Meta-Analysis in the Breech

A controversial method for grouping results from disparate studies may ultimately revolutionize how research—particularly medical research—is done; for now, the fur is flying

WHEN PRINCESS EUGENIE WAS BORN to Sarah, Duchess of York, last 23 March, she arrived in the world feetfirst. Like many obstetricians, the royal *accoucheur* treated the princess's breech presentation by cesarean section—a practice long attacked by feminists, who claim that the ancient midwife's practice of externally "turning" the baby at term is safe, effective, painless, and less costly. For years, no definitive clinical study had been able to put an end to the controversy. But by the time of Eugenie's birth, a new element had entered the fray: a controversial statistical technique known as "meta-analysis."

Put briefly, meta-analysis is the use of formal statistical techniques to sum up a body of separate (but similar) experiments. It is like an ordinary scientific review of research, except that ordinary reviews provide a qualitative—and often subjective—assessment of a few studies; meta-analysis, on the other hand, promises a *quantitative* synthesis of all available data. "It's a boon for policy-makers who find themselves faced with a mountain of conflicting studies," says Kay Dickersin, an epidemiologist at the University of Maryland. "That's what everyone likes about it and that's also what everyone is worried about."

"Meta-analysis is the wave of the future," says Thomas Chalmers, a former president of Mount Sinai Hospital who is now at the Harvard School of Public Health. "The days of the expert supposedly putting the state of the field into a review article are numbered."

If so, this change won't occur without a fight. Meta-analysis has provoked acrimony in every discipline—from psychology to physics—where it has been applied. "To some people," says Richard Kronmal, a biostatistician at the University of Washington, "it seems like little more than an attempt by statisticians to put themselves on the top of the totem pole. Individual researchers with their individual experiments see themselves reduced to becoming a cog in the great statistical wheel. And they're saying, well, no, that's not how science works."

A remarkable report, published at the end of last year in the United Kingdom, offers a window onto the possibilities raised by meta-analysis and the vitriol it evokes. *Effective Care in Pregnancy and Childbirth* is the most extensive collection of meta-analyses yet compiled. The two-volume, 1516-page, $400 work reviews more than 3000 randomized controlled clinical trials in perinatal medicine. Along the way it bluntly rejects such procedures as routine episiotomy (cutting the tissue between the vagina and anus to facilitate delivery), restricting weight gain during pregnancy (to prevent hypertension), and repeating cesarean sections routinely after a woman has had one. The study equally bluntly endorses such relatively neglected practices as vacuum extraction (rather than forceps), the use of corticosteroids for women who are delivering prematurely, and external turning for breech births.

The book has triggered an extraordinary range of reactions in the medical profession. The *Medical Journal of Australia* described it as "arguably the most important publication in obstetrics since William Smellie wrote *A Treatise on the Theory and Practice of Midwifery* in 1752." But it was denounced by the editor of the *Journal of Obstetrics and Gynaecology*: "The price of £225 should protect aspiring registrars [residents] from acquiring too many confused ideas from its pages." And one doctor called its authors "an obstetrical Baader-Meinhof gang."

The denunciations don't bother the moving spirits behind the report. "Some [obstetricians] hate it," says Iain Chalmers (no relation to Thomas), director of the National Perinatal Epidemiology Unit at Oxford and one of the report's three editors. "Of course they do. We have very strong evidence that obstetricians should do some things they are not doing, and we call into question the relevance of some of the things they are doing." In his view, obstetricians, like other researchers, base decisions on an unreliable selection of the available data, which itself is often not controlled for random error. "What we've tried to do," he says, "is to select unbiased treatment comparisons and to control random error by using meta-analysis."

