EPIDEMIOLOGIC EVIDENCE IN PUBLIC AND LEGAL POLICY: REALITY OR METAPHOR?

by
Gio Batta Gori, ScD, MPH

Washington Legal Foundation
Critical Legal Issues
Working Paper Series No. 124
June 2004

TABLE OF CONTENTS

ABOUT WLF'S LEGAL STUDIES DIVISION

The Washington Legal Foundation (WLF) established its Legal Studies Division to address cutting-edge legal issues by producing and distributing substantive, credible publications targeted at educating policy makers, the media, and other key legal policy outlets.

Washington is full of policy centers of one stripe or another. But WLF's Legal Studies Division has deliberately adopted a unique approach that sets it apart from other organizations.

First, the Division deals almost exclusively with legal policy questions as they relate to the principles of free enterprise, legal and judicial restraint, and America's economic and national security.

Second, its publications focus on a highly select legal policy-making audience. Legal Studies aggressively markets its publications to federal and state judges and their clerks; members of the United States Congress and their legal staffs; government attorneys; business leaders and corporate general counsel; law school professors and students; influential legal journalists; and major print and media commentators.

Third, Legal Studies possesses the flexibility and credibility to involve talented individuals from all walks of life — from law students and professors to sitting federal judges and senior partners in established law firms — in its work.

The key to WLF's Legal Studies publications is the timely production of a variety of readable and challenging commentaries with a distinctly common-sense viewpoint rarely reflected in academic law reviews or specialized legal trade journals. The publication formats include the provocative COUNSEL'S ADVISORY, topical LEGAL OPINION LETTERS, concise LEGAL BACKGROUNDERS on emerging issues, in-depth WORKING PAPERS, useful and practical CONTEMPORARY LEGAL NOTES, interactive CONVERSATIONS WITH, law review-length MONOGRAPHS, and occasional books.

WLF's LEGAL OPINION LETTERS and LEGAL BACKGROUNDERS appear on the LEXIS/NEXIS® online information service under the filename "WLF" or by visiting the Washington Legal Foundation's website at www.wlf.org. All WLF publications are also available to Members of Congress and their staffs through the Library of Congress' SCORPIO system.

To receive information about previous WLF publications, contact Glenn Lammi, Chief Counsel, Legal Studies Division, Washington Legal Foundation, 2009 Massachusetts Avenue, NW, Washington, D.C. 20036, (202) 588-0302. Material concerning WLF's other legal activities may be obtained by contacting Daniel J. Popeo, Chairman.

ABOUT THE AUTHOR

Gio Batta Gori is a long time defender of sound science as a foundation of fair public health policy. An Italian émigré biologists, he first was in the United States with the late Jonas Salk. After academic and industrial experiences, he contributed to national research and prevention agendas at the U.S. National Cancer Institute, where he was deputy director of the Division of Cancer Cause and Prevention, director of the Diet, Nutrition and Cancer Program, and director of the Smoking and Health Program. In 1976 he received the Public Health Service Superior Service Award. Later he directed the Policy Analysis Center of the Franklin Institute and since 1988 the Health Policy Center of Bethesda, Maryland, a study and advisory group in health policy and regulation.

After an education in the classics, Dr. Gori obtained a doctorate in biology in Italy and a Master of Public Health in epidemiology from Johns Hopkins University. He is a Charter fellow of the Academy of Toxicological Sciences, two-term President of the International Society of Regulatory Toxicology and Pharmacology, past member of the Technical Committee and the Regulatory Affairs Committee of the Society of Toxicology, a member of scientific societies and editorial boards, funding editor of Nutrition and Cancer, and editor of Regulatory Toxicology and Pharmacology. He publishes extensively on scientific and policy matters.

EPIDEMIOLOGIC EVIDENCE IN PUBLIC AND LEGAL POLICY: REALITY OR METAPHOR?

by

Gio Batta Gori, ScD, MPH

Epidemiology first acquired true fame by providing crucial assistance in conquering communicable diseases, after it became possible to identify and to control infectious agents that remain their unique and necessary causes. Those early and spectacular successes could not be repeated for diseases that have multiple origins and no specific necessary cause, including cancer, cardiovascular and most other ailments generally related to aging, and of great concern today.

Determining causation for such multifactorial diseases has proven elusive. Experimental trials in humans are beyond the pale, leaving only the possibility of observational epidemiologic surveys inevitably affected by errors, biases, and "confounders" that are only partially if at all controllable. With rare exceptions, causal epidemiologic interpretations for multifactorial diseases inevitably rely on variable judgments that cannot be objectively validated.

A tension is unavoidable between judgmental epidemiology and the sensible requirement for independently testable and objective evidence in policy making and legal proceedings, as epitomized in the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). Still, on the presumption that it portrays direct human experiences, epidemiology has been continually relied upon as a principal tool of advocacy and utilized increasingly in tort claims. Knowing how multifactorial epidemiology works is paramount in understanding what opportunities and limitations it may offer.

I. OF RELIABLE EVIDENCE

Epidemiology aims to uncover statistical correlations of exposures, behaviors, and human diseases, from which scientists can infer causality connections. In dealing with natural events, by necessity epidemiology

entails a quest for empirical truths that makes it subject to rules of testable evidence.

What constitutes truth has been debated since time immemorial, whether in metaphysical or empirical terms. Greek philosophers were first to leave a written record of speculations into the nature of evidence, and came to be the most skeptical of physical senses and perceptions. Those ancients were partial to abstract pursuits of mathematics and geometry that allowed precise statements and conclusions, but their interests in the physical world remained dubious, poetical, and contrasting. In 200 BC, Eratosthenes measured the earth's circumference and tilt and the distance of the sun and moon with uncanny precision, but his findings ultimately failed to prevail over the suffocating orthodoxy of Ptolemy's geocentric model of the cosmos, which endured for almost two millennia.

It was not until the 17th century that philosophers began in earnest to consider again what truth there could be in empirical evidence, responding to the challenge of the first scientific experiments and of their successful claims. Notorious spoilsports, philosophers like David Hume and others were quick to call precarious the validity of experimental evidence, for there could be no certainty that the next experiment would give a consistent result. Although a conjectural truism and a bit of sophistry, this argument has been tormenting philosophers down to this day, predictably without resolution.

Yet, science has progressed beyond the wildest dreams, through experiments that compare undisturbed control conditions as the starting point, with test conditions in which control conditions are modified by potential causes of change. What has made science such a manifestly successful model of inquiry is its commitment to use all possible care in securing independent and objective validation of knowledge, free of judgmental interferences. Without such a commitment, none of the technological advancements would have been possible that have made human life so much easier, rewarding, and safe in the brief span of the last 200 years.

At the same time, the evidentiary method of science is relatively new in mankind's history and culture, and has created a crisis of self-confidence in humanistic professions steeped in centuries-old rhetorical traditions that depend on dogma, artful persuasion, and subjective judgments rather than objective demonstration. Especially from the middle of the last century, practitioners in a branch of philosophy, the philosophy of science, have attempted to cast aspersion on science, eventually claiming that it is

no more reliable than astrology. Such nonsense has made unfortunate inroads into the media and public opinion, even though it is fully disclaimed by the unique and spectacular advances of science itself.

Scientific inquiry begins with the formulation of conjectures or hypotheses about possible causal mechanisms, which are then tested to be validated or rejected. At times, theoretical proofs of certain conjectures might be reached by mathematical deductions, but the essential and convincing proofs are invariably the product of direct empirical tests. It is only after validation by proper experiments that conjectures and hypotheses could be considered as creditable knowledge with any degree of scientific certainty.

