

FROM HIERARCHY TO NETWORK

*a richer view of evidence for
evidence-based medicine*

ROBYN BLUHM

ABSTRACT Evidence-based medicine (EBM) advocates the improvement of patient care through the use of current best research evidence in medical decision making. In practice, “best evidence” generally refers to where a study fits on a hierarchy of evidence, which places randomized controlled trials (RCTs) and other population-level research above laboratory research. Because population research is concerned primarily with average results obtained from large groups of people, ranking evidence on the basis of its place in the hierarchy is shortsighted and ultimately limits the ability of research results to inform the care of individual patients. The history and methodology of epidemiology reveals a close relationship between population-level and laboratory research; both types of research are necessary if we are to understand the causes of a disease. What EBM does not take into account in its hierarchy of evidence is that the same thing is true for research on the safety and efficacy of medical interventions. To maximize the information that clinical research can provide for clinical care, RCTs should be designed to elucidate within-group variability. This can only be done if the hierarchy of evidence is replaced by a network that takes into account the relationship between epidemiological and laboratory research.

THE BASIC IDEA THAT GUIDES evidence-based medicine (EBM) is uncontroversial: medical decision making regarding the care of an individual patient should be informed by the results of relevant research. Putting this idea into practice, however, has proven to be difficult. Supporters of EBM have produced numerous resources, such as systematic reviews, meta-analyses, and guidelines,

Department of Philosophy, University of Western Ontario, London, Ontario, N6A 3K7, Canada.
E-mail: rbluhm@uwo.ca.

Perspectives in Biology and Medicine, volume 48, number 4 (autumn 2005):535–47
© 2005 by The Johns Hopkins University Press

which are meant to guide physicians in making patient care decisions. Basic to the production of all of these resources is the idea that there is a “hierarchy of evidence,” with the best evidence being derived from one or more randomized controlled trials (RCTs) of the treatment in question. RCTs are designed to show the average effects of a treatment in the population under study. Yet the history of epidemiology shows that progress in population-level research depends on advances in laboratory research; both types of study are required to elucidate the causes of disease. Moreover, because the goal of medical treatment is effective causal intervention in the course of a disease or in the manifestation of its symptoms, it is important to realize that a similar relationship exists between RCTs and laboratory research. Thus the idea of a hierarchy of evidence should be replaced by a “network” model that takes into account the relationship between evidence drawn from various sources, or types of studies.

THE HIERARCHY OF EVIDENCE

To begin with, the term *hierarchy of evidence* is a misnomer: the hierarchy is actually a hierarchy of methodologies. That is, it focuses not on the actual results of a particular study or group of studies—in other words, on the evidence they provide for the efficacy of a treatment—but on how that evidence was obtained. RCTs allow for randomization of subjects to study groups and thus also provide a simple method for blinding of both subjects and investigators. Because of these features, they are held to be less likely to provide biased results than are those studies that do not allow for randomization or in which blinding is not used. Thus, hidden in the term *hierarchy of evidence* is the not unreasonable assumption that there is a link between the methodology used in a study and the strength of the evidence that can be obtained from it (though this should not be taken to imply that the use of RCTs guarantees—or is the only way to acquire—good evidence).

In addition, it should be noted that there are different versions of the hierarchy. First, there are different hierarchies for different types of clinical questions (for example, for treatment studies, studies of risk or prognosis, or studies testing the utility of different clinical decision rules). Second, different versions of these hierarchies have different degrees of detail. Some place all nonrandomized (“observational”) studies on a single level of evidence; others subdivide this level, giving different methodological weight to cohort and to case-control studies. Still, all of the hierarchies follow the same basic structure—they place epidemiological studies at the top few levels. In this article, I will follow the version of the hierarchy given by *The Users’ Guide to the Medical Literature* (Evidence-Based Medicine Working Group 2002) and will be concerned solely with treatment studies.

Although EBM operates with a broad definition of evidence, on which “any empirical observation about the apparent relation between events constitutes potential evidence” (Evidence-Based Medicine Working Group, p. 6), it relies pri-

marily on a hierarchy of evidence built on two principles that aim to ensure that the best available evidence is given priority. First, the clinical observations should be systematic; in other words, they should follow a clear methodology that has been designed to minimize bias in the measurement of outcomes. On the hierarchy presented in the *Users' Guide*, unsystematic clinical observations (such as those derived from case studies or from accumulated individual experience) provide the lowest level of evidence. Second, the research should address questions that are directly relevant to patient outcomes. Studies that examine physiological outcome measures, such as blood pressure, cardiac output, or bone density, are ranked lower in the hierarchy (p. 7). I will use the terms *population-level* and *clinical* research interchangeably to refer to research that uses epidemiological methods, in contrast to *laboratory* or *bench* research, which investigates biological mechanisms in humans or animals. I recognize, though, that many epidemiological studies incorporate physiological measurements.

