Looking for the Evidence in Medicine

An ambitious collaboration aims to review every trial of a medical treatment ever done to find out what works, what doesn't—and where future research should be focused

"There's a gold mine out there, completely forgotten and abandoned," says University of California, San Francisco (UCSF), physiologist Drummond Rennie. That gold mine is a medical one: the roughly 1 million randomized control trials of medical treatments that have been conducted over the last 50 years or so. Most of those trials—which means, in effect, the bulk of the scientific evidence on the ultimate value of treatments and interventions—have been either forgotten or simply lost. Now Rennie and his colleagues in an ambitious research collaboration are sifting through all that forgotten ore to recover its full value.

Known as the Cochrane Collaboration, after the late British epidemiologist Archie Cochrane, the 2-year-old collaboration is literally hand searching the world's literature to find and review all of the randomized control trials ever published—together with whatever unpublished trials can be located—and publishing the findings in an electronic form. Updated regularly, these reviews should synthesize the latest state of knowledge about every available therapy or intervention and give its implications for practice and research.

The several thousand volunteers, coordinated by 10 centers worldwide, who make up the Cochrane Collaboration expect their efforts to transform medical practice. Unlike drugs and medical devices, which can't be sold unless their effectiveness has been documented, actual treatments are often dictated by authority, personal experience, and habit. The Cochrane researchers hope to bring in a new force for clinical decision-making: what Oxford University epidemiologist David Sackett has described as "the conscientious, explicit, and judicious use of current best evidence."

Heavy meta. Screens from Cochrane database present a meta-analysis of anticoagulant therapy for stroke.

The collaboration shares that aim with a larger movement known as evidence-based medicine (see box), of which it is the most ambitious example. And like the movement as a whole, it is stirring controversy. "The Cochrane Collaboration is an enterprise that rivals the Human Genome Project in its potential implications for modern medicine," wrote David Naylor, chief executive officer of the University of Toronto's Institute for Clinical Evaluative Sciences, in *The Lancet* last April. But he and others fear that the Cochrane reviews will take on the weight of dogma.

"What [the Cochrane Collaboration] is trying to achieve is 100% laudable," says University of Chicago epidemiologist John Bailar, "but it may end up with a lot of apparently strong, precise results that are nowhere near as strong or precise as they appear to be." Lapses of quality control or judgment could easily slip into the enormous review process, say critics, and they also question the underlying methodology, which relies heavily on a controversial statistical technique called meta-analysis.

Naylor adds that even if the reviews are reliable, medicine can't always be done by cookbook. "Good clinical practice will always involve some elements of inference and judgment leavened by experience," he says. Other health care experts note, however, that the Cochrane Collaboration and a half dozen smaller projects aimed at narrower areas of medical practice are simply trying to help doctors do what they have always done: assess the available evidence. "It's not that physicians haven't been trying to be evidence-based in large measure," says Donald Berwick, chief executive officer of the Boston-based Institute for Health Care Improvement. "It's just that the base of evidence has gotten more and more complicated and now requires some other mechanism to digest it."

Mining the literature

The Cochrane Collaboration, which hopes to provide that mechanism, is the offspring of a project initiated by Oxford physician Iain Chalmers and colleagues in 1976: a systematic review of all randomized clinical trials having to do with pregnancy and childbirth. Between 1978 and 1984, Chalmers; Murray Enkin, an epidemiologist and obstetrician at McMaster University in Hamilton, Ontario; and obstetrician Mark Keirse, now with Flinders University in Australia, organized a search of 70 journals dating back to 1950 to identify all relevant trials in the area. Chalmers, with epidemiologists Kay Dickersin of the University of Maryland medical school, Curt Meinert of Johns Hopkins School of Public Health, and the late Thomas Chalmers (no relation) of New York's Mount Sinai Medical Center, also wrote to 42,000 obstetricians and pediatricians requesting information on any trials or data that may never have been published.

The search turned up 3500 trials, out of which the researchers selected for review only those that met specific criteria. For instance, says Enkin, trials only qualified if they used a formal randomization procedure or a quasi-random method for dividing patients between the treatment group and the control. Systematic reviews of the trials then yielded verdicts on the effectiveness of 600 different interventions and treatments.