In the words of the National Academy of Sciences:

> The task of systematizing and extending the understanding of the universe is advanced by eliminating disproved ideas and by formulating new tests of others until one emerges as the most probable explanation for any given observed phenomenon. This is called the scientific method.
>
> An idea that has not yet been sufficiently tested is called a hypothesis. Different hypotheses are sometimes advanced to explain the same factual evidence. Rigor in the testing of hypotheses is the heart of science. If no verifiable tests can be formulated, the idea is called an ad hoc hypothesis - one that is not fruitful; such hypotheses fail to stimulate research and are unlikely to advance scientific knowledge.[1]

Given the inevitable physical limitations of observation and measurement, scientists do not search for absolutes, but rather proceed to reinforce approximations. Their continuing preoccupation is not only how to increase precision, but more so how to approach objectivity, namely results that stand independently proven to be what they are.

Canons of experimental practice are the hallmark of how valid scientific knowledge is obtained, foremost requiring a rigorous curbing of personal biases and judgmental interferences. A prime advantage is that the conditions of scientific experiments can be carefully planned in advance to eliminate or control external influences.

For instance, in finding the conditions under which heat causes water to boil it would be necessary to know in advance and during the

experiment the precision of the methods for measuring temperature changes, where the measuring instruments are located, the stability of the barometric pressure, the chemical purity of the water, the shape of the vessel holding the water, the way heat is applied, and other details.

Scientific experiments are said to be reductionist, in the sense that scientists aim at reducing external influences by planning in advance how to eliminate or control any interferences that may corrupt the outcome under study. Ultimately, scientific objectivity is defined by the degree of assurance of having control over experimental corruptions.

The guide of scientific experimentation is a set of common sense rules embedded in the scientific method aiming at attaining control and if possible, the elimination of mistakes, delusions, and biases that may confuse outcomes. Most non-scientists are intimidated by the baffling difficulty of scientific research, but although the testing techniques, procedures, apparatus, and language of science can be bewilderingly complex, their common goal is to satisfy three essential requirements:

> •A warrant of identity and accuracy, i.e. the object measured in an experiment is indeed what it is claimed to be measured, and measured with sufficient accuracy.
>
> •A warrant that the effects observed are due exclusively to the causal hypothesis studied (for instance: exposure to a chemical), and not to other disturbances that interfere with an experiment's conduct and may alter and confound the results.
>
> •A warrant that an experiment and its results have been consistently reproduced in the hands of different experimenters.

Without meeting these three rules, no hypothesis or conjecture could be considered enduring and useful knowledge with any degree of objective certainty. It also could not be confidently transformed into reliable technologies or reasoned policy decisions, either public or private. Mistakes and oversights remain possible, but without conforming to the discipline of those warrants, conjectures remain *ad hoc* hypotheses that are "not fruitful," in the words of the National Academy of Sciences.

Arguably, not much more than these intuitive ground rules underpin the spectacular advances of scientific knowledge in the natural world. There are of course the imaginative creativity and inventiveness of scientists, but that belongs to the research phase of science, to hypotheses

and knowledge-in-the-making that still need to be tested for biases, confounders, precision, and reproducibility. It is not yet knowledge that objective scientists would consider fit for advancing new and reliable theories, for flying airplanes, producing effective and safe medicines and foods, or for informing personal decisions, public policies, and court proceedings. Research scientists may work creatively with hypotheses, but only rigorous and repeatedly successful experimental validations in different hands provides an objective test for proximate truth and predictivity, albeit within probability boundaries.

How does epidemiology measure up as science?

II. EPIDEMIOLOGY AND SCIENCE

Epidemiologists have long pressed the claim that their discipline is a science, counting on the common perception that science deals with proven facts. Until some fifty years ago, epidemiology centered on communicable diseases, achieving spectacular advances as epidemiologic hypotheses of causation were refined by the discovery of necessary causal entities – bacteria and viruses — in laboratory and clinical studies. Removal of such causes by sanitation, vaccination, or medication resulted in the control or elimination of the diseases in question and thus confirmed experimentally and unambiguously their causative roles.

Indeed, the material success, the consistency, and the reproducibility of communicable disease epidemiology defines it squarely as mainstream science. Similar success has been obtained for some chronic diseases of non-infectious nature in occupational settings, where causative factors could be specifically identified, and where their removal led to the control or disappearance of related diseases.

As communicable diseases waned, a surge in life expectancy led to a surge of chronic diseases linked to old age, which by now have attracted the attention of a majority of epidemiologists. In general, however, determinations of causality have been elusive for most chronic conditions that depend on a constellation of possible causes, none of which is either necessary or sufficient. More often than not, laboratory and clinical studies have been proven unable to determine specific causal mechanism for such diseases.

In attempting to fill the gap, a set of accessory postulates of causality was adopted to infer causality from observational statistics of multifactorial

conditions. They include the familiar criteria of consistency, strength, specificity, temporal relationship, and coherence as catalogued in 1965 by B.A. Hill and named after him.[2] Bereft of quantitative and qualitative benchmarks, these criteria have remained judgmental and are not linked to independent experimental verification. In the words of the Surgeon General's report on cigarette smoking: "The causal significance of an association is a matter of judgment" — justifiably a prudent judgment in that case, owing to the exceedingly robust association of smoking and lung cancer.[3]

Following the authority of the Surgeon General, a succession of professional authorities have agreed that most causality determinations in multifactorial epidemiology have been and continue to be defined by sensible judgments. To mention just a few of these authorities, in a 1970 textbook McMahon and Pugh noted that: "a causal association may usefully be defined as an association between categories or events or characteristics in which an alteration in the frequency or quality of one category is followed by a change in the other."[4] In a later textbook, Kleinbaum and associates wrote: "In epidemiology we use a probabilistic framework to assess evidence regarding causality — or more properly to make causal inferences…[but] we need not regard the occurrence of the disease as a random process; we employ probabilistic considerations to express our ignorance of the causal process and how to observe it."[5]

Doll and Peto framed even more explicitly the issue of multifactorial causality, as they wrote:

> [E]pidemiological observations…have serious disadvantages… [T]hey can seldom be made according to the strict requirements of experimental science and therefore may be open to a variety of interpretations. A particular factor may be associated with some disease merely because of its association with some other factor that causes the disease, or the association may be an artifact due to some systematic bias in the information collection……..
>
> [I]t is commonly, but mistakenly, supposed that multiple regression, logistic regression, or various forms of standardization can routinely be used to answer the question: "Is the correlation of exposure (E) with disease (D) due merely to a common correlation of both with some confounding factor (or factors)?"

> ... Moreover, it is obvious that multiple regressions cannot correct for important variables that have not been recorded at all."…….[T]hese disadvantages limit the value of observations in humans, but...until we know exactly how cancer is caused and how some factors are able to modify the effects of others, the need to observe **imaginatively** what actually happens to various different categories of people will remain."[6]

Parallel remarks are to be found in the Reference Guide to Epidemiology of the Federal Judicial Center's Reference Manual on Scientific Evidence, which states that "...epidemiology cannot objectively prove causation; rather, causation is a judgment for epidemiologists and others interpreting the epidemiologic data;"[7] ".. the existence of some [associated] factors does not ensure that a causal relationship exists. Drawing causal inferences after finding an association and considering these factors requires judgment and searching analysis;"[8] and "[w]hile the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made by using scientific methodology."[9]

Thus, while epidemiologists insist that their discipline is a science, clearly it is not a mainstream experimental science that produces reliable causal connections to fuel new scientific discoveries and successful technological advances.

More to the point, if multifactorial epidemiology does not operate in the framework of science, what warrants of reliability could it offer?

