% sponsored by nonprofits. The predictors of conclusions were source of funding, treatment effect, and double blinding. If researchers are connected to the company whose drug is evaluated, the findings are 4 times more likely to be favorable to the product than when researchers have no connections (Angell, 2004).

Swedish researchers compared full clinical trial data to what was actually published (Melander, Ahlqvist-Rastad, Meijer, & Beermann, 2003). They found that the same clinical trials were included in several publications without any indication of prior publication. Sometimes authors' names had been changed. Favorable results were 3 times more likely to be reported. What this means is that meta-analyses would unknowingly include the same trials repeatedly as if they were separate instances, thus inflating the apparent effectiveness of the drug in the meta-analyses. The Swedish researchers concluded that without access to all clinical trial data, not just published studies, one cannot make valid evidence-based recommendations.

How can these things happen in scientific research? Drug companies have become involved in funding, designing, collecting data, analyzing data, interpreting results, authoring, disseminating findings, and controlling publication. In 1990, 80% of drug studies were conducted by academic institutions. In 2000, 40% of studies were done by academics. Drug companies now hire special for-profit contract research organizations (CROs) to implement the evaluation designs of

the companies. These contract organizations are heavily dependent for funding on the drug firms. Some universities—at risk of losing hundreds of millions of dollars by losing the clinical trials—have agreed to cooperate with the companies. Academics are bound by detailed contracts with sponsors that limit independence in every way. More than 70% of funding for clinical trials now comes from the companies.

Where is the government in all this? In the early 1990s, drug companies began providing money to the FDA for drug trials, accounting for 53% of the FDA drug review budget in 2005. This allows companies to negotiate with the FDA over drug reviews. David Kessler, who began the practice as FDA commissioner under Clinton, is troubled. He says there is no doubt the practice gives industry leverage over the agency. According to the *Wall Street Journal*, regulators usually don't negotiate their activities with industries they regulate (Mathews, 2006). In fact, many of those negotiating for the drug industry are former FDA officials. Congress does oversee the agency, but many key members of Congress accept high-paying jobs as drug lobbyists when they leave office. In 2002, the industry employed 675 Washington lobbyists and makes large political contributions.

To complicate matters further, National Institutes of Health (NIH) and the FDA have allowed their scientists to accept consultant fees and stock options from companies without having to report them (Angell, 2004). Many independent academics who sit on panels that approve drugs for the FDA receive large sums of consultant money from the companies. Often these financial interests are not revealed. I think it is safe to say that government oversight of the industry has been severely weakened. I am also sure many inside the agencies lament what is happening. Internal dissent has reached the media (Committee on the Assessment of the U.S. Drug Safety System, 2006; Pear, 2004).

For their part, drug companies say that academics take too long to do studies. It costs between US\$300 and US\$600 million to develop a new drug, and each day of delay costs the company a million dollars. It is a big, expensive, and profitable business. In 2002, the combined profits of the 10 drug companies listed in the *Fortune* 500 were more than the combined profits for all the other 490 corporations put together, US\$35.9 billion versus US\$33.7 billion (Angell, 2004). And 1992 legislation requires the FDA to report to Congress whether it is meeting the short timelines for approving new drugs, shortening the time for reviewing the drugs.

A few observations are as follows:

First, these drugs and companies provide enormous benefits. I can say with conviction that they are a matter of good or bad health, even life and death, for many.

Second, bias has become a serious problem in drug evaluations, enough to threaten the great good these drugs do.

Third, government agencies (and universities) are losing control to the companies.

Fourth, our conception of bias control is inadequate to deal with the situation.

Fifth, drug evaluation has been transformed as a consequence of what evaluators do—provide honest assessments—because the findings mean so much, not because they mean so little. Something similar is possible—I would say likely—in other areas of evaluation. Indeed, drug studies provide a template for how to bias evaluation studies, even when these studies are randomized and double blinded.

