THE FALLIBILITY OF MULTIFACTORIAL EPIDEMIOLOGY

Epidemiology is the study of the occurrence, distribution, and temporal trends of disease and health in human populations. The primary ambition of epidemiology is to identify causes of diseases, which then could be made to justify various measures of prevention and cure.

Working as it does with humans, epidemiology is constrained by ethical considerations that forbid dangerous experiments in humans. In fact, with the exception of testing the effectiveness of vaccines and medicines that offer hope of improvement, epidemiology is negated the basic scientific opportunity of experiments, and is forced merely to observe superficially what goes on in the world of health and disease. Epidemiologic observations are uniquely affected by many observable and unobservable components, and become impervious to clear and valid causal interpretations.

Observational studies are plagued by the usual errors, biases, and other confounding disturbances that are only partially if at all controllable, leading - with rare exceptions - to interpretations of causality that are inevitably based on variable judgments that cannot be objectively validated.

A tension therefore arises between judgmental epidemiology and the sensible and the essential requirement for independently testable and objective evidence that justifies public health policies. Unfortunately, flawed observational epidemiologic studies have become a principal tool of advocacy and public health claims, based on the lame claim that they portray human experiences.

Hoping to be endorsed by the popular perception that science deals with proven facts, epidemiologists have long tried to characterize their discipline as a science. This may have been true until some 50 years ago, when epidemiology helped achieve spectacular advances in finding the necessary causes of infectious diseases – single infectious agents such as bacteria and viruses – which would have been impossible without the crucial contributions of other disciplines, notably bacteriology, vaccine research, and clinical studies. Removal of such individual causes by sanitation, vaccination, or medicine permitted vast natural experiments that resulted in the control or disappearance of the diseases in question, and confirmed unambiguously their causative roles – a confirmation that would have been impossible for epidemiology alone. 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 infectious diseases waned, there has been a surge of diseases that are not caused by any single and specific agent but depend on a constellation of factors, and

which therefore are called multifactorial diseases. In general, determinations of causality have been elusive for most such conditions, and laboratory and clinical studies have proven unable to determine specific mechanism for such diseases as cancer, cardiovascular disorders and many other conditions.

The fundamental evidentiary problem of multifactorial epidemiology is at least three way: a) the pervasive impossibility of a clean measurement of authentic primary data with testable and narrow margins of error; b) the impossibility of accounting for the meaning and impact of the many factors that could have a causal role for the conditions being studied; c) the extreme difficulty of obtaining consistently replicable results, given the instability of primary data, and the shifting composition and influences of potential causal factors and of biases from study to study.

In attempting to overcome the impasse, a set of judgmental criteria was adopted to infer causality from observational statistics of multifactorial conditions. They include the familiar criteria of consistency, strength, specificity, temporal relationship, and coherence – of which more later - as catalogued in 1965 by A. B. Hill and named after him.[1] Bereft of quantitative and qualitative benchmarks, these criteria have remained judgmental and 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 the case of cigarettes, owing to the exceedingly robust association of smoking and lung cancer risk .[2]

Following 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.".[3] 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

---

[1] Hill AB. The environment and disease: Association or causation? Proc R Soc Med 1965;58: 295-300.
[2] 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, DC. 1964. p. 19
[3] McMahon B, Pugh TF. Epidemiology: principles and methods. Little Brown, Boston. 1970

considerations to express our ignorance of the causal process and how to observe it.".[4]

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 regression 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."(emphasis added). [5]

Parallel remarks are to be found in the Reference Guide to Epidemiology of the Federal Judicial Center's Reference Manual on Scientific Evidence, the principal reference for instructing US courts in regard to epidemiology. The Manual states that "…epidemiology cannot objectively prove causation; rather, causation is a judgment for epidemiologists and others interpreting the epidemiologic data." [6], and ".. 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." [7] 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." [14].