III. EPIDEMIOLOGIC EVIDENCE

It should now be apparent that an irreconcilable tension exists between the Supreme Court's directions in *Daubert* and the evidentiary powers of multifactorial epidemiology, for the Court insisted that scientific evidence must not be a matter of opinion or judgment, but of objectively testable reality:

> Under the Federal Rules of Evidence (FRE), a federal trial judge is not disabled from screening purported scientific evidence; rather, the trial judge must insure that any and all scientific testimony or evidence admitted is not only relevant but reliable; the primary focus of this obligation is Rule 702 of the FRE, which governs expert testimony as to scientific knowledge; for the purposes of Rule 702, "scientific" implies

> a grounding in the methods and procedures of science, and "knowledge" connotes more than subjective belief or unsupported speculation; although it would be unreasonable to conclude that the subject of scientific testimony must be known to a certainty, Rule 702 requires that proposed scientific testimony be supported by appropriate validation - that is, good grounds - based on what is known; Rule 702's requirement that an expert's scientific testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability, that is, trustworthiness; in a federal case involving scientific evidence, evidentiary reliability is based on scientific validity.[10]

Yet, epidemiologic evidence is a matter of personal interpretation subject to no particular rules. The Reference Guide to Epidemiology of the Federal Judicial Center's Reference Manual on Scientific Evidence plainly admits as such: "[t]here is no formula or algorithm that can be used to assess whether a causal inference is appropriate..."[11]

It would be out of place to review here all that has been written about causality in multifactorial conditions, although it should be noted that virtually all arguments have ended with the explicit or implicit assessment that only sensible judgmental approaches are possible.[12-14] Less discussed within the epidemiologic profession is how judgments can be variably sensible, and how difficult it has been to reach judgments of enduring value.

Since the mid-1960s an acute need was felt to verify experimentally the validity of risk hypotheses raised by epidemiologic surveys regarding diet, alcohol use, cigarette smoking, cholesterol, and the like. Such experiments began in the early 1970s in the U.S. and Europe, and were aimed at controlling by several massive and prospective randomized trials what seemed the most important of those possible causal hypotheses. It should be noted that unlike observational surveys, randomization of participant assignment gave a truly scientific framework to those experimental trials.

Among American studies were the Multiple Risk Factor Intervention Trial (MRFIT) and the Framingham Study of the National Heart, Lung, and Blood Institute; The Minnesota Health Program; The Pawtucket Heart Program; The Stanford Five City Project; and other studies. Of international scope were the World Health Organization MONICA Trial and European Collaborative Trial on CHD Factors; The North Karelia

Study; The Helsinki Businessmen Study; the UK Whitehall Study; and many others. Altogether, scientists conducted experiments on hundreds of thousands of participants.

Epidemiologists are at a loss to explain the failure of those trials, whose results could not verify the "sensible" underlying hypotheses, and which often gave signals contrary to the hypotheses tested.[15-29] The International Union Against Cancer concluded that a series of intervention trials have "…not resulted in any substantial effect on cancer risk."[17] For cardiovascular diseases, the Minnesota Health Program reported that "[t]he overall program effects were…generally within chance levels."[18]

Also for cardiovascular diseases, the Stanford Five-City project reported that the "…conclusions about the overall effectiveness of the communitywide efforts were not always possible," even though risk factor "…changes in treatment towns exceeded those in control towns, and there was general consistency of these control-treatment differences across risk factors…."[19] The Helsinki Businessmen Study reported an unexpected "…excess mortality in the intervention group compared with the control group."[20] The MRFIT program did show negligible improvement for cardiovascular diseases, no statistical difference in overall mortality between the special intervention (SI) and the usual care (UC) groups,[21-24] and "…unexpected higher rates of lung cancer among SI as compared to UC subjects," even though SI subjects had substantially reduced their smoking.[21]

There are further examples, but virtually all trials resonate with equally distressing summations, all the more significant because derived from actual scientific experiments and not from observational conjectures. Absolutely unknown to the public, such massive failures show how uncertainty weighs heavily in the observational epidemiology of multifactorial diseases — an uncertainty that conditions much of what evidence in epidemiology is all about.

The stark reality is that, with rare exceptions, the complexity of multifactorial interactions precludes the possibility that observational studies could specify the fractional causal responsibilities of competing risk factors. Such studies lead not to specific causality statements, but to more or less plausible theories of causality, where the emphasis on one factor or another depends — *inter alia* — on the study set-up, the way biases and confounders are addressed, the manner of data acquisition, the statistical arguments applied and, last but not least, the personal leanings of investigators.

A brief inquiry into how studies of multifactorial epidemiology are conducted clarifies why they do not conform to a scientific framework, and why the evidence produced cannot be other than judgmental.

IV. EPIDEMIOLOGIC STUDIES

A scientific epidemiologic study is a prospective experiment where human subjects are randomly assigned to groups before they will be exposed or not exposed to an agent of interest, during a time adequate for the development of the expected results. Random assignment of participants to test and control groups, sufficient numbers of participants, and a careful control of potential biases and confounders from the start, all contribute to obtaining results that are in fact due to the agents studied and not to interfering corruptions.

Randomized trials are commonly carried out in testing what is supposed to be beneficial to study subjects, such as medicines, vaccines, and other healthful interventions. Additional examples include the just mentioned massive randomized trials that tested unsuccessfully the possible benefits of reducing or removing hypothetical risks.

Still, by far the bulk of epidemiologic surveys of public interest deals with addressing conjectural but presently unknown risks, making it unethical to use prospective randomized trials in testing potentially toxic agents and interventions in humans. The epidemiology of possibly dangerous effects is therefore limited to surveys in populations that have been fortuitously exposed to presumed toxic agents due to lifestyle (e.g. alcohol, smoking, diet…) or environmental contingencies (e.g. workplace, sunlight…).

In such surveys it is impossible to compare groups of subjects that are homogeneous (e.g. randomized), and it is never possible to measure with credibility the extent, specificity, and duration of the exposures considered. Epidemiologic surveys are thus observational and not experimental, because they simply observe what may have happened, without any opportunity of taking preemptive steps to ensure the independent reliability of observed outcomes.

Epidemiologic surveys also are difficult if not impossible to be faithfully replicated, because each study population is unique, each survey's design and methodology are idiosyncratic, and the complete original data are not usually released.

Thus, by definition observational surveys cannot be conducted according to the canons of scientific experiments, and therefore do not qualify as being scientific — a conclusion that we have seen confirmed by prominent epidemiologists and the guidelines of the Federal Judicial Center.

A. Study Types

Epidemiologic studies require different structures to address different survey opportunities. Cohort studies are utilized to observe differences of disease incidence (i.e. frequency) in groups of people exposed or not exposed to possibly toxic conditions. Conversely, case-control studies are utilized when it is the only feasible way to observe differences of exposure to postulated toxic factors in groups of people with or without disease.

Cohort studies can be prospective or retrospective. The former identify groups of subjects exposed or not exposed to potentially toxic conditions, and follow them over time — often years — to record the disease experience of each group. The latter identify groups of subjects with different incidences of diseases, and attempt to reconstruct the past exposures of these groups to possible toxic conditions. Both prospective and retrospective studies can stratify exposure levels, where higher incidences in relation to higher exposures are interpreted as increasing estimates of risk. Risk reduction or protection is assumed if incidence decreases at increasing exposure levels.