Threats to the Health of the Evaluation Community

Why is this happening? In my book *Professional Evaluation* (House, 1993), which discusses the rise of the field of evaluation, I employed a framework from the great French historian

Fernand Braudel, founder of the *Annales* school of history. His work *Civilization and Capitalism* traces the evolution of capitalism as an institution from the 15th through the 18th centuries (Braudel, 1981, 1982, 1984). Capitalism started from simple village markets and evolved over centuries into powerful financial institutions—banks, corporations, stock exchanges. Braudel's focus is on institutions developing over long time periods as a result of complex interactions with the environment and other institutions.

As capitalism progresses, it erodes traditional social structures. Markets are no respecters of tradition. People lose their jobs, move to find better jobs, and live away from their families, their communities, even their countries. They are adrift, far from familiar customs and beliefs. Capitalism transforms what it touches. One political economist wrote, “. . . everything has a price, and . . . its value is its price. As a consequence, markets have a profound and destabilizing impact on society because they dissolve traditional structures and social relations” (Gilpin, 1987, p. 20).

As capitalism erodes tradition, governing becomes more difficult. Historically, governments depended on traditional social structures for support. They justified their actions through precedent and custom, appeals to religion, and legitimating myths and symbols. But in advanced capitalist society, authority is demystified, and governments have difficulty maintaining legitimacy through traditional means. Increasingly, they justify their actions through nontraditional means, such as increasing material well-being, image management, and even scientific authority.

Since World War II, the use of science to legitimate government actions has grown enormously, and evaluation partakes of scientific authority when it passes judgment on programs. It is no accident that evaluation as a professional practice emerged first in the most capitalist country, the United States, where decisions about what to eat and how to live are based in part on scientific findings. Evaluation serves to legitimate government actions. Hence, the rise of professional evaluation.

As capitalism advances even further, key functions, like economic activities, move beyond government control. The United States may have reached the point where so much capacity has been drained from government that it cannot perform, or has difficulty performing, tasks considered essential to the public welfare. This erosion of capacity has resulted from transferring resources, decisions, authority, and legitimacy from public to private control. Private control of resources is asserted to be more efficient, a tenet of Chicago economics. Ronald Reagan said, “The government is not the solution to our problems; the government is the problem.” Put more simply by the man from Crawford, Texas, “It's the people's money the government collects, and the government should give it back.”

We see how drug evaluation has been transformed. Congress did not provide enough money for clinical trials and drug reviews. The FDA turned to industry for funds, and industry responded—for a price. It negotiates with the agency over its priorities. Furthermore, the companies took control of clinical trials by funding for-profit organizations dependent on the companies. Funds, decisions, and control moved from public to private hands.

In general, government has become seriously underresourced and is constrained from augmenting its resource base. It is besieged by demands for services, but cannot meet those demands. Put bluntly, government cannot control major events, nor sometimes even mitigate the effects of events, contrary to expectations. Much has been written about the erosion of trust in government. People don't trust the government for good reason. The government has lost the capacity to perform its duties.

Consider. Can the government provide national security? It was not capable of defending against the terrorist attacks of 9/11 and is not capable of winning the war in Iraq or maintaining stability in Afghanistan. Can the government defend the country's borders? It is helpless against

millions of illegal immigrants streaming in, to the point where vigilantes patrol the borders. Can it protect citizens from natural disasters? Not only did the government fail to protect the citizens of New Orleans from Hurricane Katrina, it failed to assist them during and right after. New Orleans remains a disaster zone, to the incredulity of people all over the world, who perceive the United States to be acting like a helpless, third world country. National security, public health, and border control are key government functions, even by conservative standards.

Yes, I know the Bush administration is incompetent, perhaps the most incompetent government in American history. I have written about the neofundamentalist mentality that grips the administration, a rigid mindset that cannot accept or allow information counter to its beliefs. And when government tries to do something big, as in Iraq or New Orleans, it hires private contractors over whom it can exercise neither control nor accountability (Wolfe, 2006). But something even more basic is happening. The government has been stripped of resources to the point where it cannot perform basic tasks. If the next administration is competent or liberal, it will be unable to reverse the trends that have eroded capacity.

What has this to do with evaluation? If government cannot control events, it can control information about events. Or try to. It needs to control information to mask its ineffectiveness and declining legitimacy. If the purpose of evaluation is to honestly determine the value of programs, the consequences of doing so when government actions are successful is different from the consequences when the actions are ineffective. Instead of legitimating government action, honest evaluation will delegitimize a struggling government.