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

---

[4] Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Wadsworth, London. 1982.
[5] Doll R, Peto R. The causes of cancer. J Nat Cancer Inst 1981;66:1192-1312.
[6] Green MD, Freedman DM, Gordis L. Reference Guide on Epidemiology. Reference Manual on Scientific Evidence. Second edition. Federal Judicial Center, Washington DC.2000. . p. 374
[7] Green MD, Freedman DM, Gordis L. Reference Guide on Epidemiology. Reference Manual on Scientific Evidence. Second edition. Federal Judicial Center, Washington DC.2000. p. 374

health policies. More to the point, if multifactorial epidemiology does not operate in the framework of science, what warrants of reliability could it offer? A brief inquiry into how observational studies of multifactorial epidemiology are conducted will clarify this point.

Epidemiologic Studies

Epidemiologic studies require different structures [8] to address different survey opportunities, but in general risk is measured as the rate of disease incidence of exposed subjects, relative to the incidence rate of unexposed subjects, the latter usually being defined as the control group. Thus the risk of exposed subjects is called a relative risk (RR)[9]

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 subjects, due to multiple background causes operating in conjunction with, or entirely separate from the exposure under study. Therefore, risk

---

[8] For exhaustive information consult the textbook by Rothman and Greenland (Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Lippincott Williams & Wilkins. Philadelphia, 1998). Cohort studies are utilized to observe differences of disease frequency in groups (cohorts) of people exposed or not exposed to possible hazards. Cohort studies can follow a group of subjects over time (longitudinal studies) or simply at a particular moment in time (cross-sectional studies). Case-control studies are utilized when it is only feasible to observe differences of exposure to postulated hazards 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 potential hazards, 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 hazards. Both prospective and retrospective studies may identify different levels of exposure, 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 experiences in groups with or without a specific disease. Because disease incidence is 0% in the controls and 100% in the cases, a key understanding is that in case-control 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 if found to be higher among controls. Other distinctions of studies are made. For instance, ecological studies do not compile statistics from individual members of a group, but rather compare overall statistical data of populations against generic characteristics of the same, such a dietary habits, genetic traits, geographic and environmental conditions, and the like. In general, ecological studies produce weaker clues and may call for more specific cohort or case-control studies.

[9] Relative risk (RR) is a most common index in epidemiology, along with several indexes not here illustrated, such as odds ratios (OD), hazard ratios (HR) Standard Mortality Ratios (SMR), and others.

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, the RR is 1 and therefore there is no risk differential. If the RR is above 1 the risk is said to be increased in the exposed subjects, if the RR below 1 the risk is said to be decreased in the exposed subjects, indicating that the exposure under study might be possibly protective.

Epidemiologic studies are affected by similar difficulties of design, data collection, and interpretation. Fundamental obstacles arise especially when attempting to reconstruct past conditions – dietary and body weight recollections, for instance – by asking each study subject to remember variable personal experiences, often over several decades of their prior life. Vague and untestable answers are obtained by in person or phone interviews, or from next of kin, thus resulting in databases fraught with uncertainties of unfathomable dimensions, which are conveniently assumed not to exist. Reckless as this assumption is, such illusory data are nonetheless used and subjected to statistical assessments.

Measurements, biases, and confounders. Disease are more prevalent at different ages, making it necessary to approximate equal age conditions when comparing dissimilar groups: a maneuver requiring 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 very rough approximations.

Biases are common. A selection bias occurs when the non-exposed/ control subjects mismatch the exposed/case 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 cases with a disease are apt to recall more intense and longer exposures than the controls without the disease. Recall bias and error may exacerbate when exposure information is retrieved from next of kin of deceased subjects. Exposure reconstruction from other sources may also be biased.

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 (causes?) that could also participate in an association. For instance, body weight data might be confounded by the appetite –reducing (anorectic) effects of cigarette smoking. Methods are used to uncover and reduce the effect of possible known confounders, but those effects can only be partially controlled because of inherent uncertainties, whereas no control is possible for hidden and unknown confounders.

For instance, studies dealing with overweight should consider several risk factors as potential confounders reported in the literature: for instance, studies of cardiovascular conditions face over 300 published accounts of risk factors as potential confounders. Without a credible control for at least the known confounders, epidemiologic studies could not be credibly interpreted.