Case-control studies are necessarily retrospective, as they compare past exposure experiences in groups with or without a specific disease. The incidence is 0% in the controls and 100% in the cases, and therefore a key understanding is that in such studies risks are inferred as differentials of exposure, and not actually estimated as differentials of incidence. Increased risk is inferred but not directly estimated if exposure is found to be higher among cases, and protection is inferred but not directly estimated if exposure is found to be higher among controls.[25]

B. Risk Measures

Risk is measured as a difference in disease incidence between exposed and non-exposed subjects in cohort studies, where such difference is defined as relative risk (*RR*):

RR = Incidence rate in exposed / Incidence rate in non-exposed

The *RR* ratio reflects that a certain incidence of disease is observed in both non-exposed and exposed human subjects, due to multiple background causes operating in conjunction with, or entirely separate from the exposure under study. Therefore, risk in the exposed is said to be an increment or decrement of incidence, *relative* to the basic incidence of the non-exposed subjects.

In the above equation, if the rates are the same in exposed and non-exposed subjects, i.e. *RR*=1, and there is no risk differential. If *RR* >1 the risk is said to have increased in the exposed subjects, if *RR*<1 the risk is said to have decreased in the exposed subjects, indicating that the exposure under study might be possibly beneficial to humans.

Case-control studies assess differences of exposure and not of disease incidence. Because such studies simply infer but do not directly estimate possible risk, their results are expressed by the convention of odds ratio (*OR*), namely the ratio between the odds (expressed as % or "other rate") of being exposed for the cases and the controls:

OR = Odds that cases were exposed / Odds that controls were exposed

In the above equation, if the odds are the same in exposed and non-exposed subjects, *OR*=1, and there is no inference of risk differential. If *OR* >1 there is an inference of increased risk in the cases, if *OR*<1 there is an inference of decreased risk for the cases, presuming that the exposure may possibly protect for the disease under study.

C. Attributable Risk

As noted, a relative risk value includes the baseline risk of the non-exposed and the added risk that is presumably due to the exposure studied. Thus, what risk can be attributed to the exposure is a fraction of the relative risk. Attributable risk or *AR* is calculated by the following equation:

$AR = (RR-1)/RR = 1-1/RR$

For instance, if *RR* = 1.5, the corresponding *AR* value is 0.33, meaning that the exposure studied might be responsible for 33% of the total RR value, provided that the RR value is credible.

Save for differences in terminology, the arithmetic of risk calculation is the same for cohort and case-control studies. The important difference is that the former estimate risk directly as differentials of disease incidence, while the latter only infer risk from differentials of exposure. Both types of study are affected by similar difficulties of design, data collection, and interpretation — difficulties that are far more acute for case-control studies that rely solely on vague human recollections of exposure.

D. Adjustments

Diseases are more prevalent at certain ages, making it necessary to approximate equal age conditions when comparing dissimilar groups, an operation performed through age standardization procedures. Similar procedures are used in efforts to equalize different groups for socioeconomic status, education level, race, gender, housing conditions, occupation and other common variables. Useful as they may be, such standardizations remain rough approximations.

Biases are common. A selection bias occurs when the non-exposed control subjects differ from the exposed test subjects in regard to characteristics that cannot be standardized for age, gender, etc. In fact, selection bias is impossible to eliminate in epidemiologic studies, and its presence can only be guessed but not measured with any precision.

Information bias relates to inevitable inaccuracies in data collection. Recall bias is most frequent, and is of special concern in case-control studies, where participants with a disease are apt to recall more intense and longer exposures than the control participants who don't have the disease. Recall bias and error may exacerbate when exposure information is retrieved from next of kin of deceased subjects. Exposure reconstruction from company records and other sources may also be biased in relation to the parallel reconstruction for control subjects for whom no exposure records may exist.

Differential accuracy of disease diagnostics and death certificates may affect the classification of subjects. A misclassification bias occurs

when subjects are mistakenly assigned to a group because of inadvertent or willfully wrong responses from interviewed subjects or next of kin.

Confounders are defined as hidden risk factors that could also participate in an association. Thus, poor housing conditions and respiratory problems may appear associated, when all or part of the association might be explained by the existence of confounders like smoking habits that are more prevalent in poor households. Methods are used to uncover and reduce the effect of possible known confounders, but those effects can only be partially controlled because of inherently uncertain complexities. No control is possible for hidden and unknown confounders.

For instance, studies dealing with lung cancer should consider some three dozen risk factors as potential confounders reported in the literature, and studies of cardiovascular conditions face over 300 published accounts of risk factors as potential confounders. Without a credible control for at least all major known confounders, epidemiologic studies could not be validly interpreted.

E. Statistical Considerations

Large as they might be, epidemiologic surveys usually sample only a small fraction of a population and therefore incur a statistical error of measurement. This means that a given risk estimate is not precise, but may be off the mark a certain amount of error. In other words, the true estimate may lay in the interval between the higher and lower figures that the error comports.

Given that such error is inevitable, the question of how much statistical error is tolerable arises. In what clearly amounts to an arbitrary judgment call, a standard consensus has been adopted that no more than a 5% margin of statistical error is tolerable. The complementary implication is that no less than a 95% probability of statistical certainty is acceptable. By this convention, statistical results that display less than a 5% margin of error are said to be statistically significant at the 95% level of certainty, whereas if the error exceeds 5%, the results are said to be statistically insignificant.

Statistical error is characterized by tests of significance defined as p-values or as 95% confidence intervals (*CI*).

Statistical procedures calculate the p-value of estimated risks, with the threshold of significance being at p=0.05, which is the numerical representation of a 5% margin of error. Values lower than p=0.05 are said to characterize statistically significant results at the 95% level, and vice versa. Thus, a relative risk or odds ratio of 1.6 with a p-value of 0.03 is said to be statistically significant at the 95% level, whereas a relative risk of 2.3 with a p-value of 0.07 is not.

The 95% confidence interval is more informative than p-values, although based on the same concepts and calculations. In standard form it gives a range of values within which the statistically true value of a relative risk or odds ratio is likely to be located with a 5% probability of error and a 95% probability of certainty.

In interpreting a confidence interval it is important to recall how risk or odds ratios are calculated. For both indexes a value of 1 means no change in risk because incidence is the same in non-exposed and exposed subjects, or because exposure is the same in cases and controls. Values below 1 imply risk reduction or protection, values above 1 imply increased risk. The 95% confidence interval gives an immediate impression of statistical significance, which is characterized by confidence intervals whose values are all either above of below 1, that is all in negative or positive risk territory. By contrast, intervals containing values above and below 1 define non-significant results, which remain moot because they simultaneously imply an increase and a decrease of risk. Thus:

- *RR* or *OR* = 1.9 (95%CI 1.2-4.6) means that the best estimate of the risk may be 1.9, but that its true value could be between 1.2 and 4.6, with a probability of 95%. It also means that within that range all values are statistically significant at the 95% level, because all would mean an increase of risk, the lowest value still being >1.

- *RR* or *OR* = 1.9 (95%CI 0.7–2.3)) means that the best estimate of the risk may be 1.9, but that its true value could be between 0.7 and 2.3, with a probability of 95%. It also means that some values could be <1 and could mean protection, others could be >1 and could mean risk. As a consequence the result is said to be equivocal and not statistically significant.

- *RR* or *OR* = 0.7 (95%CI 0.2-0.9)) means that the best estimate of the risk may be 0.7, but that its true value could be between 0.2 and 0.9, with a probability of 95%. It also means that within that range all values are statistically significant at the 95% level, because all would mean a reduction of risk, the highest value still being <1.

•*RR* or *OR* = 0.7 (95%CI 0.3-1.9) means that the best estimate of the risk may be 0.7, but that its true value could be between 0.3 and 1.9, with a probability of 95%. It also means that some values could be <1 and could mean protection, others could be >1 and could mean risk. As a consequence the result is said to be equivocal and not statistically significant.