Indeed, the Bush administration already controls information to an unprecedented degree, even in science. In 2004, a group of 60 prominent scientists, including 20 Nobel Prize winners, asserted that the administration had misrepresented scientific knowledge and misled the public about the implications of its policies (Glanz, 2004). The administration censored and suppressed reports, stacked advisory committees, disbanded panels that gave unwanted advice, and refused to seek independent expertise. The abuse of science has been deliberate and systematic on environmental, health, biomedical, and weaponry issues. And you see how information control played out in drug evaluations. Unhappy with impartial results, the companies devised their own system of evaluation.

I expect something similar in other areas. Here is how the process might work in education. The Bush administration is a strong supporter of private schools and charter schools, which involve transferring resources, decisions, and capacity from public to private hands. The U.S. Department of Education has sponsored two studies, both of which show that private and charter schools do not raise test scores any more than public schools do. The Department of Education suppressed the first study and disputed the findings of the second.

The next step might be studies conducted by advocates of charter schools designed to obtain favorable results. Such an analysis might obtain findings favorable to charter schools by eliminating covariates like surrogate measures of low income that are normal in such studies. Eventually, as charter schools gain resources, they may sponsor studies of themselves that show their programs favorably—just like drug companies. We are early in the transformation of educational and social evaluation, but that's how sponsor control might work, very similar to the drug company behavior. They provide a model of bad behavior.

Remedies

Clearly, we have a problem, maybe the biggest threat to evaluation since its inception. What remedies are available for salvaging honest evaluation? We can begin by looking at

what medical researchers propose for drug trials. First, the comparator drugs in these trials should be drugs already used for treating the condition, not placebos. Marcia Angell says that if she could choose only one reform, this would be it (Angell, 2004). If there is concern about the effectiveness of the old drug or side effects, the clinical trials could compare the new drug, old drug, and a placebo.

A second reform would be to strengthen the FDA. The FDA is far too dependent on the drug industry and has lost capacity to perform its regulatory mission. I am sure many inside the agency would welcome support from Congress to eliminate its dependence. FDA advisory committees should not include members with financial ties to industry (Angell, 2004). Some scientists say that these financial ties do not affect their judgment. Right. Perhaps we should allow judges to take money from defendants whose trials they oversee.

Third, drug companies should not control clinical testing of their own drugs. Imagine that General Motors controlled the evaluation of its cars. Is there any doubt its products would appear excellent? One idea is to establish an institute for drug trials within NIH that would ensure that clinical trials were properly designed and conducted before FDA approval (Angell, 2004). The institute could contract with independent researchers to conduct the trials, and data would be controlled by NIH and the evaluators. Drug companies are supposed to publish such data on government Web sites, except they don't name the drugs in the onsite data, which makes a mockery of the requirement (Zimmerman & Tomsho, 2005).

Fourth, those who conduct drug trials should be free from financial conflicts. How big a problem is financial conflict for researchers? At the Stanford medical center, 700 faculty members listed 299 potential conflicts of interest in the conduct of their research (Pollack, 2006). Stanford recently prohibited the acceptance of gifts from industry, a step already taken by the University of California, San Francisco (Moynihan, 2003a). Unfortunately, merely revealing financial ties does not prevent those ties from influencing the research. Transparency is necessary, but not sufficient. The conflict of interest is too deep.

How likely are these reforms? Not very likely. The reforms run counter to privatization and the government losing capacity. These changes would increase government control. Reform seems unlikely until we have a catastrophe in which huge numbers of people die because a drug was inadequately tested. Congress must see the problem starkly, not easy to do when surrounded by hundreds of lobbyists. Unfortunately, such a tragedy is highly likely. CEOs of companies have very short time frames. Beat Wall Street profit estimates for a few quarters and leave. They will be long gone before trouble erupts. Economists call this moral hazard, not being accountable for what you cause.

