STATISTICS IN THE PUBLIC SPHERE *

Frank van Dun

---

* Author's note: This text is a revised and expanded version of the English translation of a chapter ("Statistiek, wetgeving en beleid", written in Dutch) in the Metajuridica Course Book, which was used at the Faculty of Law of the University of Maastricht when I was teaching there. I wish to thank Francis van Dun, and Wiel Maessen and his collaborators for their translation and comments. Of course, thanks also to my former colleagues at the university, especially Sef Jansen and Han Israel, for their assistance in preparing the original text. They are not to be faulted for any errors or oversights that remain. The original caveat still applies: References to publications in applied statistics are given mostly for the purpose of illustration. Unless there is a clear indication to the contrary, they do not imply the author's endorsement or rejection of reported statistics or correlations.

It starts from a number; then there is a missing link;
and it finishes up with draconian legislation.
–John Brignell

Introduction

It is important that jurists should have some understanding of statistics and the ways they are used in the public sphere, especially in legislation, policy-making and arguments before the courts.

In this text we obviously cannot give a complete course on the science of statistics and its applications. Such an undertaking soon would lead us to the study of mathematics and the theory of probability, and of the various methods of gathering, classifying and evaluating sets of empirical data. All of that is far beyond the scope of juridical training. That training still reflects an era when judges were supposed to do justice in the individual case (which required them to look at specific causes and actions, hardly ever at statistical correlations) rather than make or implement policies designed to have significant statistical effects in the population at large.

Nowadays, the laws judges have to apply and the arguments with which they have to deal, are much more likely to involve statistics than at any previous time in history. Consequently, law students have no excuse for ignorance of some of the elementary concepts and difficulties of the subject, or for naïve credulity in the face of the flood of statistics that overwhelms all public forums, from the legislature to the courts to the universities and the media.

Accordingly, this text is no more than an attempt to cover some basic concepts and principles of statistics and rules for interpreting statistical reports. We shall do this primarily by means of simple fictitious examples. We avoid reference to published statistics, which would require a lot of background information and context, because we are not interested here in judging or evaluating particular studies.

Context: research, the media, and politics

Most people get their science from the popular media, and so do most intellectuals (including scientists when the media report about other fields than their own). However, the media are inclined to emphasise sensational findings that portend drastic changes in the conditions of life and health for many people.[1] They prefer to report these findings as soon as possible without waiting for the outcome of the usually long process of critical review that is the very essence of science. That is understandable. Analysing and evaluating complex arguments and methodologies and organising large quantities of heterogeneous data into informative lists and charts do not sell newspapers or television documentaries nearly as well as hope and fear do. Announcements of 'significant breakthroughs' and, far more often, predictions of 'looming disasters' and grim warnings about the dangers in almost everything we do or use[2], shape the image of science that the media offer for public contemplation.

Not only the news media, but also schools, political and commercial campaigns, fundraising brochures for pressure groups and activist organisations, novels and films help to propagate what has been called a culture of fear[3] and distrust. It is a culture that induces feelings of guilt and impotence and leads impressionable people to clamour for expert guidance and protection against everything and everybody, or to join one or another crusade to put their hopes and fears on the political agenda and enshrine them in constitutional provisions, legal prescriptions, and international treaties.

Political authorities tend to react ambiguously to such crusades and the culture of fear that inspires them. On the one hand, they realise that they are offered an opportunity to justify an expansion of their powers of regulation and taxation as a means to address the concerns of the public; on the other hand, they like to be seen as providential agents, commanding superior knowledge and technical know-how.

Indeed, the expansion of the State in the twentieth century went hand in hand with increased bureaucracy, professionalism and reliance on experts of various stripes and colours. The experts were hired by the State to conduct research, collect and interpret data, and formulate, implement and evaluate policies.[4] Ordinary people were led to believe that the quality of the management of the state is guaranteed by solid science and the application of scientific methods. In today's world, this belief is vital for the legitimacy of the State and the political professions. They cannot afford to see it being eroded by persistent and widely publicised claims that they are unaware of some grave danger or unable to do something about it. From the authorities' point of view, the best strategy is to jump on any bandwagon of fear sooner rather than later and to try to take command of it. They can do this by funding or subsidizing selected research and action programs and appointing trusted experts to

---

1 Ruth Kava, *Good Stories, Bad Science, A guide for journalists to the health claims of 'consumer activist' groups*, American Council on Science and Health, June 2005, http://www.acsh.org/docLib/20050610_consumerActivist.pdf

2 During the first half of 2001, the national newspapers in Great Britain published at least 150 warnings concerning all sorts of food, activities, moods, habits, lifestyles, dwelling places, appliances, medicine, surgical interventions, pregnancies, weather conditions, and what not. Health Wars: The Phantom Menace, Volume 1: An Audit of Health Scares January – June 2001, Compiled by Josephine Gaffikin (FOREST Occasional Paper 3, Audley House, 13 Palace Street, London SW1, December 2001). Similar observations can be made in other countries and with respect to other news media.

3 Frank Furedi, *The Culture of Fear*, Cassell, London 1997.

4 "As late as 1930, no more than a hundred economists worked in the federal government. By 1938, just eight years later, the figure had escalated to more than five thousand, as Franklin Roosevelt established new agencies aimed at overcoming the Depression", quoted from W. Fogel's address to the Association of American Universities Centennial Meeting (Washington, D.C., April 17, 2000). See http://www.aau.edu/aau/Fogel.html. Also, Robert Higgs, *Crisis and Leviathan*, Oxford University Press 1987 for an analysis of the role of ideology in the growth of American Government.

official commissions that are likely to become important sources of media coverage on a particular issue.

Of course, to avoid criticism of intervening ignorantly in questions relating to research, politicians are anxious to emphasise that they act on expert advice—often advice from the same people who are eligible for subsidies and grants for promoting their own lines of research and expanding their own research facilities. Hence, there is the phenomenon of a growing collusion between political authorities, academic elites, well-funded activist organisations, pressure groups and industrial conglomerates with large interests in research and development—all of them trying to manoeuvre their favourite consultants into the right places at the opportune moments.

As a result of this pre-selection process, regarding a number of 'sensitive issues' (acid rain, global warming, BSE, aids, drug abuse, terrorism, and the like[5]) hundreds of millions of euros are spent on researching a single hypothesis. Alternative hypotheses regularly are ignored by the establishment and left to a few individual researchers outside the large, heavily subsidised institutes. These researchers thereby are not only marginalized financially and professionally but also easily stigmatised as not trustworthy. When they receive even a small grant of, say, $10,000 from some interested industrial or commercial sector then they immediately face the charge of not being independent. In contrast, ten or a hundred times that amount—or being permanently on the payroll of a governmental agency (EPA), international bureaucracy (WHO), interest group (Greenpeace) or other foundation[6] that either is State-funded or has a more or less explicit political agenda—does not appear to affect the 'independence' of researchers willing to work on officially approved or 'politically correct' hypotheses.[7]

---

5 Usually such scares start with the publication of disaster scenario's, which are toned down in subsequent years. At the start of the Global Warming Scare in 1990, predictions for 2100 were in the range of 3-5°C. By 1995 they already were down to 1-2°C. If the trend continues, Greenland will have to wait a long time before it can live up to its name again. The name harks back to the medieval warm period. About 1000 B.C. Eric the Red began the Viking colonisation of what he called Greenland, to attract farmers and cattle raisers. The Vikings left in the course of the fourteenth century when falling temperatures shortened the growing season. The world (at least the northern hemisphere) was on its way to the so-called Little Ice Age, which peaked in the seventeenth / eighteenth century but may have lasted well into the nineteenth century.

The original predictions of a devastating aids epidemic turned out to be wildly exaggerated, except perhaps for Africa where other criteria and cheaper and less reliable methods were used to identify aids—which raises the question whether 'aids in Africa' is the same condition as aids in Europe or North America. The same goes for the predicted BSE-related epidemic of vCJD. 'Acid rain' seems to have lost its lustre as an environmental disaster, as does the 'hole in the ozone layer'.

6 Wealthy foundations, such as the Rockefeller, Ford and Carnegie Foundations, have been active for a long time in the fields of militant science to promote their agenda of social change not only in the U.S.A. but elsewhere as well. R.A. Wormser, *Foundations: Their Power and Influence*. Devin-Adair, New York 1958 (reissued 1993).

The *Environmental Protection Agency* (EPA) had a 1999 budget of 7.6 billion dollar. For the *World Health Organisation* (WHO) the corresponding sum for 1997-1998 was 842 million dollar. According to V. Ekamov ('Reform of the WHO', *The Lancet*, 347, juni 1996, p.1536-1537) WHO spends 75% of its budget on administrative overhead. The organisation also receives many hundreds of millions of dollars from 'private donors' (as voluntary contributions from member states, especially the U.S.A., and institutes, charities and corporations, primarily in the pharmaceutical sector). J.P. Vaughan, S. Mogedal, S.E. Kruse, K. Lee, G. Walt, K. de Wilde, 'Financing the World Health Organization: global importance of extra budgetary funds' *Health Policy*, 35, 1996, p.229-245. For 1998 Greenpeace International reported a net income of 101 million dollar. (The reader is invited to check the official web sites of these and other organisations as well as www.activistcash.com.)

7 Some industries (pharmaceutics, medical and environmental technology, among others) have partnerships with these organizations, which regularly publish studies that have been financed wholly or in part by the industries concerned. Commercial reasons—promoting medical products or technology among the professional readers of their publications—may be obvious, although they never are stated up front. Other reasons may be more sinister: to divert attention from or marginalize research that links such products and technologies (disinfectants, radiation) to certain diseases. John W. Gofman, Radiation from

Universities often are willing to tailor their programmes to the requirements of the markets for expertise with no particular concern for scientific integrity and truth. Today, many holders of bachelor's and master's degrees expect to find employment as researchers and consultants for governmental departments and non-governmental organisations (NGOs) whose main business is exerting some measure of control over one or another aspect of policy and legislation. In the competition for scarce funds only research that appears to support the aims of the suppliers of funds will do well. Science is not for sale, but research often is.

In these circumstances it is hardly surprising that much of the research that is reported in the media is part of the logistics of political and public relations campaigns. However, it is difficult for the general public to distinguish it from genuinely scientific research. While some people tend to accept every publicised finding or theory at face value and others indiscriminately reject the corn with the chaff, many are willing to let their opinions be determined by those who most effectively play on their hopes, fears, and prejudices. Knowledge is power—and so is the pretence of knowledge.

Statistics in public life

Statistics—from simple counts, crude estimates, and random polls to intricate measurements, and computations based on the results of these—undoubtedly are essential tools in many genuinely scientific endeavours. However, they now also are important tools of political propaganda and policy-making, which aim to induce or compel people to change their ways of life, their habits of action and thought. Indeed, the word 'statistics' originally belonged to the political sphere. It came in vogue no earlier than the eighteenth century, probably in Prussia. It denoted 'things relative to the State', in particular the rapidly increasing mass of figures, names, measurements, surveys and other data that the absolutist States of that period were collecting through their expanding armies of civil and military bureaucrats.