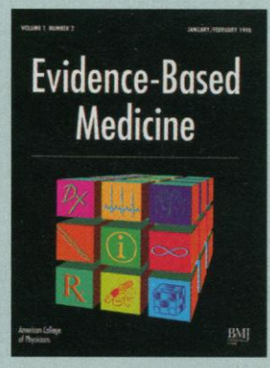
The results, first published in a 1989 volume entitled *Effective Care in Pregnancy and Childbirth*, included a number of eye-openers. One, says Enkin, was a clear message about the value of giving corticosteroids to women who are about to give birth prematurely, to speed up the maturation of the fetus. By 1981, seven randomized trials had been completed, and the results were impressive: The inexpensive, simple treatment halved infant mortality. "And yet as late as 1993," says Enkin, "one really good survey showed that only 23% of women who should have received the drug actually got it. Putting it crudely, doctors just didn't pay any attention to the evidence."

By now some, at least, are paying attention, says David Grimes, head of obstetrics and gynecology at San Francisco General

Giving Doctors the Lowdown on the Literature

Evidence-Based Medicine, the title of a new British journal, may seem an unlikely slogan for an insurgency. What is medicine based on, you might wonder, if not evidence? But the unspoken premise of the British journal and its counterpart, the American College of Physicians' *Journal Club*, is that all too often physicians rely on custom, hearsay, and dogma in choosing treatments. Instead, these journals and other programs aim to publicize the actual evidence about what does or doesn't work.

That's also the goal of the Cochrane Collaboration, a massive effort to do systematic reviews of the entire published and unpublished literature on the whole of medicine (see main text). But the Cochrane reviews are an ongoing project. In the meantime these two journals are abstracting the best systematic reviews published elsewhere and then adding expert commentary, giving doctors an early taste of this approach to medicine.



Novel idea? Journal summarizes published reviews.

This recent movement dates to 1976, when the Canadian Task Force on the Periodic Health Examination reviewed the literature on a venerable preventive-medicine practice, the annual checkup. It came to an unexpected conclusion: The annual checkup should be abandoned and replaced by age- and sex-specific preventive procedures of proven value—such as childhood immunizations and screening for hypertension and depression. That prompted the U.S. Office of Disease Prevention and Health Promotion to establish a U.S. version of the Canadian work, called the Preventive Services Task Force, that reviewed 169 commonly used clinical preventive services.

The group published its first findings in 1989 as the *Guide to Clinical Preventive Services* and extended them in a second volume of the clinical guide, published last December. Among them were some surprises. It found little evidence to support the use of electronic fetal monitoring during labor, for example, even though monitoring is done in nearly 80% of all deliveries, says task force member David Grimes, head of obstetrics and gynecology at San Francisco General Hospital. Similarly, says Grimes, "one of the time-honored parts of male physical examination has been the digital rectal examination, and we found fair evidence against doing it. ... A lot of men will be delighted to hear that news."

While some hospitals and physicians claim that these efforts are transforming clinical practice, the evidence so far is equivocal. That's the case even for the original Canadian study of the annual physical, says David Naylor, chief executive officer of the Institute for Clinical Evaluative Sciences in Toronto. "Publishing recommendations and guidelines on medical practice, even when authored by respected peers," he notes, "is no guarantee that anyone will follow them."

–G.T.

Hospital and an epidemiologist at UCSF. *Effective Care* has constituted a "sea change for us clinically. ... It's now on computer available to my staff, students, residents, and attending physicians around the clock, 7 days a week. And we use it on a daily basis. With a few keystrokes, we can find out what's out there. It provides the scientific substrate to the practice of medicine."

The favorable reception prompted Britain's National Health Service (NHS) in 1992 to fund a Cochrane Center to try and extend the strategy to other fields of medicine. A year later, the NHS helped to initiate the Cochrane Collaboration, which is now supported by centers in Oxford; Lyon, France; Copenhagen, Denmark; Amsterdam, the Netherlands; Milan, Italy; Baltimore; San Antonio; San Francisco; Adelaide, Australia; and Hamilton, Ontario—"wherever there are individuals willing to fulfill the functions expected of centers," says Chalmers. While the European Cochrane Centers are funded by their respective governments, the three U.S. centers have been getting funding wherever they can find it. The San Antonio center, for example, is supported by the Federal Department of Veterans Affairs, while the San Francisco and Baltimore centers still haven't found reliable funding sources.