A few days after I finished this article, the Institute of Medicine's Committee on the Assessment of the U.S. Drug Safety System, part of the National Academy of Sciences, released a 265-page report on drug safety. The report does not address how studies are conducted, but focuses instead on the FDA. It says the FDA is in serious disarray because of inadequate resources and too much industry influence. The agency is under extreme time pressure to review massive amounts of material. Drug safety is neglected because safety depends on tracking drugs once they reach the general population. Drugs in use are administered to a far more diverse population than in clinical trials, and it is here that many side effects appear. Clinical trials do not and cannot guarantee drug safety. Last year, there were 450,000 reports of adverse drug reactions. The FDA does not have the resources nor authority to follow up. Resources are concentrated on approving new drugs. The report recommends extensive changes to the FDA and to how companies market drugs.

Some reforms can be accomplished without government. The journals are a first line of defense. A few years ago, Jeffrey Drazen, current editor of the *New England Journal of Medicine*, thought clinical trials were not a problem, realized they were a big problem, but not his, and now

says, “We have to do something” (Armstrong, 2006; Mathews, 2005; Zimmerman & Tomsho, 2005). A dozen journal editors have proposed new standards for publication, like full data disclosure from clinical trials, transparency about financial interests, submission of original designs so reviewers can see how the study has been modified, and signed statements from authors that they have written the report.

Professional medical associations have proposed codes of ethics that include restrictions on gifts and financial ties, and some universities have issued rules about authorship and drug company ties (Mayor, 2003). As usual, student groups have been strongest in advocating ethical behavior (Moynihan, 2003b). One recommendation is that sponsors not be allowed to overrule researchers when it comes to publishing negative findings (Garattini, Bertele, & Bassi, 2003). Of course, drug companies are resistant to such changes (Moynihan, 2003b).

Finally, our conception of rigor in evaluation studies is far too weak to contend with these potent threats to impartiality. In reviewing the drug literature, I encountered several comments that the clinical trials were highly biased but that the methodologies were good. Here’s one statement: “A recent systematic review of the impact of financial conflicts on biomedical research found that studies financed by industry, although as rigorous as other studies, always found outcomes favourable to the sponsoring company” (Lexchin et al., 2003, p. 1167). How can you have a rigorous methodology that yields biased results?

Apparently, the researchers mean that the study is randomized, has hidden allocation, and is double blinded. No doubt these are important. But they do little to handle deliberate biases that arise from opportunistic choice of comparator, dosage, administration, time frame, selection of surrogate endpoints, cherry-picking data analyses, selective reporting, sponsorship, and financial ties. Our notion of rigor and good method seems to be based on a narrow, almost formulaic conception. The companies are deliberately designing studies to take on the trappings of rigor and biasing the study by other means. In other words, they are gaming our conceptions of validity and bias control and the FDA approval process.

When evaluation started as a professional practice, it was natural that evaluators would conceive biases in field studies as resulting from the same sources that social psychologists encountered in lab experiments. In Campbell and Stanley’s (1963) seminal work on experimental design, the sources of invalidity are maturation, mortality, instrumentation, selection, repeated testing, and so on, all of which are important. However, such a list is inadequate for what we face in a less innocent time.

Lab studies occur in contexts not heavily infused with politics; evaluations occur in settings with powerful political forces. Drug evaluation is riddled with politics, and those politics profoundly affect the findings. We need conceptions of bias control that address these sources of bias. We need to acknowledge vested interests in evaluations and quit pretending that they don’t affect findings. Otherwise, we look foolish or complicit.

Why should we bother? There have always been those who say we should conduct studies the way those who pay for them want. I find that course unacceptable for professional, ethical, and personal reasons. I believe the profession is at serious risk here. The world doesn’t need evaluators who have no credibility any more than it needs auditors who have none. R.I.P., Arthur Anderson. As for ethics, what about those 100,000 people who had heart attacks and strokes because the Vioxx evaluation was handled improperly? Surely, we must have ethical concern for the welfare of patients.

Recently, Alan Ryan (2004) introduced me at the Canadian Evaluation Society by saying that over the years I have reminded evaluators of their moral responsibility and alerted them to the dangers of being seduced by the agendas of those in power (Ryan, 2004). Those are some of the kindest things anyone has said about me. Others say I am just a pain in the ass—a dull one at that.