The highest levels on the hierarchy (which in practice are those that are considered in EBM) consist of study designs that provide systematic clinical observations. These studies tend to follow large groups of patients, for varying lengths of time, and to report average outcomes in each of the various patient groups. Because of this, RCTs and nonrandomized treatment studies provide only limited information about how a drug will work for an individual patient. In some cases, this is because the patient does not resemble the individuals in whom the drug was tested; she may have comorbid diseases or be taking other medications that would have disqualified her from participating in the RCT in question. In other cases, the patient may well have had similar enough characteristics to the study patients that she or he could have participated in the study; it is often the case, however, that not all study patients benefit from a drug that has been shown to be beneficial in most cases. Because RCTs tend to report only average results in the treatment and the control groups, the extent and sources of within group variability are not known. Both extrapolation of the results of an RCT to other patient groups and an understanding of the reasons for differences in outcomes within the study group require a knowledge of biological factors that may influence the effectiveness of a drug. This type of information, however, cannot come from epidemiological studies alone. Rather, it is often first discovered in the context of physiological studies on humans or animals (the second lowest level of evidence in the hierarchy) and of unstructured clinical observation (the lowest level). This suggests that the relationship between the highest and the lowest types of research in the hierarchy is much closer than usually acknowledged. EBM urges, however, that whenever possible, the evidence used in clinical decision making should come from as high in the hierarchy as possible. This approach neglects the important contribution of biological research to the interpretation of epidemiological research, including RCTs.

I do not wish to diminish the importance of large-scale studies that examine clinically relevant outcomes; nonetheless, after decades in which medical re-

search *meant* laboratory research, the pendulum may be swinging too far the other way. David Sackett (2000), for example, has written about the movement from “clinical research” to “clinical-practice research.” He notes that since the 1960s and 1970s, the skills required for medical practice in a clinical setting have become increasingly remote from those required to conduct laboratory research. At the same time, fewer medical students have chosen careers that incorporate both laboratory research and clinical medicine; the former is now carried out primarily by Ph.D. researchers. Sackett further argues that the increase in the number of physicians involved in what he calls “clinical-practice research”—primarily, judging from his examples, RCTs, but also encompassing small, qualitative studies—has occurred largely because (in contrast to bench research) this research is more easily integrated with clinical practice. In another editorial, Sackett (1999) further contrasts bench research with clinical-practice research—to the benefit of the latter: “Millions of dollars’ worth of bench research that appeared to show a reduction in atherosclerosis-related oxidative stress with vitamin E therapy was recently tested in an RCT that asked the vital question: Does the vitamin E therapy endorsed by this research really help prevent heart disease? The answer was a resounding ‘No’” (p. 1414).

Yet while the clinical-practice research that Sackett advocates is important and does hold promise to improve health care (especially if, as Sackett also advocates, responsibility for conducting such research is removed from the pharmaceutical industry), his argument ignores the fact that bench research and clinical (epidemiological) research are intimately related. The history of epidemiology shows that advances in one of these aspects of biomedical research often depends on advances in the other; this point is particularly clear in the case of infectious diseases but is equally important for understanding chronic disease.

A BRIEF HISTORY OF EPIDEMIOLOGY

Epidemiology has been defined as “the study of the distribution and determinants of disease frequency” (MacMahon and Pugh 1970). What this amounts to is attempting to give some sort of causal account of how disease arises. Although interest in epidemiologic issues dates back to antiquity, epidemiology is generally considered to have come into being in the 19th century, with the science developing most rapidly in Britain. At this point in its history, the primary concerns of epidemiology were social ones: the rapid growth of urban industrial centers, combined with the dismal living conditions of the working class, soon began to give rise to new health problems.

A landmark in the history of epidemiology was the publication of the final report of the Poor Law Commission in 1842. This document was the culmination of three years of data gathering throughout England, Wales, and Scotland, in order to examine the health of the working class and propose recommendations for improvement. The report “is filled with vivid details of existing conditions,

and contains a serious effort, district by district, to correlate these conditions with variations in mortality rates and economic status” (Rosen 1993, p. 190).