It was only later that the term passed into general use and its primary meaning shifted from the *purpose* of data management (assisting the State's rulers and administrators) to the *techniques* used in managing numerical data and extracting relevant information from them. Nowadays, the term 'statistic' evokes primarily the image of one or more series of numbers presented either in tabular form or in the form of a graph. The plural form 'statistics' is also used to denote techniques of mathematical interpretation applied to the study of phenomena, especially phenomena for which only incomplete, ambiguous or indirect data are available.

Almost all of the dire warnings that feed the culture of fear rely on statistical studies; only rarely do they refer to clinical or controlled experimental research. However, it is not until statistical studies have been supported by such research and have passed the test of scientific criticism that one should presume to speak about causes or objective relationships in the real world. By themselves, statistical correlations among various sets of 'data' hardly ever constitute scientific proof of a really existing relationship. Yet, the culture of fear thrives on the publicity the media give to 'facts' that in reality are no more than statistical correlations. True, most of these correlations are published in one or another scientific or academic journal or report, but publication does not mean either corroboration by independent studies or the ability to withstand serious criticism. Publication is not proof. Even publication in a peer-reviewed journal is no guarantee of quality, for example if the editors (who select the

---

Medical Procedures in the Pathogenesis of Cancer and Ischemic Heart Disease, Committee for Nuclear Responsibility Book Division, San Francisco, 1999. Given the costs of litigation (especially in the American culture of litigation, with its predilection for class action suits), medical and pharmaceutical companies are willing to invest a lot of money in fostering good relations with EPA, WHO, other public health authorities, and universities and other institutes of research. Having lost their financial independence long ago, universities now are encouraged to seek ad hoc funding for their research—which may include such academic malpractice as producing research that meets the client's desire for a particular finding. Individual researchers have an opportunity to endear themselves to particular clients and pursue a career outside the university as consultants or trusted experts.

reviewers), publishers, or their financiers subscribe to a mistaken theory or use their position to promote their own pet ideas to the detriment of other points of view.

Statistics play an increasingly important role in the life of individuals, groups, organisations, and nations. School reports and other evaluation forms are familiar statistics. They serve as arguments in decisions to assign persons to one place or another in school, workplace or society, to praise or condemn them for their performance, to reward or to punish them. Statistics appear in governmental or corporate policy statements to justify giving priority to one goal rather than another. Statistically identified risks are behind all sorts of standards and norms written into contracts or imposed by legislative and administrative bodies that presume to tell us what is required for health, hygiene, education, safe working conditions, building practices, driving, flying, sound financial operations, investments, and what not. Being just above or under the stipulated norm may have far-reaching consequences, financial or otherwise, for thousands or millions of people. Products, services, careers and activities are taxed and regulated, even banned, if there is an indication or suspicion that they fail to meet some statistically established norm. On other occasions failure to meet the norm is cited as a reason for being granted special privileges or subsidies.

When so much is at stake in the use of statistics one can be sure that many people will be concerned far more with getting the right statistics than getting their statistics right. Before the court of public opinion, propagandists and advocates cannot fail to appreciate how easy it is to put a particular spin on numbers to evoke a desired response; find a self-declared expert willing to make their case; continue to use long-discredited statistics as if they were scientifically established fact; appeal to opinion polls to stigmatise dissenters as isolated cranks; or distort the boundaries between disciplines to dismiss opponents as unqualified interlopers.

In big cases before the law courts, where huge financial claims are at stake, lawyers and expert witnesses often juggle with statistics of dubious quality. They know that it is difficult if not impossible for a judge to assess the merit of various strands of conflicting and often almost impenetrable technical evidence. Many potential targets of such litigation—especially the so-called *deep pockets* (corporations, hospitals, councils, large firms, and their insurance companies)—prefer to surrender to baseless demands or settle out of court because it is cheaper than being drawn into a long legal battle with no certain outcome except a lot of adverse publicity in the media. Thus, the mere threat of litigation becomes a lucrative business for lawyers, forensic experts and consultants and a useful tactic for activists who appreciate the publicity value (and the limited risks to themselves) of taking on powerful opponents and maybe bringing them to their knees. In a similar way politicians and civil servants can pressure companies, industries and communities to adopt certain policies by threatening them with regulations and standards that have little or no scientific base. Such cases of "legal extortion" are far from rare, both domestically and internationally.[8]

Of course, there are critical voices that denounce the abuse of the trappings of scientific research, and of statistics in particular, as veils for special pleading, fanciful or dishonest claims, or power grabs. Accusations of incompetent, misleading or fraudulent use of statistics are rampant.[9] The term 'junk science' is already commonplace in discussions on food hygiene, health care and the environment.[10] It is beyond doubt that in many cases recourse to

8 An economist at the World Bank: "Policy based lending is where the bank really has power--I mean brute force. When countries really have their backs against the wall, they can be pushed into reforming things at a broad policy level that normally, in the context of projects, they can't." (Quoted in Kamran Abbasi, 'The World Bank and world health: healthcare strategy', British Medical Journal, 318, April 3 l999, p.933-936.)

9 For example Hans-Joachim Maes: 'World Health Organization ... mit Entsetzen Scherz.', Deutsches Ärzteblatt, 98/25, 22 june 2001, p.1664-1666.

10 There are several individuals and organisations that report on junk science, and malpractice and fraud in the scientific community. Among them, The American Council on Science and Health (www.acsh.org),

statistical data and methods is proper, illuminating and useful, if the data and the methods meet the rigorous standards upon which genuine science insists. Unfortunately, that is not always true—and that is putting it mildly.

Still laymen—and, where statistics are concerned, most people are laymen, including most academics and many scientists—are easily impressed by numbers, percentages, risk measurements, statistically significant correlations and the like. Politicians, administrators, judges and lawyers are no exceptions. That is not a good thing. They regularly deal with legal rules, policies and arguments in which statistical data and computations, theories and conclusions play prominent roles. If they are not capable of distinguishing between good and bad statistical research, and between proper and improper uses of it, then they run the risk of propagating nonsense and doing much harm. Thousands, often millions of people may be victimized, physically or financially, by national, international or supranational governmental and advisory bodies that are prejudiced or incompetent in their use of statistics.

That risk is real. In 1991, Greenpeace and the Environmental Protection Agency used their influence and pressure tactics to stop the chlorination of drinking water in Peru. This measure immediately led to a cholera epidemic. At least eight hundred thousand people were taken ill, of which at least six thousand died of cholera. The policy was based on Greenpeace's *idée fixe* (at that time) that chloride is toxic in any concentration or dose and on EPA's classification (based on epidemiological statistics) of chloride as a risk factor of cancer.[11] Admittedly, chloride is strong stuff; it corrodes almost all metals. However, it also is a component in hundreds of substances that are vital for plant and animal life, for example salt and human digestive juices. It is well suited for decontaminating drinking water.

Another example from an entirely different field: in the twentieth century, economic and monetary policies, often based on a combination of outdated or impressionistic statistics from diverse sources[12], have more than once brought severe financial harm to millions of people. This is to be expected if a handful of persons (central bankers, the never more than a few political heavyweights that actually make national policy, and their trusted advisors) have power to make decisions that reverberate far and wide. Their personal errors are the stuff from which national and international economic and political crises are made.

Although episodes such as these invite questions about responsibility and accountability, civil and political authorities (including the courts) are reluctant to take them up. They often prefer to bow to the prestige[13] of statistical evidence and science, turning it into a veil behind which irresponsibility can flourish with impunity. Indeed, legislation and government policy *are* officially recognized[14] methods for externalising the effects of one's actions, that is for achieving one's goals at the expense of non-consenting others. Thus, there are enormous benefits to be reaped from having or peddling influence in shaping and implementing legislative action and public policy. If to the outside world the policy must be made to appear as scientifically validated then that can be arranged. Again, science is not for sale, but the mantle of science often is.

Steven J. Milloy's www.JunkScience.com, Dr. Barrett's www.Quackwatch.com, and John Brignell's www.numberwatch.co.uk.

11 Nature (November 28, 1991), U.B. Panisset, International Health Statecraft: Foreign Policy and Public Health in Peru's Cholera Epidemic (Lanham MD: University Press of America, 2000). According to Greenpeace (Archive.greenpeace.org/~toxics/reports/cholerachlorine.pdf), the epidemic should have been prevented by boiling water and improving personal hygiene!

12 See for example Richard Vedder, *Statistical Malfeasance and Interpreting Economic Phenomena*, in: The Review of Austrian Economics. Vol. 10, No. 2, 1997, pp. 77-89.

13 Prestige, from Latin *praestigiae,* conjuror's tricks.

14 'Officially recognised' means 'recognised by legislation or government policy'.

Things and numbers

Statistics is first of all a mathematical discipline. It concerns itself with the study of formal relations between sets of numbers. However, the application of statistical methods involves much more than mathematical proficiency. People, animals, material things, and their properties are not numbers--and there often is no obvious way to number them. Take any everyday object, say a house. How many houses are there in a particular village? Different observers will arrive at a different numbers, if they do not use the same criteria for identifying marginal or other dubious cases, such as huts, tents, caravans, bungalows, hotels, storage buildings, and derelict buildings; or if they do not apply the same criteria in exactly the same way—for example if one observer assumes that a thing is not a house if no one actually lives there and another assumes that it is a house if someone lives in it. Looking at aerial photographs and walking along the streets and roads of a province are different methods of counting houses that are likely to give different results. Moreover, not all of those who do the counting are equally diligent or alert: maybe there are houses (however defined) that a more careful researcher would have seen but the actual researcher did not notice.

If merely counting things often is problematical then processing the results of a count also may be a problem. If one had data on the total number of houses in every Belgian province then it would appear logical to add up these numbers to get the number of houses in Belgium. However, it would not be logical to do so if there were no assurance that the number for any province was obtained in the same way and at the same time as the numbers for the other provinces. Data for one province relating to the nineteen-nineties obviously may not be comparable to data gathered in the seventies for another province.

Problems multiply if instead of counting the houses themselves one has recourse to indirect sources (tax forms, telephone directories, data obtained from firms in the construction or repair business). Even less reliable are sources such as reports from memory; reports of numbers with no indication of how they were obtained; or impressionistic sources such as paintings or maps. Yet, no other data may be available, either because they are all that survives from the period under investigation (How many houses were there in 1800?) or because the researchers lack the time or the money to do an actual count.

Suppose that instead of houses we want to count things such as the incidences of a disease, mood, talent, or other condition that is difficult to identify or quantify. How many people suffer from angina pectoris? How many are manic-depressives? How many are smokers? How many use illegal drugs? How many are poor? How many are 'socially challenged'? How many Roman Catholics are there? How many are socialists? Clearly, there is no obvious way to count such things. Before one can number them for statistical purposes one must introduce various theoretical notions and assumptions and conventions that may be more or less controversial. Consequently, the numbers themselves are essentially controversial.