Statistical error. Large as they might be, epidemiologic surveys usually sample only small fractions of populations. This raises the possibility of statistical error, the magnitude of which is inversely related to the number of subjects in a given study. Statistical error is characterized by tests of significance defined as p-values or as confidence intervals (CI). Both relate to a predefined and arbitrary level of error that is considered acceptable, the usual convention being a 5% error (i.e. a 1 in 20 chance of error). [10] Detailed accounts of epidemiologic methodologies for

---

[10] Under such a consensus definition the threshold of statistical significance is p=0.05, that is a 1/20 value. Thus, a relative risk or odds ratio of 1.6 with a p=0.03 is said to be statistically significant at the 95% level, whereas a relative risk of 2.3 with a p=0.07 is not.
The confidence interval is more informative than p values, although based on the same concepts. In standard format it gives a range of values within which the value of a relative risk or odds ratio may be located with a probability of 95%. 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, or because exposure is the same in cases and controls. Values below 1 imply risk reduction or protection, values above 1 imply increased risk. Therefore:

RR = 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 = 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 = 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.

dealing with biases, confounders, standardizations, and statistics can be found in textbooks. [11] Still, it should be useful to acquire a perspective on statistical significance. The conventional 1 in 20 threshold of acceptable error 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 error probabilities. Margins of error for airplanes need to be extremely low. A typical car engine that has run for 100,000 miles has performed without mechanical error for over 2 billion revolutions. Several million transistors on a 1 square inch chip of silicon must perform flawlessly for years. As the basis of all atomic and chemical interactions, Feynman's quantum electrodynamics predictions are accurate to some 11 decimal places, or to a level of error of less than 1 in 100 billion. [12]

An even more perplexing realization is that statistical elaborations presented in most epidemiologic studies are based on assuming that the original data on which statistical analyses are performed are reliable, accurate, and objective measurements of real conditions. In fact this assumption is wholly unwarranted, given how much of the original data are obtained through manifestly unreliable individual recalls that guess compositions and amounts of lifetime diets, or levels of exposure to possible hazards. Such guesses are not even comparable, because different respondents use different recall metrics: what may be a large portion for one person could be small for another, and the same amount would be reported with different magnitudes. Thus, for many epidemiologic studies the inescapable conclusion is that the statistical elaborations offered are more often figments of imagination, a conclusion that generally holds true for either studies that may or may not support a particular claim, such as the presence or absence of an obesity epidemic.

Combining 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 clinical trials, a use that remains its legitimate application. However, using meta-analysis for pooling the results of structurally different observational studies is fraught with irresolvable difficulties.

In epidemiologic practice, meta-analysis gives different subjective weights to different studies. It does not pool the actual data of each study, but only the final

---

RR = 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.

[11] Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Lippincott Williams & Wilkins. Philadelphia, 1998.

[12] Feynman R. QED, Princeton University Press, Princeton, NJ, 1985.

conclusions of each study regardless of whether concordant or discordant, credible or not. The procedure does not discriminate for characteristics of each study, such as 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. Characteristics other than study size are commonly used in weighing studies, study quality being the most used despite being discretionary, judgmental, and conducive to different meta-analysis results at the hands of different analysts.

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." [13].

Therefore it should be manifest that meta-analysis can be used in epidemiology as a stratagem to contrive meaning from studies that have no apparent meaning. More importantly, the just quoted excerpt from the Federal Reference Guide on Epidemiology leads to a general but crucial warning in reading and interpreting epidemiologic reports. Most numbers in epidemiology are metaphorical proxies of uncertain real quantities, for epidemiology rarely measures reliably, and more commonly , guesses, conceives, sizes up, and appraises, all within a heavily subjective and judgmental framework.

Indeed, statistical 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 more often the illusory summaries of metaphorical primary data. Indeed, the greater the complexity of the statistical analysis in epidemiologic reports, the greater the weakness of the data is likely to be. Known as data dredging, epidemiologists like to conjure every conceivable signal out of what is usually a confused congeries of data. How do epidemiologists approach the inherent fragility of their data?