The sophisticated links drawn between housing and sanitary conditions and the incidence of disease were not, however, matched by an understanding of the biological basis of disease. To the extent that “fever”—“the portmanteau term that included typhoid, typhus, and relapsing fevers” (Rosen 1993, p. 188)—was discussed in etiologic terms, it was generally attributed to the effects of environmental “miasmas.” According to the prevailing “‘filth’ theory of diseases, miasmatic hazes rising from decaying matter, rather than contagion and microorganisms, were supposed to cause epidemics” (Ackerknecht 1982, pp. 212–13).

With the development of bacteriology in the late 19th and early 20th centuries, the focus of epidemiologic research began to shift from “sanitary statistics” to the isolation of the etiologic agents of infectious diseases. Strategies of disease prevention, too, shifted from attempts to improve living conditions (in part, perhaps, because work in this area had already led to such great improvements) to research that aimed to clarify how these agents were transmitted to human hosts, whether from other human beings (in the case of contagious diseases) or from specific environmental sources. The developing science of microbiology enabled the isolation and culture of infectious agents.

The relationship between laboratory science and epidemiology during this era became closer, as both types of scientific investigation were required in order to elucidate the causes of a disease. The importance of this interdisciplinary cooperation was not fully realized in the early days of microbiology. Rather, it was thought that laboratory work, guided by Koch’s postulates, was sufficient. These postulates provide criteria that must be met by an infectious agent in order for it to be considered the cause of a disease:

1. The parasite occurs in every case of the disease in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. It occurs in no other disease as a fortuitous and nonpathogenic parasite.
3. After being fully isolated from the body and repeatedly grown in pure culture, it can induce the disease anew [in an animal model]. (Evans 1993, p. 30)

According to Alfred Evans (1993), even now “some persons believe that if the postulates are fully met, then the risk factor is the cause of that disease” (p. ix). However, Koch initially developed the postulates on the basis of his tuberculosis research, and their applicability in the case of other infectious agents can be questioned. At the very least, the postulates must be modified for numerous other pathogens: some (including those that cause leprosy, syphilis, and malaria) cannot be grown in pure culture; others cannot be reproduced in an animal model in order to study a disease that resembles the human disease. An even

greater challenge has been presented in the study of viral diseases, since they require a cellular medium in order to reproduce.

In addition to modifying Koch's postulates, Evans emphasizes that investigation of infectious diseases requires a combination of both laboratory and epidemiological research. Initial epidemiological work can "establish the source and method of transmission of the infecting agent" (p. 31). Laboratory research can then identify and characterize the infectious agent and, finally, a case-control study can establish that the putative agent is indeed present more frequently in a group of infected individuals than in a control group (pp. 31-32).

This last step, however, requires the use of epidemiologic techniques that were not available in Koch's time. In fact, Rothman and Greenland (1998) claim that that it is only since the middle of the 20th century that epidemiology has begun to develop "a systematized body of principles" (p. 38) and a set of accepted study designs with which to analyze the occurrence of disease. While this development occurred in part because of the recognition of the complexity of infectious diseases, it is likely to be due mainly to the increasing importance of chronic disease. With fewer people dying from infections that used to be almost inevitably fatal, the burden of disease has shifted. Increasingly, medical research is focused on chronic diseases, such as diabetes, asthma, heart failure, hypertension, inflammatory bowel disease, and arthritis. Hoffman, Rice, and Sung (1996) estimate that chronic diseases account for three-quarters of health care costs in the United States; it is primarily with these diseases that EBM is concerned. Whereas, at least initially, research into the causes of infectious diseases focused on the infectious agents, understanding the causes of chronic diseases requires investigation of complex physiological processes and of how these processes are altered by those factors that contribute to the development of a disease. Examination of the ways in which epidemiologists attempt to determine the causes of disease reveals that the "hierarchical" approach to research advocated by EBM does not reflect the more sophisticated methods of studying disease used in epidemiology.