Representative samples

Often it is impossible to count or investigate all the things that satisfy a particular condition. Then researchers will look for a subset that is somehow 'representative' of the total set (or 'population') of things. With a representative sample they can say with some confidence that if 35% of the subjects in the sample have a particular characteristic then approximately 35% of the general population have that characteristic. Obviously, if the data are taken from a sample of persons then we should want to know how large the sample is relative to the *relevant* total population,[15] how the elements in it were selected, whether it is sufficiently

[15] Consider an investigation of the risk that people exposed to A will suffer from a disease B. The relevant sample is the number of people with B, which the researchers have identified; it is not the number

similar to a larger group or category of people to warrant making statements (extrapolations) about the larger group or category, and so on. Does anyone really believe that Alfred Kingsey was entitled to make general statements about the sexual behaviour of 'the human male'[16] on the basis of data collected from his research subjects (a great many of whom were prison inmates, including a sizeable quantity of sex offenders) or supplied to him by volunteers (many of whom were later identified as male prostitutes, perverts and paedophiles[17])?

Obviously, results obtained by experiments on animals should never be extrapolated to human populations[18] unless there is solid evidence that the tests affect biological mechanisms that are essentially the same in human beings and in the animals used for testing. Not only does a set of animals *not* constitute a representative sample of the human population, the animals used in laboratory tests frequently are bred or selected especially for their susceptibility to certain diseases. One reason for this practice is that the researches need to have large quantities of *diseased* tissue or organs for clinical analysis.

On other occasions researchers often will expose laboratory animals to *immense doses* of a pathogenic substance or factor (for example radiation). They do so because they are not interested in the epidemiology (see below) of a disease or condition but in the physical or chemical properties of particular biological processes. From the fact that a rat or monkey dies as a result of deliberate exposure to a massive dose of X, one should not infer that X is a risk to human health. Every substance is a poison when administered in large enough doses or sufficiently high concentrations. Lower doses may be quite harmless and in some cases beneficial. Being buried under tons of sand probably is lethal; walking on the beach on a windy day is not. Yet, that point easily gets lost when some less familiar substance (dioxin, PCB, radon) is mentioned. An unrepresentative sample of subjects vitiates any attempt to extrapolate test results to the population about which one wants to learn. So does an unrepresentative sample of exposures. Applied statistical research cannot be better than the data with which it works. That is the GIGO principle: "Garbage in, garbage out".

Averages: meaning and relevance

Once numbers are available we can use them for all sorts of mathematical calculations: simple addition, subtraction, multiplication and division, and other more complex operations. However, we should be careful to keep the mathematics apart from the things our numbers supposedly represent. As a result of a calculation we may find for example that in our city last year 3.3 people were killed per month in road accidents; or that every Dutch family has 1.92 children. Now, we easily can imagine three or four people killed, or families with one or two children—but 3.3 dead persons or 1.92 children? The result of calculations based on tallies of real things need not correspond to a tally of real things.

Of course, we all are familiar with the notion of an arithmetical average (a simple but important statistical concept). We understand the number 3.3 in the example above as follows: last year forty people died in road accidents because forty in a twelve-month period (a year) exactly correspond to an average of 3.3 per month. This mathematical truism and the

---

of people whom they have asked whether they suffer from B. Suppose 1800 people have been asked, and 50 have responded affirmatively, then the relevant sample for the investigation is 50, not 1800. Unfortunately, some researchers prominently display only the larger number, thus making it appear that they used a representative sample. Some will even keep the absolute numbers out of sight, while highlighting only percentages, fractions and other relative numbers. For example, they will report "60% of the B-patients were exposed to A" rather than "Of the 50 B-sufferers, 30 were exposed to A and 20 were not". Such practices are totally inadmissible. They are, however, quite common in media reports.

16 Alfred Kingsey a.o., *Sexual Behavior in the Human Male*, Philadelphia 1948.

17 Judith Reisman & Edward Eichel, *Kindey, Sex and Fraud* , Lafayette LA, 1990; Reisman, *Kingsey: Crimes and Consequences*, Institute for Media Education, Crestwood KY, 2000.

18 Some years ago the expression 'Frankenstein food' appeared in de media following a report on the effects of genetically manipulated food in an experiment on five rats. Zie o.a. The Bowditch Group Electronic AgBiotech Newsletter (August 13, 1998).

statistical average to which it gives rise are relevant, if there were, say, between zero to eight people killed on the road in any given month last year. However, we certainly should question the relevance of the average, if all forty people died in the same accident (they were passengers on a bus that skidded off the road into the river) or on the same day (when there was unexpected heavy rainfall on frozen ground). Dealing with rare extreme conditions is a problem all statisticians face when they use the concept of an average—and they may not agree on the criteria for identifying rare or extreme conditions.

Familiar as we are with averages, some people nevertheless are prone to misinterpret them. It makes sense to speak of averages only if there are things above average and other things below average. There are no relatively poor people ('below average') if there are no relative rich people ('above average'). Even in Billionaires' Grove some are relatively poor. Consequently, it makes no sense to *complain* that some people are relatively poor unless one is a rabid egalitarian who wants all people to be equally rich… or equally poor. An average is just a number; there usually is no reason to consider it a norm or standard of great ethical, political or other significance.

To summarize

- To do statistical research regarding some aspect or part of the real world we need abstract representations (numbers) of the things that are of interest to our inquiries.
- The application of statistics to things that cannot be counted or measured is impossible—but one can always *assign* them some number and *pretend* that it represents a good enough measure. Where there is no straightforward numerical representation of real entities or their properties the researcher must bridge the gap with theories, assumptions and decisions, all of which may be more or less controversial.
- Collecting data or information about some aspect or part of the real world is a tricky business. In many cases one only can gather data relating to a small but representative sample of things. However, there often is no obvious way to determine whether one is justified in treating the sample as representative of the whole. Here too, theories, assumptions and decisions are required.
- Even if a set of numbers has a straightforward interpretation in terms of real things or properties, it does not follow that the number that represents the result of a calculation on the set is meaningful under that interpretation. Given a series of numbers, we can always calculate their average value. That does not guarantee that the calculated average itself is a meaningful or relevant value.

Correlations

People are not immortal. However, some causes of death are more frequent than others. Progress in the treatment of certain diseases or in protecting people against accidents of a particular sort does not change that. There always will be a 'most common cause of death'. We cannot reduce the chance that someone dies, say, of cancer without at the same time increasing the chance that he dies of another cause. Every success in the treatment of cancer is a risk factor (see below) for other causes of death. Suppose that, compared to thirty years ago, relatively more people die as a result of illness. Is that a reason for alarm? Not necessarily, if it means that relatively less people die in accidents or wars.

Here we meet another important statistical concept: correlation. Particular conditions (illnesses) and events (accidents) are correlated with a particular outcome (death). We must take a closer look at this concept because it is both central to most statistical research and easy to misunderstand. A great many of the scare stories that are ever-present in the media and other public forums derive their effect from widespread misunderstanding of this concept.

Whenever we have access to two series of numbers—it does not matter where we got them or what they represent—we can compare them and note the degree to which they are

similar or dissimilar. We do this by calculating a correlation coefficient (CC), which is a numerical value in the range from –1 to +1. We shall not concern us with the formula for calculating correlation coefficients. Our only interest is in the coefficients themselves. The following extremely simple examples serve to illustrate their meaning. We compare two ordered sets of numbers, set X and set Y, each of which consists of four numbers.

X1	**1**	**1**	**0**	**0**	**CC**
Y1	**1**	**1**	**0**	**0**	**1**

Y1 and X1 are perfectly identical. We see that CC has the maximum correlation value, +1. The two series are perfectly correlated.

X1	**1**	**1**	**0**	**0**	**CC**
Y2	**0**	**0**	**1**	**1**	**-1**

Y2 clearly is not identical to X1; on the contrary, Y2 is so to speak a negative image of X1. To a 1 in X1 there corresponds a 0 in Y2, and vice versa. In short, there is an obvious difference between the two series but the difference is perfectly systematic. The CC-value here is the minimum correlation value, namely –1: the two series are perfectly but negatively correlated.

X1	**1**	**1**	**0**	**0**	**CC**
Y3	**1**	**0**	**1**	**0**	**0**

In this third example there is no correspondence whatsoever. Sometimes a 1 in X1 corresponds with a 1 in Y3, and an equal number of times it does not. The same is true for the zeros. The CC here is 0. It indicates no systematic correlation whatsoever between the two series of numbers.

We should not jump to the conclusion that CC=1 obtains only if the two series are identical. Perfect correlation is not the same as identity. Consider the next example:

X2	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**CC**
Y4	**12**	**12**	**12**	**12**	**12**	**5**	**5**	**1**

Although X2 and Y4 clearly are different, their CC has the maximum value +1. Indeed, a 1 in X2 unfailingly corresponds with a 12 in Y4; and a 0 in X2 always corresponds with a 5. Note that 1>0 and that 12>5: with a *higher* value in X2 there corresponds a *higher* value in Y4.

X2	**1**	**1**	**1**	**1**	**0**	**0**	**CC**
Y5	**12**	**12**	**12**	**12**	**25**	**25**	**-1**

Here too there is a perfect but negative correlation. The difference with the previous example is that here the *higher* value in X2 (1) corresponds with the *lower* value (12) in Y5.

Let us now consider some examples that are a bit messier than the first five:

X2	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**CC**
Y6	**12**	**18**	**22**	**15**	**29**	**6**	**3**	**0.8**

(CC=0.7953502167) Here the correlation of the two series is not perfect but still pretty high. We see that with the higher value (1) in X2 there corresponds a higher value (12 to 29) in Y6. The lower value (0) in X2 corresponds with a lower value (3 or 6) in Y6.

X2	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**CC**
Y7	**12**	**0**	**2**	**5**	**12**	**25**	**-25**	**0,2**

(CC = 0.1959828375) It would appear that there is no correspondence between the two series. Indeed, with CC=0.2 we are close to CC=0, which stands for 'no correlation

whatsoever'. However, there still is a weak correspondence in that two of the 1's in X2 correspond with the same value (12) in Y7.

Finally, consider this pair of series:

X3	-7	2.5	800	12	45.04	151	-8	CC
Y8	0.25	0.56	5	1	3.27	4.56	0.12	0.7

(CC = 0.7341022013) Again it may appear that there is no or at best a weak correspondence between the two series. However, the relatively high CC-value (0.7) indicates the contrary. Indeed, to a higher value in X3 there invariably corresponds a higher value in Y8. Moreover, the highest value in X3 (800) corresponds with the highest value in Y8 (5); the second highest value in X3 (151) corresponds with the second highest value in Y8 (4.56); and so on. The lowest value in X3 (-8) corresponds with the lowest value in Y8 (0.12).

To summarise

- A high *absolute* value (positive or negative, approaching +1 or -1) of the correlation coefficient signifies a high degree of correlation between two series of numbers. A low absolute value (in the neighbourhood of 0) signifies a weak correlation.
- Correlation refers to series of numbers. What the numbers mean, if in fact they mean anything at all, is irrelevant for the purpose of establishing the presence or the lack of correlation.

Applied statistics

In doing applied statistics we must take into account what the numbers stand for. Let us consider again the two series X2 en Y6.