---

[13] Green MD, Freedman DM, Gordis L. Reference Guide on Epidemiology. Reference Manual on Scientific Evidence. Second edition. Federal Judicial Center, Washington DC.2000. p.381

Epidemiologists And Uncertainty

Epidemiologists react to uncertainty by taking contrasting positions. A few may focus on the collection of specific and accurate data, and on controlling as best as possible for biases and confounders. This concern may be reflected in cautious, balanced, and truthful representations of epidemiologic uncertainty, which generally receives less than enthusiastic attention from the media, advocates, and policymakers.

The opposite happens for a majority of epidemiologists and long tradition of advocacy that views epidemiology as a fungible tool for promoting the financial interests of the profession or political agendas. In fact, it would be wholly unreasonable to expect epidemiologists to lead in the criticism of epidemiology and in waxing forthright about its shortcomings and uncertainties. Leveraging on easily stoked public anxieties, academic departments and government agencies become addicted to lavish public founding, whose continuation and expansion would suffer if epidemiologists were to be openly and honestly critical of their results. Often these pressures exacerbate when they are coupled with authoritarian instincts that the French aptly characterize as dirigisme.

Not surprisingly, epidemiology may be the only discipline where the presence of multiple studies with opposite outcomes is often construed as proof that a coveted hypothesis is correct. In too many instances advocacy prevails over interpretive restraint, leading to public messages and policies of questionable lineage, but that the media love. 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." [14]. The circularity of such a justification is manifest, when realizing that too many exigencies of public health are 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 is in vogue, one that shows little patience with the scientific method, while being still reluctant to be perceived as non-scientific. Its proponents claim that "epidemiologists among others have been misled by standard interpretations of the nature of science" and therefore "to control for confounders…strips away the essential historical and social context, as well as the multiple moderating influences than constitute true

[14] Rothman KJ. Modern Epidemiology. Little Brown & Co. Boston. 1986.

causation." [15]. For those proponents, causal agents are "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." [16].

A novel methodology also is sustained, 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." [17] The new paradigm focuses on broader historic, cultural, socioeconomic, and political determinants of health and disease 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." [24] 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." [22]

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 by internal consensus. In effect, this prevailing epidemiologic paradigm presents an alternative epidemiology much in the spirit of alternative medicine. The alternative leans on the meek claim that nothing better is available, proposing that good intentions alone justify the imposition of creative and usually interested conjectures. Extreme advocacy also scorns the undeniable truth that a number of questions may have no ready answers.

The counterpoint of a few 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." [18] Other sensible critics of the proposed paradigm also contend that such loose thinking invites

---

[15] Susser M, Susser E. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. Am J Publ Health 1996;86:674-677.

[16] Pearce N. Traditional epidemiology, modern epidemiology, and public health. Am J Publ Health 1996;86:678-683.

[17] Pearce N, McKinaly JB. Back to the future in epidemiology and public health: Response to Dr. Gori. J Clin Epidemiol 1998;51:643-646.

[18] Feinstein AR. Epidemiologic analyses of causation: The unlearned scientific lessons of randomized trials. J Clin Epidemiol 1989;42:481-489.

excessive reductionist assessments, which generate "the illusory comfort of perhaps metaphorical meta-theories that appear to explain everything while accounting for nothing." [19]

Both conservative and advocacy parties in epidemiology concur that causal theories is what they produce, but they differ in interpretive restraint. The distinctive characteristic of advocacy is a determination to intrude sociopolitical views in epidemiology, with the claim 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 the approximation of such truth, which is needed for responsible and justifiable public health actions.

How is epidemiology interpreted when scientific validation remains elusive?

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 hazards. In the US, 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." [20] In this quotation, scientific expertise refers not only to the use of statistics in analyzing the data, but also to deny 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 US Surgeon General. However, the important distinction is that during the fact finding phase it should be a judgment of cumulated evidence and not a judgment of conditional socio-political values. In confronting uncertainty, it is common epidemiologic practice to draw judgmental causal inferences on the basis of the Hill criteria that were previously mentioned [21]:

---

[19] Davey Smith G. Learning to live with complexity: Ethnicity, socioeconomic position, and health in Britain and the United States. Am J Publ Health. 2000;90:1694-1698.