Much of the storm comes from the fact that meta-analysis may overturn one of the most deeply ingrained traditions in science: the formation of judgments based on the "authoritative" scientific review article. "A few years ago I looked at review articles of subjects like radiotherapy for patients with radical mastectomies, coronary artery sur-

Some Meta-analyses

Subject and year	Meta-analysts	No. of studies	Findings
Effects of desegregation on academic performance of black students (1982)	T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, and P. Wortman	157	Desegregation has a tiny positive effect on reading scores and no effect on math scores. More important, formal analysis revealed glaring methodological weaknesses in all but 19 of the studies, suggesting that great effort had not succeeded in providing much of a database.
Measurement of gravitational constant (1930)	P. R. Heyl	3	Tests using different materials gave different results for the gravitational constant. An early meta-analysis, using simpler techniques than those of today, provided a more precise answer than any of the three studies taken alone.
Use of diagnostic nuclear magnetic resonance imaging (1988)	L. S. Cooper, T. C. Chalmers, M. McCally, J. Berrier, H. S. Sacks	54	At a time when nuclear magnetic resonance imaging was widely promoted as superior to computerized tomography, no good evidence existed for this belief.
Effect of coaching on SAT scores (1983)	R. DerSimonian and N. M. Laird	36	Coaching is only slightly effective. Interestingly, the probability that an experiment will find coaching to be effective is strongly tied to its methodology: observational studies find coaching of much greater use than randomized studies.
Deinstitutionalization in mental health (1983)	R. P. Straw	30	Neither the zealots for or against deinstitutionalization are right. Mentally ill patients fare equally well (or poorly) in hospitals or in alternative, less institutional settings.

gery, and emergency surgery for bleeding peptic ulcer," says Thomas Chalmers. "The opinions of the 'experts' who wrote reviews were always dependent on how they were trained, not on the body of evidence. That convinced me that on average the opinion of experts is no good."

Worse, according to Chalmers, such informal reviews can easily miss important phenomena. To prove this, Robert Rosenthal of Harvard and Harris Cooper, now at the University of Missouri in Columbia, constructed a 1980 experiment in which 41 graduate students and faculty members were asked to review seven earlier, published investigations into whether sex could account for differences in tenacity of effort. Half of the subjects were asked to use "whatever criteria you would use if this exercise were being undertaken for a term paper or a manuscript for publication"; the other half were taught meta-analytical techniques. The result: informal reviewers were unlikely to find any sex differences; statistical aggregation, however, showed a small but significant effect in favor of women.

Such problems aren't merely theoretical. They enter directly in questions of national policy, for example, an arena where meta-analysis might be used, but hasn't been so

How to Perform a Meta-Analysis

"Doing a meta-analysis is easy," says Ingram Olkin, a statistician at Stanford and the coauthor of a textbook on the subject. "Doing one well is hard."

Say, for example, that you wanted to do a meta-analysis of the many studies of drugs intended to prevent recurrence of a heart attack. There are a variety of different meta-analytic techniques you might employ, but they all share the same fundamental approach: comparing the findings from the available studies with the findings that would be expected if the effect (here, delay or prevention of a second heart attack) did not, in fact, exist.

At present, the most widely used technique is the Mantel-Haenszel-Peto method. Generally employed to evaluate clinical trials, the method assumes experiments addressing similar questions should—except for the play of chance—yield answers that point in the same qualitative direction, regardless of the precise population addressed by any individual study.

To employ the Mantel-Haenszel-Peto method, a 2 × 2 table is constructed from each included study. In the heart attack example, the table consists of the numbers of participants in the experimental group and the control group who experienced, or did not experience, another heart attack.

The figures are divided into the observed number (O) in the

	Treatment group	Control group
Suffered attack	a	b
No heart attack	c	d

experimental group with the outcome (a heart attack), and the expected number (E), which is the number of heart attacks that one would have expected if the treatment had no effect. Clearly, O is equal to a, but the expected number E is $(a + b)(a + c)/N$, where N is the total population in both treatment and control groups. (E is not simply set equal to b, because one must take into account disparities in the sizes of the two groups.) The difference $(O - E)$ is then figured for each trial. This procedure is repeated for all i trials.

If the treatment has no effect, the difference $(O - E)$ should differ only randomly from zero. Hence the grand total:

$$T = \Sigma_i(O_i - E_i)$$

should differ only randomly from zero, and as N approaches infinity, T should approach zero asymptotically. As a consequence, a non-zero T is a strong indication that the treatment has some effect. The odds ratio (exp $[T/V]$, where V is the sum of the individual variances) provides a simple estimate of the validity of the non-null hypothesis, with rough 95% confidence limits being given by $\exp(T/V \pm 1.96/S)$, where S is the number of standard deviations by which T differs from zero.