X2	1	1	1	1	1	0	0	CC
Y6	12	18	22	15	29	6	3	0.8

This time, we shall *interpret* the numbers—that is to say, treat them as symbols that refer to some really existing things or events that actually happened. For example, let us suppose that they represent the results of research in the seven towns that make up a particular region in the country. For every town we have the number of times a fast medical response unit was called to the scene of a traffic accident during the first six months of last year. These numbers are noted in Y. In X we have a '1' if the town has a mayor who is male and a '0' if the mayor is a woman. (Think of X as giving the number of male mayors in a town.).

Because the CC-value is 0.8 there appears to be a relatively strong correlation between the sex of the mayor and the number of *serious* road accidents (measured indirectly by the number of interventions of a medical response unit). However, for all we know, the correlation is merely statistical—a correlation of two series of numbers.

I have constructed this example in part to avoid complicated questions concerning the correctness or completeness of the data. The sex[19] of a mayor is almost always beyond doubt and medical interventions at the scene of a traffic accident are carefully recorded and easily separated from other calls. Since the research covered all the towns in the region, there also is no question about the representative quality of the data used for the entire region. Of course, none of this tells us anything about the degree to which these towns are comparable in size, traffic level, traffic infrastructure, commercial or industrial activity, and so on. If we should find that the towns with a male mayor coincidentally experienced an unusually high number of exceptional traffic situations (road works, sporting events, festivals, and the like) during the investigated period then that might go a long way towards explaining the difference in the numbers. Thus, the assumption that our data are correct and complete still leaves moot whether they are relevant for other purposes than merely to describe what happened during a particular period.

With respect to our example, what does the high correlation coefficient, CC=0.8, prove? Nothing! It merely gives us a value for the correlation between the two series of numbers X and Y. The numbers themselves are part of an abstract description of a situation that may or may not call for an explanation. A high correlation may be pure coincidence or the result of a factor that was not taken into account in the research project. The once familiar old wives' tale that storks deliver babies or that the presence of storks enhances human fertility provides the classic example of the fallacy of disregarding the effect of unmentioned factors. The tale relied on the observation of the positive correlation between the presence of storks' nests on the chimneys of a house and the number of children living in the house. However, assuming the correlation is correct, it does not prove that the tale is true—nor, incidentally, does it prove that babies deliver storks. It has been suggested that the correlation might well reflect the fact that larger families tend to live in larger houses and that larger houses (in the days of fireplaces and coal-stove heating) tended to have more chimneys, that is to say, more nesting places for storks.

[19]A politically correct copyeditor probably would strike out the word 'sex' here and replace it with 'gender'. However, while the sex of a person is an objectively verifiable condition, the word 'gender' is ambiguous: it may mean the sex of a person but it also may mean a pattern of behavioural, cultural, or psychological traits that is only statistically or historically related to persons of the one or the other sex. An effeminate man is of the male sex. It is anybody's guess what his gender is.

The correlation in our example obviously should not be taken as proof of the existence of a *causal* relation between the sex of a mayor and the frequency of serious road accidents in his or her town. However, sensationalism or overzealous feminism might lead some journalist or activist to pounce on our statistic to produce a newspaper report such as this:

Female mayor good for traffic safety

> Recent research at the university of X found that in our region serious traffic accidents, in which people get killed or seriously wounded, happen more frequently in towns where the mayor is male. According to the researchers, in these towns fast medical response units were called to the scene of a traffic accident on average 19.2 times during the first six months of last year. In towns where the mayor was a woman the average was only 4.5 times. The researchers emphasise that it is premature to draw any hard conclusions from their findings and that more research is required to explain the obvious discrepancy between the two averages. However, the discrepancy is large enough to warrant concern.

The headline is misleading. It suggests that there is a causal connection between a mayor's sex and the frequency of serious traffic accidents. However, that is not what the research had established. On the whole, the text of the article is an acceptable, if somewhat vague, account of the findings. For instance, it is not clear whether the last sentence is an editorial comment or a quote from one of the researchers. However, for many readers, the headline and the implicit suggestion that it reports a fact scientifically established by the research mentioned in the rest of the article will be all that they are going to remember.

Of course, people who look at our statistic from another angle might well assume that it shows that the voters who elect, or the authorities that appoint, a mayor are more likely to prefer a male to a female mayor in towns where road safety is a problem. The same caution applies: the assumption may be true but if it is then its truth is not established by our statistic.

It is hard to predict what will happen when statistical research is thrown before a public that is not alert to the difference between 'research finds a correlation between sets of data about A and B' and 'science discovers that A causes B'.

Relative risks

'N times more at risk of getting cancer', 'M percent more likely to go to prison / be unemployed / suffer from depression / getting overweight' and the like are attention grabbing headlines. The numbers N and M, whatever they may be, are likely to make an impact merely because they are featured in a headline. 'Five times as likely' sounds impressive but so does 'Fifty percent more likely'. Yet, the first corresponds with a so-called *relative risk* of 5 while for the second we have a relative risk of 1.5. 'Equally likely' translates to a relative risk of 1. We had better be clear about the meaning of statements of relative risk because they lend themselves easily to misunderstanding and abuse.

Like a correlation, a relative risk is a numerical value, which we can calculate whenever we have a sufficient number of numbers. The numbers may be meaningless or refer to nothing even remotely interesting. For example, in the previous two sentences there are altogether 6 verbs and 5 adjectives; in the verbs the letter 'a' occurs a total of 5 times, in the adjectives 3 times. So, relative to the adjectives, the risk that a verb has an 'a' in it is 1.39—which means that in those sentences the verbs are 1.39 times as (or 39% more) likely to have an 'a' than the adjectives. Alternatively, we can say that the relative risk of an adjective having an 'a' is only 0.72. Clearly it is not our ability to calculate relative risks that makes for interesting or relevant statistics; it is the availability of interesting and relevant data of good quality. A statement of relative risk merely is a way to summarise in one number a relationship among the data.

Let us return to our example and suppose that we also have statistics on the number of less serious accidents in our seven towns when the police was called to the scene of an accident but no medical response unit. Then we can derive statistical statements of the relative risk of serious accidents in the towns. Consider the following hypothetical examples:

First hypothesis:

	Accidents	
Mayor	**'Serious'**	**'Less serious'**
male	**96**	**853**
female	**9**	**341**

For serious accidents (those with the intervention of a medical response unit) the ratio between towns with a male respectively a female mayor is 96 / 9 = 10.67. For less serious accidents (police intervention only) the corresponding ratio is 853 / 341 = 2.50. The ratio of these two ratios is

$$10.67 / 2.50 = 4.268.$$

This is what statisticians call the relative risk (RR). In our case it is the relative risk of an accident being a serious accident in a town with a male mayor compared to the risk of an accident being a serious accident in a town with a female mayor. Specifically, the calculated relative risk tells us that an accident in towns of the first category is more than four times more likely to be a serious accident than an accident in towns of the second category.

The relative risk RR=4.268 is close to the ratio of the average number of *serious* accidents in the five 'male' towns to the average number of such accidents in the two 'female' towns:

$$(96/5) / (9/2) = 19.2 / 4.5 = 4.267.$$

This correspondence between the two numbers is a consequence of the fact that we have constructed our example in such a way that the average number of *less serious* accidents approximately is the same for 'male' towns and for 'female' towns:

$$853 / 5 = 170.6 \text{ and } 341 / 2 = 170.5$$

Statistically then, with respect to the *less serious* accidents there is almost no difference between the two categories of towns. Of course, there is no reason why this always should be the case. So let us play around a bit with the numbers for less serious accident and see what happens to the relative risk of an accident being serious.

Second hypothesis:

	Accidents	
Mayor	**'Serious'**	**'Less serious'**
Male	**96**	**1853**
Female	**9**	**341**

We have now a thousand *less serious accidents* more in 'male' towns than under the first hypothesis. Otherwise the data remain the same. Clearly those towns are less safe traffic-wise than in the first example. What happens to the calculated relative risk?

$$(96/9) / (1853/341) = 1.963$$

or nearly 2. Thus, despite the fact that the traffic situation in 'male' towns is worse than in the first example, the corresponding relative risk is less than half of what it was! The chance of an accident being serious in such a town is now only twice the chance in a 'female' town.

Third hypothesis:

	Accidents	
Mayor	**'Serious'**	**'Less serious'**
Male	**96**	**1853**
Female	**9**	**34**

Judging by the number of traffic accidents the situation in 'female' towns is much better than in the previous examples: the number of less serious accidents is only one tenth of what it was, otherwise the situation is exactly as under the second hypothesis. As for the RR ratio, it now is

$$(96/9) / (1853/34) = 0.196$$

or a little under 0.2. In other words, the chance of an accident in a 'male' town being a serious one is only one fifth of what it is in a 'female' town.

Note that we have made no changes to the numbers in the series X2 and Y6. It is still possible to present the traffic situation in the region as was done in the newspaper item cited earlier: "Female mayor good for traffic safety". However, as the third example in this section shows, one may feel equally justified to present it under the title "Male mayor good for traffic safety". In fact, the data do not justify either one of these headlines.

Note also that I have consistently used the expression 'the chance of an accident being a serious accident'. Popular media are not likely to use such an expression. With their commitment to 'simple and clear language', they may well prefer to use 'the chance of having a serious accident'. However, the two expressions obviously are not synonymous, as the one speaks of a relative and the other of an absolute risk. Such substitutions of one expression for another are especially likely in widely circulating statistics, which often are quoted from secondary or tertiary sources without consultation of the primary source.

Relative risk versus absolute risk

Consider the following statistic from the records of a particular airline company. It gives us, on the one hand, the numbers of employees (pilots and stewards) and passengers that died in accidents with the airline's planes and, on the other hand, the number of employees and passengers who were not involved in any accident or, if they were, survived.

	Died	Did not die
Employees	**14**	**520**
Passengers	**609**	**4123584**

For employees the risk of dying in a plane accident relative to the same risk for passengers is RR = 182.5. This is an enormous difference. Yet, it does not mean that only fools will want to become pilots or stewards. It does not mean that pilots and stewards are likely to die in a plane accident. From a statement of relative risk we cannot deduce anything concerning *absolute* risk.

Suppose we have the following statistic on the relative risk of smoking cigarettes with regard to coronary heart disease and lung cancer for a population of 100,000 smokers and an equal number of non-smokers. The (hypothetical) numbers refer to deaths in a particular year in a particular country[20]:

	Died of CHD	Died of lung cancer	Did not die
Smokers	**560**	**106**	**99000**
Non-smokers	**309**	**5**	**99500**

[20] Conversion from yearly risks to lifetime risks involves the use of special assumptions and is therefore controversial. For smokers the absolute risk of dying of lung cancer reportedly is 6% (although other numbers also circulate). However, many smokers live to a ripe old age (at or above mean life expectancy), which suggests that their cancer of the lungs is linked to their age rather their tobacco use.