[20] Green MD, Freedman DM, Gordis L. Reference Guide on Epidemiology. Reference Manual on Scientific Evidence. Second edition. Federal Judicial Center, Washington DC.2000. p.375

[21] Hill AB. The environment and disease: Association or causation? Proc R Soc Med 1965;58: 295-300.

- **Strength**. The strength of an association is a clue to causation, but a strong association is neither necessary nor sufficient to affirm causality, and a weak one is neither necessary nor sufficient to deny causality.
- **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 consistently inconsistent are unlikely to be interpretable either way.
- **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.
- **Temporality**. Effects must occur after the cause has a chance to act. This is a valid, if self-evident and trivial criterion of causality.
- **Dose-effect relationship,** which 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.
- **Plausibility**. Whether an association is biologically plausible or not remains a matter of individual speculation and is far from being objective or conclusive.
- **Coherence**. Agreement with other information may be a corollary attribute of causation, but conflicting information could be erroneous.
- **Experimental evidence**. Experimental evidence in humans would indeed constitute proof of causation, but it is very rarely available for conditions determined by many possible hazards.
- 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."** [22] **Analogy is an absolutely invalid criterion in a judgment of causation.**

In their textbook of epidemiology, 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

[22] Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Lippincott Williams & Wilkins. Philadelphia, 1998. p.27

evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non." [29]

Therefore, authorities in epidemiology negate causal persuasiveness to criteria that have no power as primary inferential determinants of causality. Indeed, it is remarkably surprising that epidemiologists have not seen fit to include prominently among their criteria of evidence those most important common-sense warrants that might begin to give some measure of confidence in epidemiologic studies: namely that what has been measured is in fact what is said to have been measured with acceptable accuracy and a stated error, that known biases and confounders have been controlled to the best possible and sufficient extent, and that the results are consistently reproducible. **None of the studies of overweight and obesity have met these elementary criteria of reliable evidence.**

Why is epidemiology receiving such inflated attention, despite insurmountable deficiencies? There are of course the usual power games of special interests, but the prime reason for the attention is that epidemiologic studies will be expected, demanded, and done in affluent societies acutely preoccupied with health and death. Epidemiologic studies will inevitably influence public and private policies, raising a critical need for guidance in judging how deserving of attention their results might be.

As a rule of thumb, attention should be proportional to the magnitude and consistency of the risks reported. For instance, the case against cigarette smoking is built on epidemiologic studies consistently reporting average lung cancer risks of smokers that are 10 times higher than for nonsmokers. By contrast, the epidemiologic risks attributed to overweight and mild obesity are in the order of 0.5-higher than the risks of normal weight people, adding that such results are not fairly consistent and are made ambiguous and questionable by a number of unaccounted confounding hazards. For instance, when mortality in relation to overweight and obesity is the issue, there are innumerable causes of death that can impinge on the results, and that the studies could not account for and chose to ignore. Likewise, when cardiovascular disease, cancer, hypertension, and diabetes are considered in relation to overweight and obesity, hundreds of known and potential hazards that contribute to those conditions are routinely ignored by epidemiologic studies. The implicit excuse given is that it would be difficult if not impossible to measure the impact of those extraneous hazards on the meaning o the results, but such a lame excuse simply confirms the lack of credibility of study conclusions. The exception, of course, is for the small fraction of people who are morbidly obese or very thin, and for whom the epidemiologic signals are strong and consistently reproducible.

So far, the discussion has provided a basic overview of the methods of epidemiology, and of its endemic uncertainties. As a strictly rational conclusion, it

is easier to refute than to sustain epidemiologic claims that are not experimental and not scientifically justified. Lacking the immediate persuasiveness of scientific experiments, the credibility of epidemiology suffers a heavy burden of documentation, the reading of which requires a very skeptical mind.