Moral responsibility aside, you can look at this situation personally. As a patient I might suffer a serious side effect from an improperly evaluated drug, but as an investor I can make a lot of money from the companies. That doesn't seem like a good trade-off to me. You may not be taking one of these drugs with your health hanging in balance—but sooner or later, you, or someone close to you, just might be.

So there it is—blowback. The consequences of evaluation are transforming evaluation itself because those subject to evaluation want to control the findings. What started during the Johnson years as an effort to evaluate programs intended to complete Roosevelt's New Deal has encountered the reversal of the New Deal during the Reagan–Bush years. The great society has morphed into the gilded age. The butterfly has entered the cocoon and emerged as a caterpillar.

Afterword

This article was presented at the American Evaluation Association conference in 2006. Since that time the media have been filled with more reports of drugs inappropriately evaluated for the reasons discussed. Unfortunately, my concerns about the drugs I have been taking were prescient. The latest study on Vytarin, a combination of a statin (Zocor) and Zetia, another cholesterol-reducing drug, has shown the combination to be ineffective and possibly harmful. On its own, each drug has been shown to be effective.

Even as Vytarin was heavily advertised on television, the sponsoring drug companies refused to release the findings of the study until Congress threatened them. The study indicates that though the combination lowers cholesterol, beneficial effects did not materialize. In fact, the buildup of plaque in the aorta of patients was greater with the combination. No one knows why. The study was conducted on a sample of patients particularly susceptible to plaque accumulation with the idea that if it retarded buildup in this group it would help more typical patients. Now that the study has produced negative findings, the companies argue the findings don't apply because they were based on an abnormal group. Hiding negative findings, changing the goals of the study, disputing the relevance of the sample—familiar stuff.

So many other drug studies have been so flawed that Congress has taken an interest, though without much import so far. Drug lobbyists remain powerful in Washington. The Institute of Medicine review that appeared in 2006 as I finished my original article indicated that the problems are political and structural. With the Bush administration business interests always dominate no matter what the issue is. Heads of government agencies have little hesitancy overruling findings and regulations that contradict administration ideology or commercial interests. According to the Institute of Medicine's report, drug company influence on the FDA is far greater than it should be.

Another issue is how the FDA is organized to conduct evaluations. One unit oversees the initial drug approval studies, which involve randomized field trials. This unit absorbs most of the resources. A different unit tracks the effects of drugs over time. This unit is seriously underfunded and underappreciated, according to the Institute of Medicine's report. Considering how many drugs interact with other drugs, as in the Vytarin study, there is no way the FDA could do randomized studies on tens of thousands of drug combinations and interactions. No budget would be large enough. Tracking the unanticipated effects of drugs on the market is the only feasible safeguard in the long run. This task should receive much more attention and funding. Admittedly, one can see why the drug companies are not particularly interested in tracking drugs once they are selling them. From a sales standpoint, the results can only be negative. Direction and stricter regulation must come from the government. A few tentative steps have been taken. Perhaps a new administration will take the problem more seriously.

Some readers have interpreted my article as being opposed to randomized trials. That is not my point. I am very much in favor of randomized controlled trials (RCTs) when appropriate, and the final stages of drug development is one of those places. In other places they may be less useful. Randomized trials are not foolproof, nor are they sufficient to prove beyond a doubt something works. No method is sufficient to provide conclusive proof. As the drug studies show, RCTs can be manipulated to obtain biased findings. Overconfidence in RCTs is naive and mistaken. However, they are still invaluable, if oversold.

My main point in the article is that evaluation studies of all types can be and are manipulated in the interests of those who stand to profit. Caution, prevention, and protection are necessary. One cannot let the market determine what happens. Market fundamentalism, the idea that markets are self-regulating and governments should not interfere, is untrue. We are paying the price in many areas, health care being one of them.

What is the relevance of this point for educational and social evaluations? I have not provided many examples, focusing instead on drug studies. The reason I focused on drug studies is that they are held up as an outstanding example of how to do scientific studies. Perhaps they are, but they can also be biased, as demonstrated. If this can be done with drug studies, it can be done with other evaluations. The application to education and social programs is straightforward. If the sponsors of programs control the evaluations and the evaluators, they may well design evaluations to further their own self-interests rather than provide unbiased information for the public interest. Such practices delegitimize evaluation to the point where evaluations are misleading, pernicious, and dangerous. I see no reason such behavior might not plague educational and social evaluation. There is no good reason to have dishonest evaluation.