Statistics may vary considerably from one country to another and from one social or ethnic group to another. See WHO's *Country Reports* (available via its web site). In Japan, with 59% smokers (15 years or older, 1994), there were 81.2 lung cancer deaths per 100.000 smokers. In the U.S.A., with 28.1% smokers (1991), the corresponding number was 305.7 per 100.000 smokers. There are many *possible* explanations for this wide discrepancy. Of course, national statistics often are not comparable. In 1976, Belgians and Mexicans smoked about 1500 cigarettes per year. In Belgium lung cancer deaths were 55 and in Mexico they were 5 per 100.000 inhabitants. Again, age seems to be the major factor: average age at death was considerably higher in Belgium than in Mexico. James Le Fanu, 'A Healthy Diet—Fact or Fiction?' in *Health, Lifestyle & Environment*, Social Affairs Unit / Manhattan Institute, London, 1991. Similar discrepancies are said to exist between lung cancer in northern and southern parts of Europe. One explanation that made headlines was Ancel Keyes' identification of the 'mediterrenean lifestyle'. Again, see Le Fanu.

The smokers' relative risk for CHD is RR=1.82. In other words, based on these data, for that particular year and country, smokers were nearly twice as likely to die of CHD as were non-smokers.[21] However, it also follows from the data that compared to non-smokers a smoker's statistical chance of *not* dying from CHD in that year was 99.5%. For lung cancer the numbers imply the relative risk of smokers dying of the disease is RR=21.3.[22] Yet the chance of a smoker *not* dying of lung cancer in that year was 99.9% of what it was for non-smokers. We see how easy it is to put a spin on the numbers: '2 or 20 times more likely' translates to 'virtually the same chance'! If all causes of death are taken into account, the data show that for a smoker the chance of surviving the year still was 99.5% of the chance a non-smoker had.

Problems of classification and confounding factors

Unlike the airline example, where there was an unambiguous criterion for classifying a person either as an employee of the airline or as a passenger on its planes, the smoking example raises an obvious problem concerning the classification of a person as a smoker or a non-smoker. Between the life-long non-stop chain smoker and the person who has never smoked anything there is a whole range of smoking habits. Where do we draw the line? Moreover, diagnosing death is one thing, diagnosing a particular disease as the cause of death is quite another thing. We must have some confidence in the ways the causes of death were determined, if we are going to use the available data for our statistical calculations. Which doctors or diagnostic methods are better at avoiding *false positives* (a patient is diagnosed as having a disease when in fact he does not have it), and which are better at avoiding *false negatives* (the disease is present but is not detected)? Which level of confidence in the data do we require? Do the available data satisfy our criteria? Different researches may have different views on these questions.

Assuming we have made the distinction between smokers and non-smokers in an appropriate way, it is clear that if a smoker dies of CHD one cannot say simply that his smoking habit caused his coronary heart disease (or caused it to be fatal). Non-smokers too die of the disease. In other words, a statistic of smokers versus non-smokers is not the same thing as a statistic of the effects of smoking versus the effects of not smoking. This is inevitable because, besides smoking, other factors (from genes and the natural environment to working conditions and lifestyles) are known to have an effect on a person's health or life expectancy. And then there are the factors that are not known! It is virtually impossible, of course, to have data on the extent to which each one, or any combination, of these confounding factors affects a particular individual person, let alone all the people in a given population. Thus, to move from 'a smoker dies' to 'his death is attributable to his smoking' one must *interpret* the data and *adjust* the statistics to avoid the conclusion that a smoker faces only one cause of death. However, to do so one must invoke assumptions and theories that may be more or less contestable.

Statistics may be based on data about lots of individual persons but before the data can be used in statistical calculations they must be abstracted from those persons. In statistics, the persons disappear from view and only the numbers remain. Consequently it is generally illegitimate to base an individual's chances on a statistic for a group or category to which some researchers have assigned him. A 33% divorce rate for the general population does not imply that the marriage of Mr and Mrs Smith has a one in three chance of breaking up.

In some cases people will not be moved by statistics that omit mention of confounding factors. A statistical finding that there is a strong positive correlation between a certain type of food and health or longevity makes no impression on people who are allergic to it. One man's poison may be another's medicine—and who is willing to forsake his medicine merely because it is not beneficial for others? Nevertheless, it may happen that overzealous

[21] The relative risks that are cited in the literature for smoking and CHD are in the range of 1 to 3.

[22] This is in line with published relative risk of smoking cigarettes for lung cancer.

regulators withdraw a drug from the market merely on statistical grounds, even if it is of proven value to those patients for whom it is intended. In some cases, however, the focus on a single factor (smoking, fast driving, CO_2 emissions) is so intense that it blocks serious consideration of other factors (relating to health, road safety, climate change) and consequently leads to a waste of resources that might have been employed to more effect for other purposes.

Epidemiological research

Epidemiology is the study of pathologies using statistical methods. Its reputation was made in the nineteenth century when statistics were first used successfully in the fight against the recurring epidemics of cholera, especially in big cities like London. By carefully registering data on the place of residence of Londoners that succumbed to the disease, William Farr (1807-1883) eventually led medical and scientific research to focus on the quality of drinking water and to identify the bacillus responsible for cholera. Note that Farr's correlations between incidences of cholera and certain water fountains did not identify the cause of the disease. In fact he believed that the cause was foul or polluted air (vapours[23]). However, his contemporary John Snow (1813-1858), a physician who used Farr's statistics, had become convinced that cholera was a water-born disease when he had noticed that the victims in one of the epidemics predominantly were users of water of one particular water company, whereas the clientele of a competing company that supplied the same neighbourhood largely remained untouched.[24] In the 1854 epidemic Snow acted on his conviction by insisting that the handle of the pump in Broad Street (in Soho, where the incidence of cholera was very high) be removed. This forced the locals to get their water elsewhere. The results were dramatic: the epidemic receded quickly. Snow's hypothesis, which became known as the germ theory, eventually gained the upper hand, when scientific methods for analysing and comparing water samples were perfected, controlled laboratory experiments could be set up, and first Filipo Pacini (in 1854, whose findings were ignored) and then Robert Koch (in 1884) identified the bacillus that caused cholera.

Statistical methods were successfully applied also to other epidemics. Epidemics are relatively rare but significant *specific* events with possibly devastating consequences. They happen in certain places at particular times. Hence, there is good reason to suspect that an epidemic has a specific cause and that there are specific ways in which it spreads from one person, animal or plant to another. This is not true for common illnesses, causes of death, or social and economic conditions—the sort of things to which epidemiological methods eventually were applied with increasing frequency. Indeed, for many of these diseases and conditions there is no reason whatsoever to suspect that they are related to a single cause or a well-defined set of circumstances. Consequently, the fact that one person's physical or social condition is of the same general sort as another's does not justify the conclusion that it has the same cause as the other's, or that one of them communicated it to the other. This is true if all the various possible causes are known; it is even truer in the far more usual case where we do not know all the possible causes.

As currently used, the word 'epidemiology' no longer denotes the contribution statistics can make to the study of epidemics but the use of statistical methods to analyse data regarding almost every aspect of human life (primarily with respect to health issues, but increasingly also social, economic and cultural matters[25]) as well as animal or plant life.

Instead of a clear focus on the causes of a particular event (an epidemic) epidemiology now resorts to 'imaginative observation'[26] to juxtapose one statistic to another in the hope of

---

[23] Many social reformers in nineteenth century England subscribed to the "miasmatic " theory, which in its extreme form held that everything that smells bad is [a cause of] disease.

[24] On Snow and Farrar, see http://www.ph.ucla.edu/epi/snow.html (consulted August 3, 2005).

[25] This extension of statistics to matters in which human action and decisions are of paramount importance is a consequence of using positivist methodology where it is not applicable. Positivism assumes that the progress of science consists in testing hypotheses by checking whether they are in agreement with empirical facts (usually in the form of statistics). Hence, positivist sociology, economics and political science are forever asking for statistics—and statisticians and epidemiologists are increasingly willing to meet this demand. For the reasons why this approach is misguided in the field of human action, see Ludwig von Mises, *Human Action* (The Scholar's Edition), Auburn AL, 1998, 105-118

26 R. Doll & R. Peto, *The causes of cancer*, Oxford University Press 1981. The authors had no qualms in advocating expensive research and intrusive government interventions on the basis of their 'imaginative

finding positive or negative correlations between them, and then presenting these as lists of 'risk factors'. In this manner several hundred risk factors for cardiovascular diseases have been named: alcoholism, abstinence, smoking, high level of consumption of milk, low level of consumption of milk, noise, snoring, age, obesity, extramarital sex, English as a first language, slow facial hair growth, taking the pill, being an only child, being the fifth or later child in a large family, coffee, urban residence, early menopause, low social standing, and what not (and who cares?). Similar lists exist for all sorts of cancers. Practically all of these correlations are extremely weak. More often than not, they cannot be replicated in other studies. For only a few of them, is there even a hint of a plausible causal mechanism that would explain the correlation. Not surprisingly, they virtually never lead to any scientific or medical breakthrough.

Although some of these risk factors (tobacco, alcohol, the pill, extramarital sex, low social standing) may have been investigated at the behest of activists who realised the propaganda potential of buttressing their cause with numbers, most of them clearly are the result of frivolously applying an existing method to any available set of data. However, once the numbers are available, anyone can use them to propagate one theory or another, or this or that product, treatment or policy. Anyone can make up a theory and get—or manage to get—some media attention for it. Of course, the fact that there is some published statistical correlation that would be relevant, if the theory were true, does not warrant the conclusion that the theory is true. Yet, precisely that conclusion is commonly accepted in media stories and commercial and political propaganda.

Publication bias

We should be on our guard with respect to the published results of epidemiological studies even when we have every reason to assume that the studies themselves are conducted carefully and competently. What is published hardly ever is the full story. With the following fictitious case we shall illustrate how the phenomenon known as publication bias may lead to a distorted presentation of otherwise impeccable research.

Let us assume that we have observations about the presence of a factor F (say, a chemical substance) and a disease V on an island. The substance has been introduced on the island some years previously and since then has found many applications in the life of the islanders. Because there are indications that the number of V-patients has increased in recent years, some people say that with increased exposure to the substance comes an increased risk of having the disease. However, they have no clinical evidence to back up their assertion. Medical science knows of no mechanism that directly or indirectly links F and V as cause and effect. Nevertheless, some people are concerned and there is some debate on the question whether F is indeed a significant factor in explaining the rise in the incidence of V.

Let us assume further that there are a million people living on the island, one hundred thousand in each of its ten provinces. Ten epidemiologists, one in each province, undertake to investigate the relation between F and V. We assume they are diligent and competent and work independently of one another.

Finally, we assume that we (you the reader and I) know everything there is to know about the factor F and disease V. Thus, we know that one in ten thousand people suffer from V (in

---

observation'. They also had no qualms in presenting their relative risks as *causes* (in direct violation of the epidemiological guidelines established by Richard Doll's mentor A. Bradford Hill, see below in the text), while deliberately ignoring age, by far the most important 'risk factor' of cancer. In his later years, Doll (who died in 2005) became a successful consultant and expert witness for governments and businesses. Peto is best known for his massive statistics and scare mongering on smoking ("1 billion people to die from tobacco-related diseases by 2100"). The book set the tone for what was to come by insisting that cancers almost certainly can be avoided: you'll die unless you eat, do and live as you are told—the message that makes modern epidemiology the darling of governments, commercial advertisers and activists.

the absence of F) and that the presence of F increases the risk of suffering from V with approximately 50%. *We* know this, but the people on the island do not!