Each center's primary purpose is to marshal volunteer review groups, which focus on specific conditions—stroke or schizophrenia, for example. Reviewers, each of whom is responsible for studying a particular treatment or intervention, are committing themselves to a grueling task, and they sign on for life, says Chalmers. "We spend a lot of time trying to dissuade people from getting involved, so that if they do, they know with absolute certainty what they're getting into." Those who persist attend workshops to learn how to develop a review protocol—a set of questions to be answered and an approach to answering them—and use the collaboration software, which, says Chalmers, "forces you to take a systematic approach to doing the review." They then work with editors and external reviewers to refine their protocol and carry it through to a final review.

The review groups amass material for their reviews by, among other methods, hand searching all the journals in the world likely to publish randomized control trials on their particular subjects—or having the journals perform the searches themselves. The collaborators are also trained to dig up unpublished data by every possible means, says Dickersin, including pressuring pharmaceutical companies to release results on studies that they may have preferred to keep quiet.

In the past 2 years, nearly 100 systematic reviews have been completed, covering conditions including stroke, subfertility, schizophrenia, and parasitic diseases. Completed reviews are being published in the Cochrane Database of Systematic Reviews, which is available on floppy disks, CD-ROM, and, by the end of the year, over the World Wide Web. But it will take decades of work for the project to review the entire width and breadth of medical practice, Chalmers says. And reviews, in any case, will never actually be completed, but rather will be living documents. The Cochrane volunteers will watch the literature, updating the database and revising the original review when necessary—indeed, that's why the volunteers sign on for life.

The last word?

Each review, in theory, represents the best possible assessment of each treatment's value. But critics say the results may not be as definitive as they look. While some of the reviews give no more than a qualitative conclusion, others will go a step further and do what's called a meta-analysis when the literature search yields enough sufficiently similar studies. In meta-analysis, a single quantitative conclusion is extracted by synthesizing the data of many different trials.

That's where many epidemiologists start to worry, including Bailar. "The idea of meta-analysis is immensely appealing," he says, "and I wish it worked." Bailar argues that because of the extraordinary difficulty involved in doing it right, it often yields "a

result that appears to be of the highest obtainable precision but is error-ridden, and possibly quite unreliable. ... A careful, thorough, thoughtful, old-fashioned narrative review by somebody who really understands the subject is often a much better way to go at it."

Even one of the developers of modern meta-analysis, Oxford epidemiologist Richard Peto, worries about the Cochrane Collaboration's effort to do meta-analysis on an industrial scale. Good meta-analysis, he says, takes painstaking attention to detail and often requires involving the authors of the original studies. "You get the data from them," he says, "you go over the data in detail, do a lot of to and fro getting rid of errors, and then you get them involved in the interpretation both of their own study and the overview. This just isn't possible in the Cochrane Collaboration."

Members of the Cochrane Collaboration respond that a systematic, standardized analysis of the available evidence, whatever its warts, has to be better than no analysis at all, and that the collaboration has put a lot of work into setting up protocols to minimize the potential problems. "We've got a handbook of how to do [the reviews] that's about 3 inches thick," says Rennie. And even when mistakes creep in, he says, they have a good chance of being rooted out, because the collaboration will be subjecting its conclusions to what might be called perpetual peer review and refereeing. With conventional journal articles, says Rennie, later criticism—even evidence that an article is worthless—may have little effect. "It's as if the *Titanic* hit an iceberg but went sailing on," he says.

In contrast, the San Francisco center, run by Rennie and health policy researcher Lisa Bero, is setting up a system that should allow anyone reading a Cochrane review to e-mail criticisms to the center for relaying to the authors. If necessary, says Bero, the authors will revise their original review, which could mean coming to entirely new conclusions on the basis of new evidence. "It has to be done," says Rennie. "An archive is dead if it doesn't change and doesn't have ongoing review, criticism, and quality control."