Such meta-analyses are often presented visually by charting the odds ratios and their associated confidence intervals. The example below, taken from a 1988 overview of 25 clinical trials of vascular disease and antiplatelet agents*, such as aspirin, is typical of the genre.

The closer the assembled trials cluster around 1 (no effect), the less likely is any effect. Here, a vertical dashed line goes through every trial and its concomitant error bar, demonstrating the power of meta-analysis to find statistical agreement in what looks like great disagreement. ■ C.M.

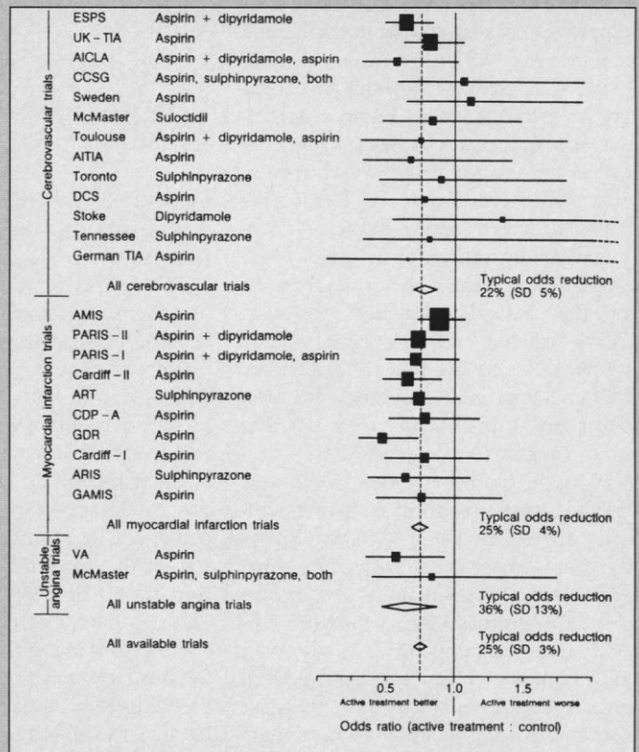


FIG 3—Odds ratios (active treatment:control) for first stroke, myocardial infarction, or vascular death during scheduled treatment period in completed antiplatelet trials. —■—=Trial results and 99% confidence intervals (area of ■ proportional to amount of information contributed). ◇=Overview results and 95% confidence intervals. Dashed vertical line represents odds ratio of 0·75 suggested by overview of all trial results. Solid vertical line represents odds ratio of unity (no treatment effect).

*"Secondary prevention of vascular disease by prolonged antiplatelet treatment," *British Medical Journal* **296**, 320 (30 January 1988).

far been exploited with much frequency. The National Acid Precipitation Assessment Program (NAPAP), a 10-year, half-billion-dollar interagency research program created in 1980 and due to publish its final report late this year, has spent much effort aggregating data on the causes and effects of acid rain. Despite the presumed importance of NAPAP's recommendations, meta-analysis wasn't considered. Instead, researchers were asked to "estimate (using expert judgment) the quality of the information base" on a scale of zero to four stars—as if they were reviewing restaurants.

"Is a famous scientist's estimate of one star worth the same as a young guy's three stars?" asks one obvously skeptical NAPAP participant. "What if one expert ranks a hypothesis with two stars, but another ranks the opposite hypothesis with two stars? Should you rank the net at zero? How do you combine a one-star estimation and a four-star? Do they honestly expect Congress to think this means anything?"

Formal, mathematical efforts to overcome the problem of subjective analyses by combining different experiments date back to the work of English geneticist Sir Ronald A. Fisher in the 1920s. But meta-analysis was first employed on a large scale in the United States in the early 1970s, when social scientists tried to assess New Frontier and Great Society programs. Richard Light, a statistician at the Harvard Graduate School of Education and the Kennedy School of Government, who with David B. Pillemer, a psychologist at Wellesley College, wrote a well-regarded introduction to the field of meta-analysis, recalls that the original assessments often "sharply conflicted, one with another"—baffling scientists and policymakers alike. "There was and still is a terrific need to pull this stuff together," he says. "Look at the dispute today over whether Head Start is effective."