CAUSAL ANALYSIS AND CONFOUNDING IN EPIDEMIOLOGIC RESEARCH

I have previously referred to MacMahon and Pugh's (1970) definition of epidemiology as "the study of the distribution and determinants of disease frequency." These two topics are related, respectively, to two central concepts in epidemiology: incidence and exposure. An incident is the occurrence of a new case of the disease under study. It is generally described in terms of a particular population and limited to a particular time. Thus, epidemiologists speak of "incidence rates," or the number of new cases of a disease in a population in a given period of time (usually years). Incidence rates have been called the "basic building blocks for epidemiologic inferences" (Rothman and Greenland 1998, p. 4). A major task of epidemiology is to determine incidence rates; this is not easy, as

considerable time and effort may be required to make these measurements. The determination of incidence rates fulfills the first goal of epidemiology: incidence rates characterize the distribution of disease in a population. (The distribution of disease is also characterized by prevalence rates, which reflect the total number of cases of a disease in a population at a given time. Together, prevalence rates and incidence rates give information about the cure rate or the mortality rate of a disease.)

Incidence rates, however, do not give any indication of the causes of disease. This brings us to the second central concept in epidemiology, that of exposure. In epidemiologic terms, an exposure is any factor that may play a causal role in the development of a disease. These include not only exposures in a layperson's sense of the term—for example, exposures to infectious disease agents or carcinogens—but also behaviors (smoking, exercise), interventions (educational programs, medical therapies) and genetic traits. Isolation of factors that are truly exposures is the second goal of epidemiology, since exposures are determinants of disease. Thus, according to Rothman and Greenland (1998), determining the incidence rate of a disease in a population “is just the first step in accumulating epidemiologic knowledge” (p. 4). The ultimate goal of the epidemiologist is to establish the causes of disease.

This goal is rendered vastly more difficult by the fact that epidemiology is not generally an experimental science. The epidemiologist must try to identify causal factors “in the wild,” rather than in the controlled environment of a laboratory. Unlike a scientist engaged in animal research in the lab, an epidemiologist cannot expose an experimental group to a putative exposure and compare outcomes to those in a control group, and she cannot regulate the other exposures with which the members of each group come into contact. Further, many epidemiologic studies must work backwards, from known instances of disease to their causes, rather than from known exposure to disease.

These difficulties have spurred the development of the modern epidemiologic methods referred to by Rothman and Greenland. In the most common type of epidemiological study, a cohort study, scientists divide a population into two (or more) groups, according to whether members have a disease of interest, and then compare the groups. If, for example, the ill group is (on average) older, or less wealthy, or contains a higher proportion of women than the control group, then age, lower socioeconomic status, or being female can be viewed as a risk factor that possibly predisposes individuals to a disease. It is important to note, though, that identifying risk factors should not be seen as a goal in itself (except to the extent that it contributes to preventive efforts, such as increasing workplace safety by removing toxic materials, or encouraging healthy lifestyle choices); rather, it is best seen as a preliminary search for causal factors. As Rothman (2002) points out, age, socioeconomic status, and gender do not themselves *cause* disease; rather, they are characteristics possessed by an individual that may *appear* to be causal because they are statistically associated with the disease of interest.

This type of misleading relationship is known as confounding; by controlling for confounding variables, investigators can distinguish between statistical risk factors identified in early cohort studies and factors that are truly causal.

Rothman illustrates the issue of confounding by using Stark and Mantel's (1966) work linking birth order and Down syndrome. In the 1960s, these researchers found a strong trend toward increased prevalence of Down syndrome as birth order increases. Yet this trend does not indicate a causal factor; being the fourth- or fifth- born in a family does not *cause* Down syndrome. Rothman (2002) notes that: "The effect of birth order . . . is a blend of whatever effect birth order has by itself and the effect of another variable that is closely correlated with birth order" (pp. 101–2). It turns out that this other variable is maternal age: because birth order and maternal age are highly correlated, mothers giving birth to their fourth or fifth baby are older, on average, than those giving birth to their first or second. The apparent trend with birth order was found to be entirely due to confounding by maternal age.

Yet the effect of maternal age is itself a confound. Since age itself cannot cause anything, Rothman notes that it simply serves as "a proxy for as yet unidentified events that more directly account for the occurrence of Down syndrome. When these events are identified, we would ultimately find that mother's age has no effect once we take into account the biological changes that are correlated with age, which is to say that the apparent effect of mother's age is also confounded by factors as yet unknown" (p. 105).