Detailed accounts of epidemiologic methodologies for dealing with biases, confounders, standardizations, and statistics can be found in textbooks.[31] Still, it should be useful to acquire a perspective on statistical significance. The 5% convention is equivalent to a 1 in 20 threshold of acceptable error, which would be disastrous in most everyday activities. Would it be sensible to drive a car if 1 time in 20 the brakes failed, or it turned left when attempting to steer right? Science valid enough for reliable applications must attain much lower margins of error, usually far less than one in a million. As an extreme benchmark, Feynman's experimental account of quantum electrodynamics predicts accurately to some 11 decimal places, or to a level of error of less than 1 in 100 billion.[32]

An additional warning is that statistical error and certainty are both figures of probability that refer to the accuracy of the numerical context of the data available, and tells nothing about the reliability of the data themselves. For instance, statistics cannot know or determine whether exposure recall data are credible or not, nor whether biases or confounders may be present. Statistics is blind to whatever influences may be corrupting the underlying data of a study.

F. Multiple Studies

Meta-analysis is the statistical technique used to pool results from different studies. Originally it was developed for summarizing the results of homogeneous randomized clinical trials, which remains its legitimate application. However, using meta-analysis for pooling the results of diverse observational studies is fraught with irresolvable difficulties.

The procedure gives different weights to studies, primarily in relation to their size. However, meta-analysis does not pool the discrete data that originated each result, but only the final results of each study regardless of whether they are concordant or discordant, credible or not.

The procedure does not discriminate according to characteristics of each study, such as its design, data collection, standardizations, biases, confounders, adjustments, statistical procedures, etc. Meta-analysis,

therefore, produces only a weighted average of the final numerical results of the studies, but does not standardize, relieve, or control for differential corruptions that may be present in each study. If characteristics other than study size are used in weighing studies (e.g. study quality), those characteristics are likely to be discretionary, judgmental, and conducive to different meta-analysis results when handled by different analysts.

Statistical tests of homogeneity for a group of studies do exist, but again they relate only to numerical homogeneity and say nothing about other determinant characteristics. The Reference Guide to Epidemiology of the Federal Judicial Center's Reference Manual on Scientific Evidence warns that

> [a] final problem with meta-analyses is that they generate a single estimate of risk and may lead to a false sense of security regarding the certainty of the estimate. People often tend to have an inordinate belief in the validity of findings when a single number is attached to them, and many of the difficulties that may arise in conducting a meta-analysis, especially of observational studies like epidemiologic ones, may consequently be overlooked.[33]

Therefore, with the exception of its use for summarizing homogeneous randomized clinical studies, it should be manifest that meta-analysis can be used as a strategy to contrive meaning from studies that have no apparent meaning.

More importantly, the above excerpt from the Reference Guide on Epidemiology leads to a general but crucial warning in reading and interpreting epidemiologic reports. Numerical displays in epidemiology should be seen as having an analog rather than digital meaning. Most numbers in epidemiology are metaphorical proxies of uncertain real quantities, for epidemiology rarely measures reliably, and more commonly evokes, conceives, assesses, sizes up, adjusts, rounds up, and appraises.

Indeed, numerical transformations and renditions impart an undeserved sense of accuracy and credibility to a background of vagueness caused by study design deficiencies, asymmetries in data collection, statistical error, biases, confounders, limitations of adjustments and standardizations, prejudice, and more. Tests of statistical significance are equally speculative, being no more than approximate summaries of metaphorical primary data.

As a further cautionary note, the greater the complexity of the statistical analysis in epidemiologic reports, the weaker the data is likely to be. Known as data dredging, epidemiologists sometimes squeeze every conceivable signal out of what is usually a congeries of data.

How do epidemiologists approach the inherent fragility of their data?

V. EPIDEMIOLOGISTS AND UNCERTAINTY

Epidemiologists react to uncertainty by taking contrasting positions. At one end, the mainstream profession is focusing on the specificity and accuracy of data collection, and on controlling as best as possible for biases and confounders. This concern may be reflected further in cautious, balanced, and truthful representations of epidemiologic uncertainty to the media, the public, and policymakers.

At the opposite end, a long tradition of advocacy views epidemiology as a fungible tool for the advancement of goals aligned with personal, cultural, and political views, most often reflecting a consuming enthusiasm for what the French would call *dirigisme*. In too many instances advocacy has prevailed over interpretive restraint, leading to confused and quarrelsome debates.

Advocacy positions are typically supported on the grounds that "[d]espite philosophic injunctions concerning inductive inference, criteria have commonly been used to make such inferences. The justification offered has been that the exigencies of public health problems demand action and that despite imperfect knowledge causal inferences must be made."[34]

The circularity of such a justification is manifest, when one realizes the large number of exigencies of public health that have been created by epidemiologists on the basis of knowledge that is marginal, if not wholly conjectural.

Undoubtedly advocacy has valid roles, but it should be apparent that its legitimacy is proportional to the factual reliability of what is being sustained. Epidemiologists are divided on this issue. A new "paradigm" of epidemiology has been proposed, one that shows little patience with the scientific method, while being still reluctant to be perceived as non-scientific.

Proponents claim that "epidemiologists among others have been misled by standard interpretations of the nature of science"[35] and therefore "to control for confounders…strips away the essential historical and social context, as well as the multiple moderating influences than constitute true causation."[36] For those proponents, causal agents are to "be seen as resulting from mechanisms that are internal to the population under study and that operate dialectically, rather than involving regular associations between externally related independent objects."[37]

On the other hand, a novel methodology is proposed, where "the solution is not to abandon scientific standards, but rather to apply them more rigorously" even though "it is inappropriate to falsely dichotomize research methods as: quantitative vs. qualitative, hard vs. soft, deductive vs. inductive, or objective vs. subjective."[38] The new paradigm focuses on broader historic, cultural, socioeconomic, and political determinants of health and disease[35-41] because "[r]igid adherence to an arcane view of science… [is] likely to promote narrow disciplinarian sectarianism at a time when an even more multidisciplinary ecumenical approach to public health challenges is required."[41]

Eventually epidemiologists are seen as "professionals in the sense traditional to medicine, the law, and the clergy. That is, society accords them a privileged and autonomous function funded on special training."[42]

It should be apparent that individual decisions and public policy may have serious problems with a priesthood of epidemiology that claims privileged knowledge, and is inclined to resolve causal theories dialectically and through internal consensus. In effect, the proposed paradigm presents an alternative epidemiology much in the spirit of alternative medicine. The alternative extends the weak claim that nothing better is available, to the assertion that good intentions alone justify the imposition of creative and usually interested conjectures. Extreme proponents of this stance also scorn the truth that some questions may have no ready answers.

The counterpoint of sober epidemiologists is that in most advocacy positions "scientific principles… are…. disregarded not because they are difficult to understand, but because they are difficult to carry out….[T]he customary excuse for ignoring scientific principles is the argument that they are not necessary in epidemiologic research…[and] that no additional scientific principles need be invoked because each epidemiologic procedure has its own distinctive standards…established by a consensus of appropriate authorities."[43] Other sensible critics of the proposed paradigm also contend that such loose thinking invites excessive reductionist

assessments, which generate "the illusory comfort of perhaps metaphorical meta-theories that appear to explain everything while accounting for nothing."[44]

Both mainstream epidemiologists and advocates concur that causal theories is what they produce, but they differ in interpretive restraint. The distinctive characteristic of the advocacy approach is a determination to inject idiosyncratic sociopolitical views into epidemiology, claiming that this makes epidemiology "balanced and responsible." The claim, however, is unsustainable, for the intrusion of ideology in the evidence-gathering process negates the original element of factual truth, or approximation of such truth, that is needed for responsible public health actions.