By 1976, the quest for statistical methods for "pulling stuff together" had met with enough success that Gene V. Glass, a psychologist now at the University of Arizona, coined the term "meta-analysis" to describe the process of synthesizing results from separate but similar experiments.

But these new techniques weren't adopted instantaneously. Indeed, at times it must have seemed to the fledgling meta-analysts that no one was listening. As they fumed, "the oat bran syndrome" was repeated over and over again. In that syndrome, a small experiment finds a promising effect, the scientists involved appear on "Nightline," an entrepreneurial industry is created to take advantage of the supposed findings, and then, months later, a second, contradictory study is splashed on the front pages. "It's appalling how many times this has happened in medical research," says Richard Peto, head of the Cancer Studies Unit at Oxford and meta-analysis's most prominent exponent (see article on facing page). "Good treatments are ignored; useless treatments are disseminated—and much of it is because people have not properly analyzed data that have already been gathered."

One reason for the lag in adopting the method is that it is deeply ingrained in all who use statistics that you can't compare apples and oranges—that data from different studies cannot be pooled. "There must be more than a dozen studies of the effects of TV on children," says Light of Harvard. "Each one was done with a different protocol, with different sets of kids, and with different definitions. You simply can't throw together all of them together."

> *"[Meta-analysis] is going to revolutionize how the sciences, especially medicine, handle data. And it's going to be the way many arguments will be ended."*
>
> —Thomas Chalmers

Meta-analysis, its proponents explain, does not throw together experiments. Rather, it groups many individual studies and uses them, collectively, to compare what has been observed with the null hypothesis: the hypothesis that the effect sought in an experiment is, in fact, absent. Take, for example, the question of whether watching television has an effect on children's behavior. In a meta-analysis all the various studies done on that question would be gathered and compared, one at a time, with the null hypothesis—in this case the hypothesis that television has no effect on behavior.

If the null hypothesis is true, the series of comparisons in the meta-analysis should differ only randomly from a zero effect. Adding them together should give a result near zero, because the chance results will cancel each other out. But if the experiments consistently observe something new, such as an increase in violent acts, the comparisons should add up quickly, providing a sharp contrast to the null hypothesis (see article on p. 477). The great virtue of this method is that clear-cut results can emerge from a group of studies whose findings initially seemed to be scattered all over the map. Yet the technique is hardly without pitfalls, many of which stem from the fact that meta-analysis itself is not an experiment.

"Fundamentally, meta-analysis is observational in nature," says Richard Kronmal, a biostatistician at the University of Washington in Seattle. "It is subject to all the pitfalls of observational studies." An observational study (in contrast to an experiment, where conditions are manipulated to bring a certain phenomenon to light), must accept what is there, regardless of its quality. In the case of meta-analysis, "what is there" consists of studies done by other investigators.

"You get good studies mixed in with bad studies," Kronmal says. "You get studies with missing data or confused definitions. How much weight to assign each one is not easy to decide, and you never know if you've got all the studies." This observational nature raises a host of specific problems.

For one thing, to obtain all available information, meta-analysts must stringently search for unpublished experiments. It is widely believed—though exact proof remains elusive—that tests with negative results are much less likely to be submitted for publication, or, even if submitted, to make it into print. Hence overlooking the unpublished material may lead to a bias in favor of positive results. Fortunately for meta-analysts, that problem diminishes as the number of included experiments rises. In one review of 345 studies, Rosenthal and a collaborator calculated that 65,123 similarly sized but unpublished studies would have to exist to overturn their conclusions.

But even if the bias toward positive results is licked, aspiring meta-analysts must worry about whether they are averaging the results of poorly and excellently conducted studies. On the other hand, as any researcher knows, experiments with imperfect protocols can accurately reflect the real world, and even impeccably conducted experiments can go awry. Thus rejecting "bad" studies risks taking a biased slice of the universe of data.