First, each of our epidemiologists goes to the medical archives in his province to collect data on those who suffered from V in the last year before F was introduced. This tells them how many people had V in an environment that was free of F. The results are listed in the following table. On average each province had 10.2 patients diagnosed with V. As each province is home to 100.000 people, this result is nicely in line with the fact [known to us] that one in 10.000 people has the disease.

1	2	3	4	5	6	7	8	9	10	Aver.	F
8	10	6	10	11	10	16	11	7	13	[10.2]	No

Next, the epidemiologists review the situation today, some years after the introduction of F. The results are in the following table.

1	2	3	4	5	6	7	8	9	10	Aver.	F
18	17	20	6	15	12	23	14	21	14	[16]	Yes

A comparison of the two sets of data gives us the following table:

1	2	3	4	5	6	7	8	9	10	Aver.	F
8	10	6	10	11	10	16	11	7	13	[10.2]	No
18	17	20	6	15	12	23	14	21	14	[16]	Yes
Percentage change										Aver.	
+125	+70	+233	-40	+36	+20	+44	+27	+200	+8	[72.3]	
Relative risk of F for V										Aver.	True
2.25	1.7	3.33	0.6	1.36	1.2	1.44	1.27	3.1	1.08	[1.733]	[1.5]

The average relative risk (1.733) approaches the true relative risk (1.5), which is known to us but not to the people on the island. Note, however, that the calculated relative risks vary significantly, from 0.6 to 3.33, for the individual provinces. This was to be expected given the observed changes in the numbers of V-patients. These changes vary from a decrease of 40% to an increase of 233%.

Remember that the researches work independently. Accordingly each one of them intends to publish his results. However, it is unlikely that researcher 4 will find a medical journal that will publish his findings. After all, he found a negative result, a relative risk lower than 1. According to his findings, there is nothing wrong with F, as far as its effect on V is concerned. On the contrary, his research suggests that F diminishes the risk of V with more than a third. Of course, in itself this may be an interesting finding, but the editors or publishers may well think that it must be an anomaly: all the other findings submitted to them show a relative risk greater than one.

Moreover, given the limited resources at their disposal, they are more likely to publish papers that confirm the suspicions against F than papers that place F in the same category as the many thousands of other substances that do not noticeably increase the incidence of V. Thus, researchers 2, 5, 6 7, 8 en 10 also should not expect to find a ready outlet for their findings. They suggest relative risks in the range 1.08 to 1.7. These RR-values fall well below the threshold (RR>3) that serious epidemiology requires for studies such as those in our example (see below). Indeed, the epidemiological research gave no information about other factors that causally might affect the incidence of V, or about biases in the methods used to collect and interpret the data. For example, there was no information on the time it took for the use of F to spread to the different provinces after it was introduced on the island. There also was no information on possible changes in health insurance policies that might have led

doctors and hospitals to be more inclined to diagnose V in dubious cases than they had been before.

In contrast, the findings (RR=3.33 and RR=3.1) submitted by researchers 3 and 9 are above the RR>3 threshold, which suggests a lower probability that the results were produced by the coincidence of confounding factors and biased methods. Researcher 1, who found RR=2.25, is an intermediate case. The editors and publishers probably would have considered his result (RR>2) significant, if they had been reasonably sure that it was not vitiated by confounding factors or bias. However, in the absence of such assurance, they may well decide that it would be irresponsible to publish results that do not meet the criteria that generally are accepted as safe in studies with so many unknowns.

Hence, there is a good chance that only the findings of researchers 3 and 9 will be published, especially if we consider large-circulation publications (journals for medical professionals, and general science magazines). Such publications typically reach an audience beyond a small circle of people with enough expertise to evaluate research on their own. They also are the most likely sources of information for other media (newspapers, radio, television) that seek to inform the general public, including politicians and other policymakers.

Let us assume, then, that the editors and publishers reject the findings of researcher 4 as anomalous and impose the RR>3 norm for the others' submissions: only the research submitted by 3 and 9 gets published. To the outside world, then, the "science" on the relation between F and V looks like this:

3	9	Aver.	F
6	7	[6.5]	No
20	21	[20.5]	Yes
Change (in %)		Aver.	
+233	+200	[+216]	
RR		Aver.	True
3.33	3.1	[3.2]	[1.5]

Note what has happened to the average value of the relative risks. At 3.2 it no longer approximates the true value [1.5]. The wider media may believe they are justified in claiming that all 'serious scientists' (that is to say, those whose work was published in 'reputable publications') agree that F at least triples the risk of V. Of course, it is likely that one of the other epidemiologists writes a letter to the editor pointing to the different outcome of his research However, there is a good chance that he will be considered a 'dissident from the accepted view' or dismissed as not being an authority because his research was not published. Yet, we know (from the way we have constructed our example) that each of our ten epidemiologists was fully competent and diligent. From a purely scientific point of view, their research was equally impeccable.

Our example illustrates the pitfalls of assuming that publication is the distinguishing mark of sound science. We should *not* assume that the interests of publishers of research findings coincide with the interests of science itself, even when the publishers' decisions are taken in good faith and are quite reasonable and defensible in themselves. Publication bias does not mean that there is a willful attempt to bamboozle the public. However, it is clear that it may leave the public with an incomplete and misleading impression of what research has brought to light, and so spark exaggerated worries and baseless panics. We easily can imagine that on our island some politicians, bureaucrats, pressure groups and militants for one or another cause will not hesitate to use the published statistics in their never-ending quests for attention, votes, budget increases, subsidies and support.

In our example we have assumed that the substance F is an additional risk factor for disease V (true RR=1.5). Publication bias merely *exaggerated* the perceived relative risk. However, it would be easy to construct an example showing how publication bias can create

the impression that F creates a serious risk, even when in reality it is not a risk factor at all (with true RR=1).

Note also that in our example publication bias resulted from the rejection of research that was considered an anomaly or failed to establish a relative risk above some threshold (RR>3). The use of this relatively high standard was justified (at least in the publishers' mind) by the paucity of data on confounding factors and methodical biases. However, publication bias can also be the result of using RR-thresholds that are far too lax, given the data and methods used in calculating risk ratios. If that is the case, much, possibly all, of the published research may be unreliable. Yet, the sheer number of publications may convince people that there would not be all this smoke if there were not a fire somewhere. This is the more likely if the studies purport to identify risk factors for serious diseases or conditions. Then it is a safe bet that someone will invoke the so-called precautionary principle to take preventive measures even if there is no indication to believe that the published studies actually found evidence of a looming danger. (We shall consider the precautionary principle in the last section.)

To sum up

- Even impeccable research can create a false impression merely because the published material may not cover all the research that is relevant to a particular question. It follows that conclusions, even those that are based on an exhaustive study of *published* material, may be widely erroneous if that material reflects publication bias.
- Publication bias obviously will be especially significant in publications intended for a public of non-specialists (a fortiori the general public). Where criteria such as 'social relevance', 'political correctness', and 'respect of public opinion' come into play publication bias has the potential of creating the sort of context in which research and researchers can derive significant rewards from their ability to conform to, or at least not to offend, prevailing prejudices. As noted before, this is the more likely the more the funding of research becomes politicised or dependent on the agendas of large organisations.
- However, publication bias also affects specialised publications. No journal or magazine can do without a publication policy to determine which submissions it will accept and which it will reject. Especially in disciplines that are dominated by one or a few 'leading journals', editorial policies may introduce bias in favour of, or against, certain topics, hypotheses, methods of research, and even particular institutions and individual researchers, that can have a long-lasting effect. Its effects should be gauged in the context of the publish-or-perish culture that prevails at many academic and research institutions, where researchers (especially the younger ones) are treated as employees and judged by their ability to produce lots of publications in established journals.[27] This arguably creates a bias in favour of 'swimming with the flow' of the current establishment, while contrarian or unorthodox thinking is relegated to 'obscure journals', small informal groups, or the Internet.

[27] For a general discussion focussed on the U.S.A, see John W. Sommers (ed.), *The Academy in Crisis: The Political Economy of Higher Education*, The Independent Institute, San Francisco 1995.

Statistical significance *versus* scientific relevance

When a statistical finding is published, it usually comes with an announcement that it is statistically significant at some level or other. By the sound of it, such an announcement is reassuring: it suggests that the statistical correlation to which it is attached is not a fluke but a revelation of something real and possibly important. However, what does it mean?

Relative risk again

One aspect of statistical significance we have mentioned already. In the example of the epidemiological research into the relation between substance F and disease V, we noted that people (in our case editors and publishers of scientific or professional journals) would consider the degree to which the research had been controlled for confounding factors and possible sources of bias. The better the evidence that this control had happened, the more willing they should be to accept lower relative risks. The more doubts they have on this score, the more they should insist on high relative risks.

As a rule—but it is no more than a rule of thumb—relative risks RR>3 are considered significant for positive risks. They indicate a 200% increase of the risks associated with the factor under investigation. If there is a negative risk (RR<1) then exposure to the factor appears to diminish the risk that the disease or condition will occur. To indicate at least a 200% decrease, or a threefold reduction, of the risk a relative risk would have to be RR<0.33. In other words, relative risks between 0.33 and 3.0 will *not* be considered significant. For research that takes into account and controls for confounding factors and bias, the thresholds of significance will sometimes be relaxed to RR>2 for positive and RR<0.5 for negative risks.

However, in epidemiological studies it often is impossible to control for even one, let alone more than a few confounding factors (an unknown number of which are unknown factors). Such studies moreover rely heavily on methods with inbuilt but hard to detect bias: questionnaires, interviews and voluntary reports loaded with words that may mean one thing to one person and another thing to another person, diagnostic records and certificates issued by doctors and official bodies for purposes unrelated to the question that is being investigated, and so on and so forth.

One should not be misled by the fact that epidemiologists often claim to have controlled their data for ten, twenty or more other factors. This almost always means only that they have adjusted their numbers by somehow taking into account the results of already published epidemiological studies. However, those other studies were up against the same sort of problems and are therefore not to be taken at face value.

While mathematical juggling of numbers from different sources of doubtful reliability is a cheap way to produce a 'new study', there is no reason to expect that it will produce a reliable statistic. In any case, it is not the same as controlling for other factors, as this is understood in laboratory settings and controlled clinical experiments. By the appearance of their published results (tables of numbers, percentage signs, equations, symbols and abbreviations), epidemiological studies may look like scientific experiments but they hardly ever are even remotely similar in design or execution.