How is epidemiology interpreted when scientific validation remains elusive?

VI. INTERPRETING EPIDEMIOLOGY

As noted, ethical considerations and the stark reality of complex multifactorial interactions preclude the possibility of controlled experiments to specify the causal responsibilities of competing risk factors.

The Reference Guide to Epidemiology of the Federal Judicial Center's Reference Manual on Scientific Evidence concurs that in epidemiology "[w]hile the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made by using scientific methodology."[45] In this quotation, scientific expertise refers generally to the use of statistics in analyzing the data, but the statement negates that epidemiologic studies could follow the scientific method.

Fact finding in epidemiology is mostly "a matter of judgment," in the previously cited words of the Surgeon General. However, the important distinction is that during the initial fact finding phase it should be a judgment of cumulated evidence and not a judgment of conditional values — the latter coming later within risk management decisions in the context of public health policy and action, where advocacy may also find its place.

In order to carry some weight, epidemiologic surveys need to document quantitatively the exposure to the agents being considered, to provide accurate and definitive pathological data, to show an appropriate study design and choice of test and control groups, and robust statistical and analytical constructs.

Studies should also demonstrate successful control over possible confounders, systematic and random errors, and selection and information biases. Results should provide evidence of a dose-response gradient, and be consistent with the results of similar studies. Such standards are rarely met, and a majority of studies are published without sufficient information for uninformed readers to know what details are missing.

Still, it is common practice to draw causal inferences from single studies or groups of studies on the basis of the Hill criteria previously mentioned and listed below:

A. Strength

The strength of an association is a clue to causation, although a strong association is neither necessary nor sufficient to affirm causality, and a weak one is neither necessary nor sufficient to deny causality.

B. Consistency

Consistency of results from different studies is an obvious attribute of true causal relationships. Yet, false associations also could be repeatedly consistent because of a consistent correlation with different but related causes. There is no criterion to distinguish whether a consistent association is true or false in epidemiology, but epidemiologic associations that are inconsistent are quite likely to be untrue.

C. Specificity

Specificity requires that a cause leads to a single effect, which is seldom the case in multifactorial epidemiology. Smoking for instance leads to many different effects.

D. Temporality

Effects must occur after the cause has a chance to act. This is a valid, if self-evident and trivial criterion of causality.

E. Biologic gradient

A dose-effect relationship is a useful but not dispositive criterion of causation. An observed dose-response gradient could be due to the presence of biases and confounders, and not to the variables at issue.

F. Plausibility

Whether an association is biologically plausible or not remains a matter of individual speculation and is far from being objective or conclusive.

G. Coherence

Agreement with other information may be a corollary attribute of causation. However, conflicting information could be erroneous.

H. Experimental evidence

Experimental evidence in humans would indeed constitute proof of causation, but it is very rarely available for multifactorial conditions.

I. Analogy

Authorities in epidemiology comment that "whatever insight might be derived from analogy is handicapped by the inventive imagination of scientists who can find analogies everywhere."[46] Analogy is an absolutely invalid criterion in a judgment of causation.

In their textbook, Rothman & Greenland summarize Hill's criteria as follows:

> As is evident, the standards of epidemiologic evidence offered by Hill are saddled with reservations and exceptions. Hill himself was ambivalent about the utility of these 'standards' (he did not use the word criteria in the paper). On the one hand, he asked, 'In what circumstances can we pass from this observed association to a verdict of causation?' (original emphasis). Yet, despite speaking of verdicts of causation, he disagreed that any 'hard-and-fast rules of evidence' existed by which to judge causation: 'None of my nine viewpoints [criteria] can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non.'[47]

Therefore, authorities in epidemiology negate causal persuasiveness to criteria that have no power as primary inferential determinants of causality, but simply remain *ex post facto* attributes of causal conclusions that had been previously validated experimentally.

Indeed, it is remarkable and surprising that epidemiologists have not seen fit to include prominently among the above criteria those most important warrants that might begin to give some measure of confidence to epidemiologic studies: namely again that what has been measured is what is said to have been measured, and measured with acceptable accuracy and statistical error, and that known biases and confounders have been controlled to the best possible and sufficient extent.

Is epidemiology coming to a dead end? Not any time soon, for epidemiologic studies of chronic diseases have a moral imperative of their own. Such studies will be done in affluent societies acutely preoccupied with those very diseases, the results will be published, and they will influence public and private policies.

VII. EPIDEMIOLOGY AND POLICY

Societies and individuals regularly and prudently adopt protective measures against probable risks. However, the problem with multifactorial epidemiologic risks is that they may be conjectural rather than measurably probable, thus creating unforeseeable odds against insuring against them. In fact, most public health, safety, and environmental regulations may qualify as being politically prudent, but very few can warrant to have become effectively successful.

Only a few exceptional regulatory policies have confirmed the related epidemiologic studies by proving effective, notably those that have controlled diseases linked to certain industrial hazards now strictly regulated in the workplace. Exceedingly high risks also have successfully warned of hazards, as was the case for cigarette smoking.

Still, despite the shortcomings of epidemiology, a combination of well conducted and controlled epidemiologic surveys, laboratory and animal findings, and clinical reports, have been construed as grounds for prudent public health action. Known in policy circles as "weight of evidence," the approach considers internal quality attributes of individual studies, and compares their final results against a set of external qualifiers of causation: again the Hill criteria.

Epidemiologic uncertainties raise two main political problems: how to resolve cost and benefit considerations in allocating limited resources; and how to decide how much uncertainty is acceptable in restricting the autonomy of free citizens. Clearly, both decisions require ways to measure

and rank the relative uncertainty of different warnings, namely the setting of priorities.

Although much has been said about the use of Hill's criteria in establishing the credibility and rank of epidemiologic reports, public policy decisions have relied most often on quality attributes of studies, namely again the adequacy of design, data collection procedures, matching of test and control groups, confounder and bias control, medical diagnostics, statistical procedures, and other attributes that could speak to the robustness of a study and the influence of possible corruptions. Not to be overlooked in establishing policy priorities is the level of political pressure exercised by different constituencies and advocacies, on the basis of valuations that may have little to do with objective realities.

Often, the significant feature for policy decisions has been the reported magnitude of risk, and especially of attributable risk. Some epidemiologists have suggested that only relative risks above 2 or 3 should be taken seriously. The reasoning of such suggestions is not apparent, however, because the plausibility of relative or attributable risks values can only be a function of how well studies might have been conducted and controlled. Plausible risks would then be projected against cost benefit considerations, the size of the population potentially exposed, the age and gender of the exposed, and other considerations of social, cultural and economic nature.

Although the process is ostensibly anything but scientific and objective, considerations of epidemiologic reports in public policy are inevitable on account of their human substrate and of the immediate political relevance of the collective anxieties they raise. Provisionally tenable as a political exercise of prudence in a population context, such arguments could not apply to the relevance of epidemiology in ascertaining causation for individual instances of disease, where epidemiology is virtually powerless.

VIII. POPULATION VERSUS INDIVIDUAL RISK

Statistical, logical, biological, and common sense considerations argue against using epidemiology to infer individual causalities. If epidemiologic studies or a related public health policy should say that an exposed population might have the probability of a 20% increased risk of a disease, it does not mean that every individual in the population has a 20% increased risk for the disease.