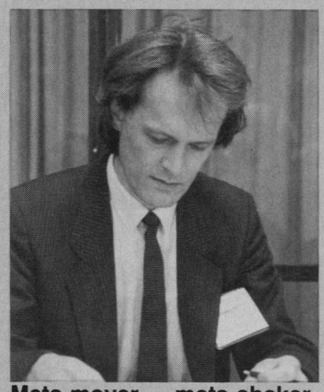
For clinical trials, the most prominent use of meta-analysis, Peto believes such worries can be minimized by controlling the individual experiments' selection bias, which is the introduction of bias into the selection of the group under study. He insists that only properly randomized trials can be put into a meta-analysis and then focuses on what is called an "intention-to-treat analysis"—that is, including all of those who are randomized in the analysis, regardless of whether it is believed they complied with the experimental regimen.

Most statisticians agree that controlling selection bias in the experiments under review is essential. But it does nothing to address the arguments of what Lawrence

Richard Peto: Statistician with a Mission

The most prominent exponent of meta-analysis is rail-thin, with angular features and a sheaf of straw-colored hair. Despite not having a doctorate, Richard Peto, director of the Cancer Studies Unit at Oxford University, is helping to push medicine toward a reevaluation of its principal research tool: the clinical trial. Peto, says Charles Hennekens of the Harvard School of Public Health, "is the Mozart of the clinical trial. The metaphor is rather exact, both to the quality and quantity of his output, and his willingness to [Hennekens laughs] suffer fools gladly."

Peto completed an undergraduate degree in mathematics at Cambridge amidst the turmoil of the Vietnam era. Unsure whether he wanted to spend his life with something as apparently irrelevant as mathematics, he spoke to Richard Doll. Doll was then perhaps the world's most renowned epidemiologist, although Peto didn't know that at the time. Doll convinced the disaffected younger man that biostatistics was a way to use his skills for something humanly worthwhile.



Meta-mover, meta-shaker.
Richard Peto of Oxford.

In this back-door fashion, Peto joined the third generation of British statisticians who invented the rules for clinical trials. The first was Sir Austin Bradford Hill, whose work in the 1940s established the methodology for ethical experimentation involving human beings. Hill and Doll, his protégé, established the link between smoking and cancer. When Hill retired in 1964, Doll (who is now Sir Richard) took his place. When Doll was appointed Regius Professor of Medicine at Oxford, Peto followed, his lack of a Ph.D. forcing Doll to hire him in the guise of a computer programmer.

By the time Doll and Peto joined forces, medical research had become a huge, international affair. Peto surveyed the stream of data crossing his desk and was dismayed. "Nothing ever produced any important results," Peto says. "Eventually I started wondering why that was. And the reason was that trials were too complicated and, above all, too small."

A basic problem was that cures had been found for most of the easy diseases—only the tough ones were left. Cancer, heart disease, and the like have multiple causes and are unlikely to have a single cure. Therefore, Peto reasoned, doctors today should be looking for modest, incremental improvements. "Unfortunately," he says, "almost nothing that was being done put you in a position to observe those benefits reliably."

Suppose, for example, that a new treatment can reduce the occurrence of heart attack by 20%. Given that heart attacks kill half a million people every year in the United States, that one-fifth reduction would stop thousands of premature deaths. Peto asked: How big would a trial have to be to observe such a reduction reliably? The answer is . . . VERY BIG. To observe a 20% reduction in 1 year, a randomized, placebo-controlled clinical trial would have to enroll more than 200,000 subjects. "It's perfectly clear," Peto says, "that these trials that appear in the newspaper, where they give something to a few dozen cancer patients and announce how many got better—90% of them tell you nothing and are a waste of time."

One way out, Peto concluded, was meta-analysis: a means of adding together similar trials to provide a large enough sample for statistical reliability. Having finally got a substantive university post in 1975, Peto and his collaborators laid out rules for such a statistical summing up. As in the case of many reformers, there was an impatient edge to Peto's presentation, and his blunt style gave him the reputation of an enfant terrible. "Doctors are used to being surrounded by adoring admirers," explains William K. Hass, a New York stroke specialist who has worked with him. "And when Richard gets into his Newtonian mode—well, it gets their attention, but they don't always like it."