Traditionally, these factors have been sought by studying physiology. Recently, however, advances in genetics and molecular biology have offered the promise of a new level at which to identify causal factors in disease. It is now possible, in some cases, to identify individuals with and without a genetic marker of interest and to track the incidence rates of a disease in each group. Similarly, biochemical and physiological measurements can monitor changes in people who have been exposed to a risk factor of interest and thus track the development of disease. Thus, as Rothman points out: "As the layers of confounding are left behind, we gradually approach a deeper causal understanding of the underlying biology" (p. 105).

Understanding the underlying biology, however, requires both population-level studies (Sackett's "clinical practice research") and laboratory studies that elucidate the biological mechanisms underlying disease. As Evans (1993) pointed out in the case of infectious diseases, both levels of analysis are necessary; the same is true in the case of the chronic diseases that are currently the focus of much medical research.

FROM HIERARCHY TO NETWORK

It could be argued that, even though understanding biological mechanisms may be necessary in order to understand the causes of a disease, it is still the case that

epidemiological studies (generally RCTs) are sufficient to guide medical decision making. Particularly in the case of pharmaceutical treatments, the argument might run, preclinical research is sufficient to explain how a drug acts at a physiological level. What is important is to conduct RCTs that determine whether the drug *works* to cure or treat disease. This is the line of reasoning implicit in Sackett's arguments cited earlier.

What this argument ignores, however, is that the purpose of medicine is effective causal intervention in the course of a disease in an individual patient, and that the accurate estimation of a drug's effects in a broad population does not go far in achieving this goal. Whether a drug is prescribed to prevent, cure, or slow the course of a disease, or to manage symptoms, it effectively becomes another causal factor in the set of causal factors (or exposures) that have already contributed to the patient's state of health. And perhaps the chief lesson of modern epidemiology is that the effects of a single causal factor operating at a population level may be very different in different individuals. Just as not all smokers get lung cancer and some individuals with a genetic predisposition to a disease live long, healthy lives, it is also the case that some patients fail to benefit from, or are even harmed by, exposure to a drug that has been shown in an RCT to be, on average, beneficial. By determining which factors in a population are regularly associated with health or disease, epidemiological studies aim to uncover the various causal stories linking exposures with health outcomes. RCTs should aim to do so as well.

As mentioned earlier, however, this is not how RCTs are generally conducted. Instead, they tend to report average results for each study group, at least for the types of continuous variables that are the outcomes measured in many trials. Moreover, it is increasingly the case that between-group differences in these outcomes are statistically small (though clinically important). In an article discussing the implications of the need to detect smaller differences in treatment outcomes, Peto and Baigent (1998) offer suggestions for how best to conduct RCTs. Most importantly, they say, clinical trials must become bigger: "if one is trying to decide how many millions of future patients should be treated, it may often be appropriate to randomize at least as many thousands . . . or even tens of thousands" (p. 1170).

It might appear that by conducting such huge trials, we would be in a position to gather a great deal of information, not only about the average effects of the drug, but also about the variability in its effects in different subgroups of the patients in the study. However, Peto and Baigent disagree with this approach. They suggest that in order to enroll the numbers of patients that RCTs now increasingly require, trials must be "extremely simple and flexible . . . simplify the entry criteria, simplify the treatments and simplify enormously the data requirements." They further claim that many studies collect "ten or a hundred times too much information per patient" and worry that this may make clinical research impracticable, arguing that "those who sponsor, perform and regulate therapeutic research need to find ways of making trials much simpler and much larger."

In my view, Peto and Baigent are traveling in exactly the wrong direction. After all, “moderate benefits,” as shown in the average result in an experimental group in an RCT, might be seen because individual patient results fall symmetrically and normally around the mean (as Peto and Baigent appear to assume) or because some subgroups of the study population fare much better or much worse than the average. Without paying attention to the possibility of within-group variation, we will not know whether the observed results are due to one or the other of these extremes (or to some other type of distribution). And the only way to measure within-group variation is to gather the information from each patient that may mark them as a member of a relevant subgroup and then conduct analyses of these data. Further, the recommendation that the entry criteria be simplified and broadened only exacerbates the problem of within-group variability. While clinical trials should enroll a broader range of patients than is currently common, there is no point in doing so unless we also attempt to determine whether (and how) the inclusion of patients with different characteristics will affect the results obtained from the study.