Obviously not all individuals in the population will develop the disease and it is not possible to state in advance who will develop the disease and who will not. It can only be said that single individuals in the population will either develop the disease or not, namely that individual chances are either 100% or 0%. In this regard Rothman writes:

> There is a tendency to think that all of us are subject to a 10 percent probability of lung cancer if we were to become heavy smokers, as if the outcome, aside from smoking, were purely a matter of chance. It is more constructive, however, to view the assignment of equal risks as reflecting nothing more than our ignorance about the determinants of lung cancer that interact with cigarette smoke. It is likely that some of us could engage in chain smoking for many decades without the slightest possibility of developing lung cancer. Others are or will become 'primed' by presently unknown circumstances and need only to add cigarette smoke to the nearly sufficient constellation of causes to initiate lung cancer. In our ignorance of these hidden causal components, the best we can do in assessing risk is to assign the average value to everyone exposed to a given pattern of known causal risk indicators. As knowledge expands, the risk estimates assigned to people will approach one of the extreme values, zero or unity.[48]

Another line of reasoning against applying population inferences to individuals stems from the concept of attributable risk. For simplicity, let us assume that the number of the exposed and unexposed is the same. The unexposed may show x human subjects with the disease, or a rate X. The exposed may shows x + b subjects with the disease, or a rate X+B. The rate B is taken to be the excess or attributable risk of the exposed population. The word "excess" is key. It means that in the exposed population there is the same rate of diseases as in the controls, *plus* an excess rate that might be attributed to the exposure. Clearly, some indistinguishable exposed subjects would have the disease regardless of exposure, at the same rate X incurred by the control population.

Therefore, in confronting a diseased individual from the exposed population, it cannot be said that this individual disease was caused or contributed to by the exposure, for it could be also attributed to the generic causes of the same disease operating in the non-exposed population.

An argument could be made that it should become easier to infer individual causality as the attributable risk increases, but such attributable risk never equals total risk. For instance, it would not be possible to define objectively and quantitatively the relative causal contributions to the disease in a bladder cancer patient who experienced an undefined occupational exposure to aromatic amines and had been a heavy smoker. Individual attribution of causation could only be possible in the exceedingly rare instance of individually-specific biologic evidence of causation attributable to the exposure.

Misguided attempts to invoke epidemiology in the definition of individual causation have become an all-too frequent occurrence in legal proceedings involving torts and personal injuries.

IX. EPIDEMIOLOGY IN COURT

In court, epidemiology appears mostly in administrative or tort cases, where it could relate either to claims of harmful exposures, or to disputes about regulations claimed to be excessive or insufficient.

Regulations are often based on collective or population-based prudence; prudence is judgmental and epidemiology is likewise subjective. Thus, the epidemiologic arguments discussed in the policy section above, matched with extant statutory language and with cost and benefit econometrics, allow the courts to interpret what might be considered "fair" levels of regulation.

More often it is tort cases that invoke the assistance of epidemiology in resolving contentious issues of causation. On this point, the Reference Guide to Epidemiology of the Federal Judicial Center's Reference Manual on Scientific Evidence is plagued with ambiguous conclusions.

The irresolvable problem is that epidemiologic risks pertain to populations but not to individuals within those populations, and therefore it should not be permissible for a litigant to use an epidemiologic risk in assigning a probability of causation for an individual instance. The "more likely true than not true" legal threshold of causation has been translated into an over 50% threshold, meaning an attributable risk over 50%, equivalent to a relative risk over 2.0 or double the risk at baseline. The rationality and fairness of such a threshold of causality is ostensibly problematic as a measure of whether a tort occurred, especially when it could be justified only by metaphorical epidemiologic statements.

Yet many courts have accepted this reasoning in assigning causation and culpability. The first argument against such reasoning is that it is illusory to accept a risk of 2.0 as being real at face value, given all the uncertainties that multifactorial epidemiology entails. Other counterarguments are those just considered in regard to population versus individual risks. A polled group showing a 60% preference for proposition X cannot mean that single individuals in that group are 60% in favor of X, because each may have different levels of agreement, or may be either for or against the proposition.

Again, individuals in an exposed group are at risk both for the exposure and for the baseline risks of non-exposed individuals — which makes it impossible to determine whether individual occurrences may be due to baseline or exposure risks, even for highly attributable risks.

In the end, it should not surprise if the interpretations of epidemiology in the courts have not been and could not be as uniform and consistent as it might be desirable. With *Daubert*, the Supreme Court has raised the bar in requiring that the evidence be independently testable and not a matter of subjective judgment. Still, epidemiology cannot be transformed by a writ of the Supreme Court and the tension will remain, to be resolved within the framework of individual trials.

A drastic solution would be to accept the views of German courts, where evidence that is not testable is generally not admitted — a solution most unlikely for a more elastic American law very much tied to the tyrannical tradition of precedent. Most likely, *Daubert* will stand as a powerful incentive for the courts to be more probing of epidemiologic evidence, and more cautious in its acceptance. With that, *Daubert* also will moderate the *ipse dixit* framework of expert testimony, and will put a new burden of epidemiologic literacy on attorneys on all sides of disputes.

The benchmarks for assessing epidemiologic studies remain fairly simple: Can the studies warrant to have measured what they say to have measured, and with acceptable statistical error? Have the studies controlled for biases and confounders to the best of their ability and knowledge? Are the results concordant in different studies?

CONCLUSION

In closing, two main questions come to mind: what to make of current epidemiologic reports, and how scientists can improve future

epidemiology. The present discussion has provided a basic overview of the methods of epidemiology, and the methods to interpret and respond to its endemic uncertainties. A superficial conclusion is that it should be easier to refute than to sustain epidemiologic claims, although there is obviously more to it. Does epidemiology provide evidence or metaphor?

Epidemiology is largely metaphor used to attempt to define individual risks, but the metaphor could be variously credible as a flag for population or statistical risks. Although still not objective, epidemiology's credibility depends on how well a report or a group of reports show robust study designs and a convincing control of biases and confounders. Lacking the immediate persuasiveness of scientific experiments, the credibility of epidemiology suffers a heavy burden of documentation, the reading of which requires a skeptically trained mind.

As for the future, many epidemiologists agree that epidemiology is not exempt from an injunction requiring that truth be told.[49-56] Epidemiologists should be explicit in disclosing how and when epidemiologic conclusions are not scientifically objective, but the product of what epidemiologists of Doll and Peto's caliber qualify as imaginative interpretations.

A careful openness would go a long way in restoring the reputation of epidemiology, now marred by aggressive claims that are factually unsustainable and that have contaminated the transparency and fairness of public policies and of court proceedings.

Advocacy should not interfere with fact finding, and truthfulness should be restored to epidemiology even as it means admitting the full extent of its uncertainties. The temptation to overstate and to claim possession of inscrutable knowledge would not speak well of the intentions and good will of the profession.

ENDNOTES

1. National Academy of Sciences, RESPONSIBLE SCIENCE: ENSURING THE INTEGRITY OF THE RESEARCH PROCESS, Vol. 1, National Academy Press, Washington, D.C., (1992), at 38.

2. Hill A.B., *The environment and disease: Association or causation?*, 58 PROC. R. SOC. MED. 295-300 (1965).

3. U.S. Surgeon General, *Smoking and Health Report of the Advisory Committee to the Surgeon General of the Public Health Service, U.S. Department of Health, Education, and Welfare. Public Health Service Publication* No. 1103, Washington, D.C. (1964).