"Peto has unquestionably made major contributions," says Kay Dickersin of the University of Maryland. "But there are also many people who think he has exaggerated his claims for meta-analysis and unfairly minimized its problems. But I don't think you'll find many statisticians who will object to the other part of his message, which is his insistence that trials are too small."

If doctors wanted to learn which treatments give the modest benefits possible for complicated diseases such as cancer, Peto said, they were going to have to learn how to run big, simple trials. "People thought, if you do a clinical trial, you do it precisely, and you collect an enormous amount of information on every patient. But you don't have to. Really, what's involved? One-half gets A, one-half gets B, and then you count the dead bodies."

Despite Peto's capacity to irritate doctors, he has managed to get this idea of large, simple trials into practice. In 1978, he, Doll, and ten collaborators managed to convince more than 5000 British M.D.'s to participate in the biggest prospective clinical trial ever undertaken in the United Kingdom. It was undertaken to learn the effects of aspirin on heart attack, which half a dozen smaller studies had failed to establish. The results were inconclusive, largely because there aren't enough M.D.'s in Britain to achieve statistical reliability. To obtain better data, Peto and Doll assisted Hennekens in forming the Physicians' Health Study in the United States. Published in January 1988, this study of 22,000 doctors found conclusive evidence of aspirin's benefits for heart attack.

Meanwhile, Peto was pushing ahead with more detailed studies of heart disease. To that end, he began assembling the Antiplatelet Trialists Collaboration in an effort to collect all trials on aspirin and other agents with similar mechanisms of action. By March 1990, when the Antiplatelet Trialists Collaboration had its second main meeting, Peto's conception had grown to include 207 trials.

Coordinating that much data is a Herculean job. Peto appeared gaunt as he addressed the gathering. But he radiated elation: meta-analysis had worked. Meta-analytical techniques had allowed his team to group the separate experiments into what amounted to a single vast clinical trial. Statistically, the results were sound; the probabilities that effects were due to chance were vanishingly small.

"The marvelous thing," he says, "is that we now have a possibility of putting a big dent in vascular disease, the major cause of death in the well-doctored world. When you think that this has been missed for all these years—well, I think we're going to see some real changes in behavior."

■ **C.M.**

Hedges of the University of Chicago calls "random effects modelers," who say that Peto's methods implicitly presume that the experimental interventions are equal. For example, comparing studies of drugs to surgery for cancer inherently assumes that the surgical procedures do not vary importantly, an assumption that Hedges says "may be disagreed with."

Proponents of meta-analysis acknowledge these problems, but argue that they are less significant than they might seem. Says Ingram Olkin, a statistician at Stanford who co-wrote one of the earliest textbooks on meta-analysis, "They're secondary, or even tertiary, compared to the problems with traditional informal reviews. The other way of doing things is inexcusably unscientific." In studies of children and television, for example, one might worry whether one can compare tests on groups from different backgrounds or those exposed to different shows. But meta-analysts say that the effects of television on children would be unlikely to differ qualitatively from group to group—with middle-class kids, say, being stimulated by the box into delinquency, and the poor being nudged toward sainthood.

As for those who might worry that the definition of antisocial behavior might vary across studies, the meta-analyst's retort is that definitions should not differ vastly among trained observers. "I don't want to startle my friends in the physical sciences," Light says, laughing, "but social scientists would broadly agree that hitting, biting, scratching, and shouting imprecations are antisocial behavior."

A further concern is that meta-analyses will be used to close off clinical trials before definitive results are in. None of the statisticians contacted by *Science* could cite a case in which a well-conducted meta-analysis had produced incorrect or misleading results. Yet none were prepared to argue that if a meta-analysis of several small studies shows a particular effect clearly, it is a waste of time and money to prepare a large, conclusive experimental trial. "If a meta-analysis jumps to a conclusion based on a lot of poor studies, then is it unethical to do a further study?" asks Kronmal. "I know of people who have refused to put their data into a meta-analysis for just that reason—they're afraid it will close off a subject prematurely."