In carefully designed and executed controlled experiments, relative positive risks between 1 and 2.0 (or negative risks between 0.5 and 1) may be significant, because good design reduces the risk of confounding factors and good execution includes the use of state-of-the-art techniques of measurement. The scientific value of such experiments moreover rest on the fact that they can be repeated easily by other researchers, using other sets of data, and systematically modified to control for hidden biases in the design or execution of the original experiments. In contrast, epidemiological studies rarely are repeatable—and not only occasionally because of ethical objections, say, against manipulating people. More often the reason is inherent in their data and methods. Hence, they often contradict one another, as

when one study finds that drinking coffee increases the risk of getting cancer (or ending up with a low paying job) and another concludes that it decreases the risk.

A word of caution about the rising fashion of using computer simulations is in order here. Such simulations easily can be repeated and modified systematically (if the source code is made public and sufficiently documented). They can be immensely useful. It is legitimate to combine scientifically established relationships and the results of well-controlled experiments into computer programs and then to use these to simulate the behaviour of more or less complex systems under different circumstances. However, it is not legitimate to add to, let alone substitute for, these elements putative relationships and hard-to-evaluate data and still pretend that one is simulating the behaviour of a really existing system. One cannot presume that a computer program is a model of something merely because it produces output that seems to vary when the input varies—if it were, every computer game would be a model of something. Moreover, even a good model will produce rubbish when it is fed garbage. To the extent that the model itself is garbage, it will turn even solid input into rubbish. In research as elsewhere, high-tech is no guarantee of high quality.

P-values

Another aspect of statistical significance concerns the probability that the results were not due to mere chance. Thus we are not looking at the quality of the data and the methods for collecting and interpreting them but only at a mathematical property of the data.

Statisticians will refer to the so-called p-value of a result to express this aspect of statistical significance. As a general rule, that value should not be larger than 0.05. If it is larger then the probability that the result is due to mere chance is larger than 5%, and this is considered unacceptable. The announcement that a statistical finding is significant at the $P<0.05$ level therefore means that the chance that it is due to mere chance is less than one in 20. Obviously $P<0.05$ is not holy writ. For many purposes it may be much too lax.

The idea behind the p-value criterion is simple. We know that a throw of a (fair) die has a one in six chance of producing three eyes.[28] Now, someone presents us the result of an experiment: he found that a *red* die has a one in six chance of showing three eyes. Obviously, the chance that his outcome is due to mere chance is 1, way above the 0.05 threshold—in fact it is a certainty. The next day he returns with the announcement that he found that a *yellow* die has an 85% chance of showing three eyes. That outcome certainly is statistically significant. However, it does not prove any correlation between the colour of a die and the result of rolling it. That is because we have not controlled for other factors (weight, material, sharp versus rounded edges, the presence of magnets near the place where the experiment took place, and so on). To find correlation between colour and outcome we need to *randomise* as many of these factors as we can. Only if the statistically significant result holds up in a sufficiently randomised trial should we conclude that correlation exists.

There is an obvious problem here: it may not be possible—physically, economically, or ethically—to experimentally randomise other factors. In epidemiological studies it is possible, if at all, only on a modest scale. That is why epidemiological studies require data about large samples, if they are to inspire some confidence. It is reasonable to expect that large samples (selected for a particular attribute) *naturally* will be more randomised with respect to other attributes than small samples. However, it is no more than a hope that it will be *sufficiently* randomised, if the size of the sample is small relative to the relevant total population of things.

In short, statistical relevance, as measured by a low p-value, in itself expresses only a mathematical relation between numbers. It does not prove that there is a high probability that a correlation between real factors exists, unless we can be sure that the numerical

[28] In most real life situations we have no idea of what 'mere chance' means. Then we have to guess or have recourse to assumptions—which introduces elements of subjective belief or theoretical bias.

correlation was obtained from the study of a sufficiently randomised sample. Thus, a numerical correlation may be statistically significant without being of any scientific value.

Confidence intervals

Another measure of statistical significance is the so-called confidence interval, usually indicated by an expression such as 95%CI 2.5 – 3.5. This means that the true value of a relative risk has 95% chance of being in the interval between RR=2.5 and RR=3.5. Obviously, 95%CI 2.5 – 3.5 indicates more confidence than either 90%CI 2.5 – 3.5 or 95%CI 1.8 – 4.6: its CI-index is higher than the second measure's index, and its interval is narrower than the third measure's interval.

Moreover, we cannot have any confidence in a statistic that finds a positive relative risk (say, RR=1.21) if its 95% confidence interval has a *lower limit below or equal to 1*. For example, with 95%CI 0.9 – 1.9 there is a good chance that there is no risk at all or even a negative risk.

Similarly, we cannot have any confidence in a statistic that finds a negative relative risk (say, RR=0.7) if its 95% confidence interval has an *upper limit equal to or above 1*. For example, with 95%CI 0.4 – 1.09 there is a good chance that there is no risk at all or even a positive risk.

A wide interval does not inspire confidence: RR=30 (95%CI 5-180) means that the true relative risk may be up to 6 times lower or higher than the calculated relative risk.

To sum up:

- Whether or not one should consider a relative risk, p-value or confidence interval significant depends on one's evaluation of the quality of the design and the execution of the study that produced it, and in particular of the data and methods used in collecting and interpreting them.
- For truly scientific studies, a layman should not hazard such an evaluation because he lacks the knowledge and the experience to judge what is possible and appropriate in one research context as against another.
- However, much of the epidemiological research that ends up in the public sphere is so bad that it only takes a modicum of common sense to see through its scientific pretence. (This probably explains the sustained efforts to devalue common sense, which derives from an accumulation of lived experiences, and replace it with that overblown, fabricated and easily manipulable surrogate called 'public opinion'.)

Correlation is not causation

If we are given long lists of data covering many heterogeneous things then it is quite likely that we shall find some statistically relevant correlations. Much of the new epidemiology rests on computer-assisted analyses of large databases, with the computer signalling any correlation that satisfies the $P<0.05$ criterion. This procedure, known as a data dredge or mining the data, automatically will yield statistically relevant results.[29] No thought, let alone science, is required. Unfortunately, many of these findings get published (if only in the popular press) if someone takes the trouble of inventing a theory that *might* explain the correlation. Such ad hoc theorizing obviously has no scientific merit. However, it easily can make the correlation seem plausible enough to suggest that more research and more funds are needed—or attract the attention of some sales manager or interest group.

[29] Some databases seem to be maintained for no other reason than to make data mining possible. The most egregious example is Harvard University's Nurses' Health Study, which has been collecting data on the diet and lifestyle of more than a hundred thousand nurses in the U.S.A. See http://www.channing.harvard.edu/nhs.

Austin Bradford Hill, the founder of modern medical statistics, formulated nine criteria[30] that must be considered before moving from the statistical significance of a correlation to its relevance in scientifically establishing a causal nexus:

- *Strength*: The association should be strong enough so that we can rule out other factors.
- *Consistency*: Different researchers should be able to replicate the results under different conditions.
- *Specificity*: The exposure should be associated with a very specific disease as opposed to a wide range of diseases.
- *Temporality*: The exposure should precede the disease.
- *Biological gradient:* Increasing exposures should be associated with increasing risk of disease.
- *Plausibility*: There should be a credible scientific mechanism that can explain the association.
- *Coherence*: The association should be consistent with the natural history of the disease.
- *Experimental evidence:* Physical intervention should show results consistent with the association.
- *Analogy*: It should be possible to find similar results in other areas or using other approaches.

One should always keep these criteria in mind in evaluating the relevance of statistical research, especially in fields where for some reason or other rigorously controlled experiments are out of the question. However, in the decades following Hill's admonition, epidemiologists increasingly ignored his criteria, preferring to suggest causal relations where none could be inferred.[31] Even with sufficient randomisation, a statistical correlation can only reveal that there is a statistical correlation between two real things. It does not prove that the one thing *causes* the other. In science, to speak of cause-and-effect one must have a good understanding of *how* the cause produces the effect—in other words, one must have identified at least the basics of the mechanism that brings about the effect when it is activated by the cause. To say that A causes B but nobody knows how, is an oxymoron.

It should not come as a surprise that a recent review[32] found that one in three of 49 highly quoted studies published between 1990 and 2003 in three leading medical journals were contradicted by later research. As the reviewer concludes, "Controversies are most common with highly cited non-randomised studies, but even the most highly cited randomised trials may be challenged and refuted over time, especially small ones." It is an open question whether the results would be better for less prominent studies or studies in less prominent journals—but who is willing to bet on that possibility? Moreover, this was a review of medical science. What would we find if we were to turn to social science, let alone its penumbra of cultural, critical, and gender studies, and assorted ideological interpretations?

An infamous episode

We should be wary when research findings are touted as being statistically significant. Nevertheless, we should not ignore measures of statistical significance. There are research findings that do not meet even the formal requirements of statistical significance, yet claim to be scientifically relevant. That is the mark of junk science.

---

[30] A. Bradford Hill, "The Environment and Disease: Association or Causation?" Proc. Royal Soc. Med. 58:295 (1966)

[31] James LeFanu, *The Rise and Fall of Modern Medicine*, Little Brown 1999, p59.

[32] John P. A. Ioannidis, *Contradicted and Initially Stronger Effects in Highly Cited Clinical Research JAMA.* 2005;294:218-228.

Another reason to look carefully at such measures is to see if they have been tampered with. By lowering the norms recommended by scientific practice, it is easy to suggest statistical significance where no honest researcher would find it. The most flagrant case of this type of fraud was the research presented by EPA on so-called passive smoking. When this research was cited in a court case[33], the judge William Osteen had this to say.

> The record and EPA's explanations to the court make it clear that using standard methodology, EPA could not produce statistically significant results with its selected studies. Analysis conducted with a .05 significance level and 95% confidence level included relative risks of 1. Accordingly, these results did not confirm EPA's controversial a priori hypothesis. In order to confirm its hypothesis, EPA maintained its standard significant level but lowered the confidence level to 90%. This allowed EPA to confirm its hypothesis by finding a relative risk of 1.19, albeit a very weak association.

EPA lost the case. However, governments and government agencies are jealous of their powers. Given their influence in the legislative process, they easily can erect hurdles to discourage or avert legal challenges of the scientific expertise to which hey lay claim.[34] Judge Osteen's denunciation of EPA's egregious attempt to manipulate public opinion did not stop the flood of regulations that the passive smoking scare spawned. Once bad science, junk science, or fraudulent abuse of science has been written into the laws, citizens are left virtually defenceless. 'Numbers put into an Act of Parliament become enshrined as icons.'[35]

---

33 Flue-Cured Tobacco e.a. *v.* U.S. Environmental Protection Agency, in The U.S. District Court for the Middle District of North Carolina (Winston-Salem Division), July 17, 1998, W. Osteen J. (p.77)

[34] The US Court of Appeal, 4th circuit, finding that his court had no jurisdiction to review EPA's actions under the Administrative Procedure Act, vacated Judge Osteen's decision on December 11, 2002. This reversal may set a precedent for denying citizens the possibility of seeking judicial review also under the Information Quality Act. See http://www.thecre.com/pdf/20050124_insideepa.pdf.

35 John Brignell, *Sorry, Wrong Number. The Abuse of Measurement* (Brignell Associates / European Science and Environment Forum, s.l., 2000), p.140.