At the very least, both critics and enthusiasts of the Cochrane Collaboration say the reviews should provide important information about what isn't known in medicine. Or, as Naylor puts it, "The wonderful thing about the systematic approach of reviewing available evidence across every imaginable aspect of clinical practice is that it will, one hopes, nail down those areas where we just don't know what we're doing"—whether blood-thinning or clot-busting drugs are beneficial in acute strokes, for example.

It should also tell doctors when a procedure is so well documented that it should become a part of standard practice. Says Dickersin, "The Cochrane reviews should finally tell us when we have enough evidence to believe something."

–Gary Taubes

WOMEN AND TENURE

No Women Chemists at Women's College

When Joan Valentine studied chemistry at Smith College in the mid-1960s, it didn't seem strange that all her professors were men. "That was the norm" in the sciences, Valentine says, and the elite women's college in Northampton, Massachusetts, was no exception. Today, however, it is. While Valentine went on to become the first tenured woman chemist at the University of California, Los Angeles, and the five schools comparable to Smith—the so-called "sister colleges"—now have between one and five tenured women chemists apiece, Smith, in the past 38 years, has had none. In February, the school extended this streak, rejecting physical chemist Sharon Palmer's tenure bid, despite unanimous support for her from the department.

The rejection—due to what the school says is a poor publication record—has prompted distress both on and off campus. That a women's college can't seem to find a qualified woman is "really sending the wrong message," says Rebecca Pinto, a senior chemistry major at Smith who has helped to circulate petitions in support of Palmer. "She is what we aspire to be, and she didn't get tenure—it's really dispiriting." Valentine, too, is pained about her alma mater. "I couldn't possibly, in good conscience, recommend that any young woman go there" to study chemistry, she says. And the case has renewed many concerns about the demands for extracurricular commitments that seem to come between women scientists and their research.

Because Palmer, 37, is the senior of two tenure-track women in her eight-member department, "women students identify with her, and there is a lot of advising she has to do because the students go to her first for certain sorts of problems" such as career decisions or conflicts with other instructors, says George Fleck, a tenured chemist in the department. Palmer is also "extremely strong" at guiding undergraduate research projects, says Petra Turowski, the department's other untenured woman. "And when you spend almost all of your time training students, you get very little [publishable data]."

"SIX SISTERS" FULL-TIME COLLEGE CHEMISTRY FACULTY

	Tenured		Tenure-Track	
	Women	Men	Women	Men
Barnard	1	1	2	2
Bryn Mawr	2	2	1	0
Mount Holyoke	3	1.4	1	2
Smith	0	6	2	0
Vassar	3	1	0	4
Wellesley	5	4	3	4

"Committee assignments and service work that draw [women faculty members] away from research" is a serious problem that hasn't faded as more women have entered science, says Marge Cavanaugh, a program director in the National Science Foundation's division of chemistry. Valentine adds that for protection from these ballooning demands, young faculty members need good advice from colleagues and department chairs. "If Smith's chemistry department fell down in getting the message to Palmer that 'this is what will be required for promotion,' then something very strange was going on," she says.

Smith chemistry chair Robert Linck says Palmer "did indeed receive guidance" on setting aside time for her own research. And he along with Smith's five other senior chemists—all men—recommended her for tenure. Still, the school's tenure committee balked, saying that only three of Palmer's 13 published papers had been produced during her 8 years at Smith.

John Connolly, dean of the Smith faculty and a member of the tenure committee, would not comment specifically on Palmer's case. He did say "there is adequate support at the college for all untenured members of the faculty to develop both their teaching and their research. ... That's not to say these means work equally well for every candidate." He adds that most departments at Smith employ roughly equal numbers of men and women, and indeed three of the college's four tenured physics professors and five of its 13 tenured biology professors are women. The chemistry department, undergoing a scheduled review this month, "has already raised this as a problem," Connolly says.

Palmer has until next spring to pack her bags, but says she's considering filing a grievance with the college over the handling of her tenure case. "I've had students tell me, 'I always wanted to be an academic scientist, but having seen what you go through, now I don't,' " Palmer says. "I should be showing them that it can be done."

–Wade Roush