4. McMahon B., Pugh T.F., EPIDEMIOLOGY: PRINCIPLES AND METHODS, Little Brown, Boston (1970).

5. Kleinbaum D.G., Kupper L.L., Morgenstern H., EPIDEMIOLOGIC RESEARCH, Wadsworth, London (1982).

6. Doll R., Peto R., *The causes of cancer*, 66 J. NAT. CANCER INST. 1192-1312 (1981) (Emphasis added).

7. Green M.D., Freedman D.M., Gordis L., REFERENCE GUIDE ON EPIDEMIOLOGY. REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (2D ED. SECOND EDITION, Federal Judicial Center, Washington D.C. (2000), at 374.

8. *Id.* at 375.

9. *Id.*

10. 509 U.S. 579 (1993).

11. *Supra*, note 7, at 375.

12. Rothman K.J. (Ed.), CAUSAL INFERENCE, Epidemiologic Resources Inc., Chestnut Hill, MA (1988).

13. Shaffner, K.F., *Causing harm: epidemiological and physiological concepts of causation*, in ACCEPTABLE EVIDENCE: SCIENCE AND VALUES IN RISK MANAGEMENT. Mayo D.G., Hollander D.R., (Eds.), Oxford University Press, NY (1991).

14. Cole P., *Causality in epidemiology, health policy, and the law*, 27 ENVIR HEALTH RPT. 10279-85 (1997).

15. Werkö L., *The enigma of coronary heart disease and its prevention*, 221 ACTA MED SCAND 323-33 (1987).

16. Werkö L., *Analysis of the MRFIT screeners: a methodological study*, 237 J. INT. MED. 507-518 (1995).

17. Hakama M., Beral V., Cullen J., Parkin M., *UICC workshop on evaluating interventions to reduce cancer risk,* 43 INT. J. CANCER 967-69 (1989).

18. Luepker R.V., Murray D.M., Jacobs D.R., et al., *Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Health Program*, 84 AM. J. PUB. HEALTH 1383-1393 (1984).

19. Fortmann S.P., Flora J.A., Winkleby M.A., Sachooler C., Taylor CB, Farquhar J.W., *Community intervention trials: Reflections on the Stanford Five-City project experience*, 142 AM. J. EPIDEMIOL 576-586 (1995).

20. Strandberg T.E., Salomaa V.V., Vanhanen H.T., Neukkarinen V.A., Sarna S.J., Miettinen T.A., *Mortality in participants and non-participants of a multifactorial prevention study of cardiovascular diseases: a 28 year follow up of the Helsinki Businessman Study*, 74 BR. HEART J. 449-454 (1995).

21. *Multiple Risk Factor Intervention Trial Research Group. Mortality rates after 10.5 years for participants in the Multiple Risk factor Intervention Trial. Findings related to a priori hypotheses of the trial*, 263 JAMA 1795-1801 (1990).

22. *The Multiple Risk Factor Intervention Trial Research Group. Mortality after 16 years for participants randomized to the Multiple Risk factor Intervention Trial*, 94 CIRCULATION 946-951 (1996).

23. Shaten J.B., Kuller L.H., Kjeselberg M.O., Stamler J., Ockene J.K., Cutler J.A., Cohen J.D., *Lung Cancer Mortality after 16 years in MRFIT Participants in Intervention and Usual-Care Groups*, 7 ANN. EPIDEMIOL 125-136 (1997).

24. Shaten J.B., Kuller L.H., Kjeselberg M.O., Stamler J., Ockene J.K., Cutler J.A., Cohen J.D., *Lung Cancer Mortality after 16 years in MRFIT Participants in Intervention and Usual-Care Groups*, 7 ANN EPIDEMIOL125-136 (1997).

25. Other types of epidemiologic surveys are seldom of policy or legal interest. Cross-sectional studies take a snapshot of the prevalence of disease in a given population at a given time, which is then correlated with habits and exposures of possible causal relevance. Ecological studies do not compile statistics from individual members of a group, but rather compare overall statistical data of groups against generic characteristics of the groups, such a dietary habits, genetic traits, geographic and environmental conditions, and the like. In general, ecological studies produce preliminary clues for further cohort or case-control studies.

26. Feinlieb M., *New directions for community intervention studies*, 86 AM. J. PUBL. HEALTH, 1696-1698 (1997).

27. Stallones R.A., *Mortality and the Multiple Risk Factor Intervention Trial*, 117 AM. J. EPIDEMIOL, 647-560 (1983).

28. Taubes G., *The soft science of dietary fat*, 291 SCIENCE 2536-45 (2001).

29. Taubes G., *Epidemiology faces its limits,* SCIENCE 164-69 (1995).

30. *Do epidemiologists cause epidemics?* (Editorial), 341 LANCET, 993-94 (1993).

31. Rothman K.J., Greenland S., MODERN EPIDEMIOLOGY (2D ED.), Lippincott Williams & Wilkins, Philadelphia (1998).

32. Feynman R. QED, Princeton University Press, Princeton, N.J. (1985).

33. *Supra* note 7, at 381.

34. Rothman K.J., MODERN EPIDEMIOLOGY, Little Brown & Co., Boston, MA (1986).

35. Susser M., Susser E., *Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology*, 86 AM. J. PUB. HEALTH 674-677 (1996).

36. *Id.*

37. Pearce N., *Traditional epidemiology, modern epidemiology, and public health*, 86 AM. J. PUB. HEALTH 678-683 (1996).

38. Pearce N., McKinaly J.B., *Back to the future in epidemiology and public health: Response to Dr. Gori*, 51 J. CLIN .EPIDEMIOL 643-46 (1998).

39. McKinlay J.B., Marceau L.D., *To boldly go…* 90 AM. J. PUB. HEALTH 25-33 (2000).

40. Koopman J.S., *Emerging objectives and methods in epidemiology,* 86 AM. J. PUB. HEALTH 630-32 (1996).

41. *Supra* note 37.

42. *Supra* note 21.

43. Feinstein A.R., *Epidemiologic analyses of causation: The unlearned scientific lessons of randomized trials*, 42 J. CLIN. EPIDEMIOL 481-89 (1989).

44. Davey Smith G., *Learning to live with complexity: Ethnicity, socioeconomic position, and health in Britain and the United States*. 90 AM. J. PUB. HEALTH 1694-98 (2000).

45. *Supra* note 7, at 375.

46. *Supra* note 31, at 27.

47. *Id.*

48. *Supra* note 34, at 12.

49. Coughlin S.S., *On the role of ethics committees in epidemiology professional societies*, 146 AM. J. EPIDEMIOL 209-13 (1997).

50. Weed D.L., Coughlin S.S., *New ethics guidelines for epidemiology: Background and rationale,* 9 ANN. EPIDEMIOL 277-80 (1999).

51. Jackson J.R., *Activity in 1998 in the development of standards of good epidemiologic practices*, 72Suppl.3 INT. ARCH. OCCUP. ENVT'L. HEALTH M81-M85 (1999).

52. Buncher C.R., Collins J.J., Halperin W.H., *Possible progress and unresolved conflicts resulting from guidelines on good epidemiologic practices*, 33 J. OCCUP. MED. 1261-64 (1991).

53. MacMahon B., *A code of ethical conduct for epidemiologists?*, 44Suppl 1 J. CLIN. EPIDEMIOL 147S-149S (1991).

54. Couglin S.S., Beauchamp T.L. (Eds), ETHICS AND EPIDEMIOLOGY, Oxford University Press, NY (1996).

55. Beauchamp T.L., *Moral foundations*, in: ETHICS AND EPIDEMIOLOGY. Coughlin S.S., Beauchamp T.L., Eds., at 24-52. Oxford University Press, N.Y. (1996).

56. Gori G.B., *Considerations on guidelines of epidemiologic practice*, 12 ANN. EPIDEMIOL 73-78 (2002).