Terror, utopianism and power

The abilities to instil fear (terrorism) and to exploit fear (mobilizing for wars on terror) have always been recognised as marks of political power. People submit out of fear for their rulers or because they are led to believe that their rulers will be able to eliminate the causes of their fears. In modern Western societies the culture of fear is the background for the rise and expansion of the military, police and information gathering powers of the regulatory state, also known as the warfare-welfare state, the Nanny State, the Therapeutic State, the New Paternalism or simply Big Government. Arguably, the culture of fear is not a relatively inconsequential intellectual fashion. It may be an inevitable consequence of the unravelling of the alliance between knowledge (science) and faith that is the historical foundation of Western civilisation. What this implies for the future of science itself is anybody's guess. However, it is clear already that the esteem in which faithful science once was held does not extend to the fearful science that feeds one scare-mongering campaign after another.

Faith and science

Nearly one hundred years ago, C.K. Chesterton[36] wrote that the problem with people who have lost faith is not that they believe nothing but that they believe anything. Paraphrasing him, we might say as well that the problem with people who have lost faith is not that they fear nothing but that they fear everything. It makes sense. Faith comprises three basic attitudes:

- We cannot have and therefore should not aspire to superhuman knowledge and power.
- There are certain things—reality, truth, freedom, justice and other intellectual and moral values—which we have to accept even if their respectability cannot be proven scientifically.
- Despite our obvious limitations we are capable of dealing with the problems of existence in this world.

Arguably, science was born of that faith. However, to appreciate the argument it is important to distinguish between faith and belief. 'Belief', in this context, stands for the stories or myths by means of which people learn (or learn not) to have faith. There are, of course, many systems of belief that inspire faith—different mythological packages may contain the same basic truths—although some undoubtedly are more effective than others in this respect.

Unfortunately, some people are unable or unwilling to deal with the mythologies that constitute either their own or others' beliefs. They cannot handle myths. They will not be bothered with the effort to separate the truths the myths contain from the events they relate. Hence, their only option is, either to believe the stories as literal reports, or to reject their truths on no other ground than that all or some of the facts in the stories never happened.

Whichever side they choose, they are bound to disparage the notion that myths may have objective truth value even if they are wholly or in part works of imagination rather than historical records. The time-honored way of educating children by means of stories and mythologies that provide an intergenerational frame of meaning and reference becomes overburdened with efforts to instill belief or disbelief. This is unfortunate because building faith has to start in childhood, long before an introduction to science begins to make sense. Without an already established moral framework, could an education in science produce anything but monsters?

The unraveling of the alliance between science and faith stems from this inability to tell faith from belief and science from belief in science. Believers on both sides came to accept the notion that faith and science were deadly enemies. Rather than a vital support in a world full of uncertainties, science worshippers imagined that faith was the cause of our

[36] In his witty classic essay *Orthodoxy*, London 1908.

uncertainties. By promoting that view as the essence of enlightenment they insinuated a far more radical message: We shall not be able to cope with the problems of life unless we dare to raise ourselves to the position formerly occupied by no man (or, as their opponents prefer to say, by God). Then and only then shall we be masters of the universe, possessing the knowledge we need to control everything and to achieve full emancipation from every source of frustration and fear. The zealots could not contribute much to the advance of science, but they certainly felt up to jeering what they saw as its enemy.

There was a tiny little problem with their position: the control over everything includes the control over people, and that makes it rather interesting to know whether we shall be among the controllers or among the controlled. However, they told us that Science would solve that problem as well: through social engineering and eugenics it should be able to create a new breed of men that would accept unquestioningly the prescriptions of Science. (A strange thought: science without questions, blind devotion to science! But then it was an expression of scientism—that is, science worship—not of science itself.)

In retrospect it turned out to be easy to do away with faith, especially when our first attempts to govern ourselves by the light of our own genius led us into a period of devastating world wars and the most brutal dictatorships in history.

Could science fill the void, now that it was supposed to have made faith redundant? It soon became clear that it could not offer anything beyond the evolutionary truism that the judgement of history is written by the victors, in science no less than in politics. It is therefore more 'rational' to side with the strong than with the weak. Just make sure you guess correctly which side will win—that is to say, make sure to guess correctly the side that most of the others guess will win. Run with the pack! What else could a scientistic surrogate for morality advise?

For the first time in history, science found itself in a world without faith, and scientists were supposed to sail without a moral compass. Respect for reality? There is no reality except what we say it is. Truth? There is no truth, only the remorseless competition of partisan interests. Consequently, science is militant or it is irrelevant—and to be militant it has to strike terror in the hearts of men to arouse them from complacency and mobilise them for the cause it serves.

For the rest of us, what did the destruction of faith mean? We were supposed to forsake the superstitions of the past: the idea that we could cope with our own lives and the idea that others would share with us at least those universal notions of moral conduct and right thinking that faith-based civilisations had tried to instil in the young. If faith has no scientific basis then it has no basis at all.

Thus, we found ourselves in a world in which it was deemed irrational not to live in fear of everything and everybody, including ourselves, and irrational not to seek therapy, expert guidance and the protection of the mighty—whom we fear most of all and try to placate with daily demonstrations of submission. The signs are everywhere: There are more dangers between heaven and earth than your common sense can imagine; therefore surrender your common sense even if it is the only sense you have. You are what you eat, so eat what we tell you to eat lest you die with no one else to blame but yourself. Left to yourself you are a menace to yourself, society, and the planet: your only salvation lies in doing what you are told. Submit to our terrorism lest you fall victim to theirs. After all, *we* have the expertise and the science to deliver on our promises; who are *you* to argue with that?

Fear and power: the precautionary principle

Of course, the expertise and the science are not always what they are cracked up to be. However, it seems we are not permitted to see that as an opening for taking our chances on faith. On the contrary, the culture of fear admonishes us that inadequate expertise and insufficient science make it all the more imperative that we submit. This is called the precautionary principle:

> Where there is uncertainty as to the existence or extent of risks of serious or irreversible damage to the environment, or injury to human health, adequate protective measures must be taken without having to wait until the reality and seriousness of those risks become fully apparent.

In short, it is better to prevent some hypothetical danger than to be sorry if someday it should occur. A problem is what would be a problem if it were a problem—regardless of the lack of evidence that it is a problem. *Thou shalt live in fear!* …at least, vote and spend as if you do.

We all know from experience that the costs of prevention often are not worth the trouble. Rather than stay at home as a precaution against all that may happen on the road or in the workplace, we consider buying accident and liability insurance—or just take our chances. However, the precautionary principle has become a political shibboleth of the religion of fear. It is not really about taking precautions. Rather it is about setting enforceable priorities. Spraying DDT may seem like an efficient precaution against malaria to those who are most at risk, but it should not be tolerated if *our* concern is preventing pollution of the environment.[37] The 1991 cholera epidemic in Peru was a clear example of misplaced priorities in taking precautions. [38]

The precautionary principle now is included in laws, policies, and treaties, and in the constitutional documents of international bureaucracies and supranational authorities. It has been adopted in the process of empowering the European Union.[39] However, as it is not a genuine principle that applies in every case, there always has to be an authority that will decide where it applies and where it does not apply. In the context of the EU, the European Commission has announced that it will use the principle with the widest possible discretion:

> The precautionary principle is not defined in the Treaty, which prescribes it only once - to protect the environment. But *in practice*, its scope is much wider, and specifically where preliminary objective scientific evaluation indicates that there are reasonable grounds for concern that the potentially dangerous effects on the *environment*, *human, animal or plant health* may be inconsistent with the high level of protection chosen for the Community.[40]

What lies outside the range of this *concern* (whose concern?) that is based on *preliminary evaluations* about *potentially* dangerous effects that *may* be inconsistent with a *chosen* (by whom?) *high* level of protection (against what?) of the environment, humans, animals and plants? And those are only the criteria that will be applied *specifically*!

As the Commission interprets the precautionary principle, it is a free pass for the arbitrary political and administrative use of statistical data about everything and everybody—a pretext for laying claim to a totalitarian authority (as if that were not the first danger against which precaution is in order).

---

37 L. Mooney, & R. Bate (eds), Environmental Health: Third World Problems—First World Preoccupations (Butterworth, London, 1999); Washtimes.com/world/20020616-11558965 .htm (16 June, 2002). However, see http://timlambert.org/category/science/ddt/ for another point of view on DDT.

38 See note 11 above. Smoking (a positive risk factor for lung diseases) reportedly is a negative risk factor for Alzheimer's Disease and other malfunctions of the nervous system. Graves AB, van Duijn CM et al., 'Alcohol and tobacco consumption as risk factors for Alzheimer's disease' Int. Journal Epidemiology, 20, 1991; (2 Suppl 2) p.S48-S57. Apparently, individuals should not decide for themselves against which things they'll take precautions.

39 Article 130r(2) of 1997 Amsterdam Treaty (97/C 340/01) OJ C 340 of 1997-11-10:

40 'Communication from the Commission on the Precautionary Principle' van 2 februari 2000.

Utopian salvation

Appeal to the precautionary principle is not the only technique for staking such totalitarian claims. Perhaps the most scandalous argument for totalitarian power is the 'definition' of health that was written into the constitution of the World Health Organisation:

> Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition.[41]

WHO's definition of health includes not only absence of disease and infirmity but also happiness, wealth, social esteem, intelligence, absence of frustration and fear from any cause whatsoever, and who knows what else. That is not a definition of 'health' but of 'whatever you like'. The major function of the constitution of an organisation is to limit its powers in view of the purpose it is supposed to serve. WHO's constitution does the opposite: it stipulates a goal for the attainment of which no amount of funds and no range of powers will ever be enough. It is the constitution of an organisation aspiring to totalitarian powers.

The rest of the text merely serves to remove any remaining doubts about that aspiration. What on earth is *the highest attainable standard* of complete well-being? Would the lowest attainable standard of *complete* well-being not be more than enough? Who but WHO would proclaim the enjoyment of complete well-being a fundamental human right? Such a right could be guaranteed only in Utopia. To invoke it in a constitutional text of an organisation is to suggest that the organisation is capable of lifting mankind out of the real world and into the realm of utopian fantasy. Obviously, WHO is not capable of doing that. It is not even capable of eliminating disease and infirmity, that is to say, securing health in the common down-to-earth sense of the word.

The mere fact that such shameless nonsense as the EC's interpretation of the precautionary principle or WHO's definition of health can survive in the public sphere is an indication of the low standards of moral and intellectual integrity that prevail in it. It is folly to make such nonsense the arbiter of correct science.

Those who set out to make legislation and politics scientific will only succeed in politicising science. There is a pattern here. When the modern state made its appearance on the world stage some five hundred years ago, many thought it would be a means for enforcing justice; instead it ended up justifying force and in the process destroyed the common understanding of justice. According to present prevalent views, justice is no more than a personal opinion. To the extent that it has meaning in public life it is the ruling opinion, propagated by its control of education and enforced by its control of the state's police powers. Is that the fate that awaits science?

41 Constitution of the World Health Organisation (Introduction). [Emphasis added] See http://www.yale.edu/lawweb/avalon/decade/decad051.htm