One simple way of increasing the amount of information that can be gained from RCTs is to regularly conduct subgroup analyses. Current EBM orthodoxy cautions against conducting such analyses, or at least suggests that there should be only a few such analyses conducted and that those should be planned a priori. The rationale for this is that minimizing subgroup analyses also minimizes the problem of finding supposedly significant results that are actually due to chance. The argument then goes that the more comparisons being conducted, the more likely some of the significant results obtained will be due to chance. While this sort of false positive result does occur, avoiding them is neither as important nor as difficult as the standard approach would suggest. First, there is a trade-off in statistics between the risk of false positive errors and the risk of false negatives. In the case of medical research, we might well want to accept a greater risk of false positives; as Kristin Shrader-Frechette (1994) has argued, there is actually an ethical imperative in applied research (including medical research) to decrease the risk of false negatives. While Shrader-Frechette is concerned primarily with research on the environmental impact of certain practices, this advice is even more appropriate for medicine, where (as the recent controversy over the adverse effects associated with Vioxx[®] shows) downplaying relatively weak trends in outcome data may literally have life-or-death consequences. Thus, when the putative differences between subgroups have a potential effect on patient safety (as when a particular subgroup appears more likely to suffer adverse effects from a treatment regimen), it may well be better to err on the side of caution and take the effect seriously. Second, even if a small number of the subgroup analyses conducted in a study do give apparently significant results that are actually due to chance, the probability of finding the same chance finding twice when conducting the same (or a similar) study greatly decreases. Thus, consistently running subgroup analyses in all clinical trials will provide real

information about within-group variation, and about the risk factors (or exposures) associated with that variation.

The ultimate aim, however, should be to link the results of clinical research with the underlying biological mechanisms that cause them. This brings us back to the question of the relationship between risk factors and causes: risk factors can be sought through subgroup analyses and the causal story can then be filled in through a combination of epidemiological and laboratory research. Doing so both provides additional evidence for statistically weak associations and contributes to the goal of effective treatment by elucidating the effects of medical intervention. Yet EBM tends to avoid discussion of the biological processes that give rise to observed statistical results. In fact, it generally advises against basing judgments on the “biological plausibility” of a finding, since an ad hoc “just so story” can be given for any finding. The fact that such stories are possible, however, does not imply that all appeals to biological plausibility are equally substantiated. We do have a great deal of knowledge about the physiological and molecular mechanisms that underlie disease, and that increased knowledge will ultimately separate the good from the bad appeals to “plausibility.”

Again, both conducting subgroup analyses and combining epidemiological and laboratory research are necessary if we are to make real progress in understanding the causes of disease—and the effects of causal intervention in disease treatment. This picture of medical research is, however, highly idealized. In particular, it could be argued that the specification of causes in medicine is largely a pragmatic matter. Jonathan Rees (2002) has argued that the cause of disease should not be viewed as “some objective God’s eye summary of pathophysiology, but rather an operational statement of where we think the Achilles’ heel of a disease might be” (p. 699). He notes that pernicious anemia is a highly complex disease, but that given our ability to successfully treat the disease with vitamin B₁₂ injections, we are justified in operationally defining B₁₂ deficiency as the cause of the disease.

Rees raises an important point: we need not know the complete “God’s eye” story in order to treat disease—and indeed cannot afford to wait for it. Successful intervention *can* occur with only a sketchy knowledge of disease mechanisms: insulin was saving lives long before the details of the pathophysiology of diabetes were understood. It was only with further research, however, that type 1 and type 2 diabetes were distinguished, with the result that non-insulin-based treatments could be developed for the latter and the unpleasant consequences of long-term insulin use could be reduced through the development of better strategies for insulin therapy. This example suggests that the identification of the Achilles’ heel of a disease and successful intervention at that point are not always as simple as in the case of pernicious anemia, and that our operational definition of the cause of a disease may well change over time. Vitamin B₁₂ injections are an effective and relatively benign form of treatment; in cases where the treatments are not entirely satisfactory or may themselves have adverse effects, it is best not to be

too easily satisfied by our identification of a locus for successful intervention. (To do him justice, Rees is not arguing that physiological research into the course of disease is not necessary, only that it is not sufficient to guarantee advances in clinical research; in a sense, his argument and mine are complementary.)