References

- Als-Nielsen, B., Chen, W., Glud, C., & Kjaergard, L. (2003). Association of funding and conclusions in randomized drug trials. *Journal of the American Medical Association*, 290, 921.
- Angell, M. (2004). *The truth about the drug companies*. New York: Random House.
- Armstrong, D. (2006, May 15). How the New England Journal missed warning signs on Vioxx. *Wall Street Journal*, pp. A1-2.
- Berenson, A. (2006a, May 13). Follow-up study on Vioxx safety is disputed. *New York Times*, p. B3.
- Berenson, A. (2006b, May 31). Merck admits a data error on Vioxx. *New York Times*, p. C1-2.
- Bodenheimer, T. (2000). Uneasy alliance: Clinical investigators and the pharmaceutical industry. *New England Journal of Medicine*, 342, 1539-1544.
- Braudel, F. (1981, 1982, 1984). *Civilization and capitalism: 15th to the 18th centuries* (3 vols.). New York: Harper and Row.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Committee on the Assessment of the U.S. Drug Safety System. (2006). A. Baciú, K. Stratton, & S. P. Burke (Eds.), *The future of drug safety: Promoting and protecting the health of the public*. Washington, DC: Institute of Medicine, National Academies.
- Garattini, S., Bertele, V., & Bassi, L. L. (2003, May 31). How can research ethics committees protect patients better? *British Medical Journal*, 326, 1199-1201.
- Gilpin, R. (1987). *The political economy of international relations*. Princeton, NJ: Princeton University Press.
- Glanz, J. (2004, February 18). Scientists accuse White House of distorting facts. *New York Times*, pp. D1-2.
- Harris, G. (2006, September 29). F.D.A. says Bayer failed to reveal drug risk study. *New York Times*, pp. A1, A9.
- House, E. R. (1993). *Professional evaluation*. Thousand Oaks, CA: Sage.
- Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. (2003, May 31). Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *British Medical Journal*, 326, 1167-1170.
- Mathews, A. W. (2005, May 10). Worrysome ailment in medicine: Misleading journal articles. *Wall Street Journal*, pp. A1-2.

- Mathews, A. W. (2006, September 1). Drug firms use financial clout to push industry agenda at FDA. *Wall Street Journal*, pp. A1-2.
- Mayor, S. (2003, May 31). World Medical Association reviews doctors' links with drug companies. *British Medical Journal*, 326, 1165.
- Melander, H., Ahlqvist-Rastad, J., Meijer, G., & Beermann, B. (2003, May 31). Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: Review of studies in new drug applications. *British Medical Journal*, 326, 1171-1173.
- Moynihan, R. (2003a, May 31). Who pays for the pizza? Redefining the relationships between doctors and drug companies. 1: Entanglement. *British Medical Journal*, 326, 1189-1192.
- Moynihan, R. (2003b, May 31). Who pays for the pizza? Redefining the relationships between doctors and drug companies. 2: Disentanglement. *British Medical Journal*, 326, 1193-1196.
- Pear, R. (2004, February 22). Taking spin out of report that made bad into good health. *New York Times*, pp. D1-2.
- Pollack, A. (2006, September 12). Stanford to ban drug makers' gifts to doctors, even pens. *New York Times*, p. C2.
- Pollack, A., & Abelson, R. (2006, May 22). Why the data diverge on the dangers of Vioxx. *New York Times*, pp. C1, 5.
- Ryan, A. (2004). Foreword to E. R. House: The role of the evaluator in a political world. *Canadian Journal of Program Evaluation*, 19(2), 1-16.
- Wolfe, A. (2006, July/August). Why conservatives can't govern. *Washington Monthly*, pp. 15-16.
- Zimmerman, R., & Tomsho, R. (2005, May 26). Medical editor turns activist on drug trials. *Wall Street Journal*, pp. B1-2.