A disturbing sign, some critics say, was the willingness of the National Cancer Institute to issue a Clinical Alert in May 1988 based on a meta-analysis of four unpublished studies of cytotoxic chemotherapy in premenopausal women with breast cancer. "The studies were unpublished and therefore un-peer-reviewed," Dickersin points out. "Yet they were trying to change the way every doctor in the nation treated breast cancer." Although the advice was backed 7 months later by another, bigger meta-analysis from Peto's group, the incident alarmed statisticians. "There was a lot of crabbing about it," Dickersin says. "And justifiably so—this is not the way to use statistics."

In spite of these potential hazards, meta-analysis clearly fills a critical need in science: the need to reconcile conflicting research results. "In some of the physical sciences," says Olkin, "you can replicate experiments identically. But in many fields, you can only repeat them, which always introduces some variation." Variation produces uncertainty, and meta-analysis is one way of dealing with that uncertainty.

> *"[Meta-analysis is] a boon for policy-makers who find themselves faced with a mountain of conflicting studies . . . That's what everyone likes about it, and that's also what everyone is worried about."*
>
> —Kay Dickersin

As a result, the technique is rapidly gaining popularity. According to Dickersin, meta-analyses were published at a rate of one or two a year until the end of the 1970s. Now, she says, the rate has "taken off—there are hundreds of them."

Among all these hundreds of meta-analyses, some of the most dramatic results have come from Peto. His 1980 meta-analysis of the apparently contradictory clinical trials of aspirin and coronary disease helped reverse the then-dominant belief that the drug had little effect on vascular disease. A 1985 meta-analysis by Peto, Rory Collins of Oxford, Salim Yusuf of the U.S. National Heart, Blood, and Lung Institute, and four other U.S. epidemiologists of 33 trials of intravenous streptokinase for acute heart attack led to what Collins calls "an absolutely fundamental reappraisal of how cardiologists should approach this condition." Peto and his colleagues at Oxford are now coordinating a meta-analysis of 207 trials of antiplatelet drugs that includes a staggering 115,701 patients—easily the biggest drug test ever undertaken, and conceivably the biggest medical experiment ever performed.

But Peto hardly has the field to himself. Meta-analyses now appear in disciplines from marketing (to synthesize studies of advertising) to meteorology (to take an overview of more than 750 cloud-seeding experiments), and from education (to evaluate studies on subjects such as class size and coaching on Scholastic Aptitude Tests) to epidemiology (where, for instance, Thomas Chalmers is examining studies of the health effects of power lines). "It's absurd that it's not being used more," Chalmers says. "Somebody should use it to end this whole battle over asbestos, for instance."

Moreover, meta-analysis may—in the not-so-distant future—have a profound impact on the way that all trials are done. Olkin, Hedges, Thomas Chalmers, and Joseph Lau of the Boston Veterans Administration Medical Center are proposing what amounts to a national registry of experiments. As they are completed, they will be automatically added into a computerized meta-analysis, much as the Oxford Database for Perinatal Trials is currently doing in the field of obstetrics. When the computer signals that the aggregate data is approaching significance, a committee will decide whether further study is needed, or the case can be considered to be closed. Although Chalmers believes that a national registry is inevitable, he believes that getting the proposal funded will not be easy—no extant governmental agency is broad-based enough to handle it comfortably.

An example of the need for such a registry is provided by Iain Chalmers's retroactive meta-analysis of the effects of diethylstilbestrol (DES), a synthetic estrogen administered to prevent miscarriage. In the 1970s, DES was discovered to cause vaginal cancer in the offspring of women who took the drug. (The median age of diagnosis is an incredible 19; the usual treatment is radical surgery, with vaginectomy.) According to Chalmers, a meta-analysis of the trials completed by 1955 would have strongly militated against the continued administration of DES.

Between then and now little has changed in the eyes of statisticians; there is no particular reason to assume that another case like that of DES would not occur. "We might have enough data to answer a question sitting around for 5 years before somebody notices," Thomas Chalmers says. "In a society that generates as much scientific data as ours, it's absolutely foolish not to put it together properly. I can't believe that we'll really continue going on as we are."

■ CHARLES MANN

Charles Mann is a free-lance writer based in New York City.