Finally, as the example of diabetes shows, treatment advances may come in one of two forms: they may involve the development of new treatments or the refinement of old ones. In the first case, advances in genetics and molecular biology are often cited as having great potential to provide a new understanding of the cause of some diseases and therefore to open new possibilities for treatment. Already, cancer therapies have been developed that target specific oncogenes implicated in certain subtypes of tumor, including the development of Herceptin® for the treatment of HER-2 positive metastatic breast cancer (Blay et al. 2005). These cases offer hope that identification of further patient differences at both the physiological and the molecular levels may well result in the development of new treatments targeted at specific subgroups of patients. It is equally important, however, to remember that individualized treatment does not necessarily require the development of new drugs; similar research might well show that different patient groups benefit from different doses or combinations of drugs, or have different risks of certain adverse effects. Here again, these differences might be traced to genetics, or they may be due to the effects of comorbid diseases, drug interactions, or environmental or lifestyle factors.

Although it is possible for improved understanding of biological mechanisms to result in better treatment, it is necessary for clinical research to start taking an active role in pursuing this goal. Unfortunately, EBM's mistrust of biological explanations, insofar as it affects the way in which clinical trials are conducted and analyzed, may become a self-fulfilling prophecy. Only by making consistent efforts to link the results of clinical trials to what is known about the biological processes affected by the drug can clinical practice research contribute to tailoring treatments to individual patients.

CONCLUSION

Extrapolation from the success of a drug in an RCT to the utility and safety of the drug in the case of a particular patient is not always straightforward. By enrolling a broader range of patients in RCTs and paying attention to within-group variability in outcomes, RCTs will ultimately provide better evidence for the care of individual patients. While the use of a broader study population may mean that it is more difficult to arrive at the precise outcomes estimates, this is not a great loss because these estimates tell us relatively little about the actual effects of a drug in the case of an individual patient. The hierarchy of evidence, with its focus on RCTs and meta-analyses, assumes that these precise estimates are the best evidence to inform medical care; a better approach to using the results of medical research in clinical practice is one that takes seriously the

methods that have been identified in epidemiology for identifying the determinants of disease—and extends these methods to understanding the amelioration of disease as well. The hierarchy of evidence should thus be replaced by a network in which the results of studies from different disciplines, using different methodologies, must be considered together to arrive at an assessment of current best research evidence. On this network view, secondary resources, such as the systematic reviews and meta-analyses developed by practitioners of EBM, would take into account not only the between-group differences observed in RCTs, but also within-group differences and physiological studies that elucidate the biological differences that may account for outcome differences. Given the success that EBM and related enterprises have had in improving the reporting standards of clinical trials and in assembling multi-disciplinary groups to compile secondary resources for clinicians, the next challenge should be to develop a broader evidence base that reflects the nature of the research that provides that evidence.

REFERENCES

- Ackerknecht, E. H. 1982. *A short history of medicine*, rev. ed. Baltimore: Johns Hopkins Univ. Press.
- Blay, J. Y., et al. 2005. Targeted cancer therapies. *Bull Cancer* 92(2):E13–18.
- Evans, A. S. 1993. *Causality and disease: A chronological journey*. New York: Plenum.
- Evidence-Based Medicine Working Group. 2002. *Users' guide to the medical literature: A manual for evidence-based clinical practice*. Chicago: AMA Press.
- Hoffman, C., D. Rice, and H. Y. Sung. 1996. Persons with chronic conditions: Their prevalence and costs. *JAMA* 276(18):1473–79.
- MacMahon, B., and T. F. Pugh. 1970. *Epidemiology: Principles and methods*. Boston: Little, Brown.
- Peto, R., and C. Baigent. 1998. Trials: The next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 317(7167):1170–71.
- Rees, J. 2002. Complex disease and the new clinical sciences. *Science* 296:698–701.
- Rosen, G. 1993. *A history of public health*, expanded ed. Baltimore: Johns Hopkins Univ. Press.
- Rothman, K. J. 2002. *Epidemiology: An introduction*. New York: Oxford Univ. Press.
- Rothman, K. J., and S. Greenland. 1998. *Modern epidemiology*. Philadelphia: Lippincott-Raven.
- Sackett, D. L. 1999. Time to put the Canadian Institutes of Health Research on trial. *CMA J* 161(11):1414–15.
- Sackett, D. L. 2000. The fall of “clinical research” and the rise of “clinical-practice research.” *Clin Invest Med* 23(6):379–81.
- Shrader-Frechette, K. 1994. *Ethics of scientific research*. Lanham, MD: Rowman & Littlefield.
- Stark, C. R., and N. Mantel. 1966. Effects of maternal age and birth order on the risk of mongolism and leukemia. *J Natl Cancer Inst* 37:687–98.