Risks, Causality, and the Precautionary Principle

Paolo Vineis
Micaela Ghisleni

1. From knowledge to prevention

In a letter to The Veterinarian magazine of 1874, Mr. Priestman reported that during a week characterized by extremely dense fog he had been looking after a herd of cows at the yearly London cattle fair. ''On Tuesday, the first foggy day, many animals were having difficulty breathing as early as seven o'clock in the morning; in the evening most of the animals were showing signs of distress. Sheep and pigs were not affected…''. The animals which were worst off were slaughtered and showed signs of acute bronchitis, ''with lobular lung congestion and emphysema'' (Bates, 1994, 1). An accompanying editorial suggested there could be reason to worry about human health as well, but nothing was done until the famous events of December 1952 (that is, 70 years later) when more than three thousand deaths occurred within the London population because of the exceptional pollution. The question arising from the cattle episode is: when is evidence (even if indirect, for instance in animals) sufficient to justify preventive intervention?

To tackle the issue we will consider an exemplary story, the history of research on pellagra. The pellagra case is exemplary for various reasons. Firstly it constitutes a virtually unique case in the history of medicine, in that a single individual, Joseph Goldberger, managed, in the course of his lifetime, to formulate the correct hypothesis, to verify it empirically, both through descriptive data and through controlled trials, and finally to implement effective preventive solutions. Another reason why the story is edifying is that, in this case, not only did Goldberger manage to build a convincing ''causal web'' which pointed at diet as being the central element for inducing pellagra, but he also succeeded in demonstrating almost conclusively the mechanism through which it happened. From a scientific point of view the demand for evidence could never end. Even having identified the cause of pellagra in niacin (nicotinic acid) deficiency, we can further question the biochemical mechanisms through which the deficiency causes those specific symptoms, in a never ending argumentative chain. Clearly such an approach of systematic doubt within public health can bring to a complete standstill. The transposing of scientific evidence to guidelines for prevention is a very delicate issue, beginning with the nature and quantity of evidence needed to start preventive action. At times measures can be effective even in the absence of any clue as to the biological causes or the mechanisms involved. Preventive measures were found for many diseases before or even a long time before the true cause or mechanism of action were discovered, as showed in the following table.

The message put forward by the table is clear: in many cases waiting for definitive scientific evidence would have entailed a serious delay in implementing those preventive measures which, in retrospect, have been shown to be essential to reducing the rate of the diseases considered.

1.1. *How are risks identified: Goldberger and pellagra*

Pellagra is a disease which manifests itself in skin lesions, digestive problems, which can also be very serious, and, in very advanced stages, neurological and psychiatric problems. In 1907, when Goldberger was given the task of investigating it, there were at least 100,000 cases in the south of the USA, and many thousands in Italy and Spain. Most researchers

Topoi **23**: 203–210, 2004.
© 2004 *Kluwer Academic Publishers. Printed in the Netherlands.*

Table comparing the date of discovery of a preventive measure and the date of discovery of the causal agent (from Wynder, 1994, 2, modified).

Disease	Discoverer and date of discovery of the preventive measure	Date of discovery of the cause	Causal agent
Scurvy	Lind, 1753	1928	Ascorbic acid deficiency
Pellagra	Casal 1755	1924	Niacin deficiency
Cancer of the scrotum	Pott, 1775	1933	*Benzopyrene*
Smallpox	Jenner, 1798	1958	Orthopoxvirus
Puerperal fever	Semmelweiss, 1847	1879	Streptococcus
Cholera	Snow, 1849	1893	Vibrio Cholerae
Professional cancer of the bladder	Rehn, 1895	1938	2-naphthylamine
Yellow fever	Finlay, 1881	1928	*Flavivirus*
Cancer of the mouth (in tobacco chewers)	Abbe, 1915	1974	*N*-nitrosonornicotine

thought the disease had infective origins, following a vogue which was widespread at the beginning of that century. For instance, although Eijkman's research in Java clearly showed that beriberi was caused by diet, the search for a microorganism as its cause continued for many years. Eijkman showed that in 27 prisons where the convicts were fed polished rice, devoid of cariosside rich in vitamin B1, the rate of beriberi was of 1 every 39 prisoners, while in 74 other prisons where brown rice was eaten, the incidence was of 1 every 10,000. (The final isolation of vitamin B1, or thiamin, occurred in 1936 thanks to Williams, a chemist who worked for the telephone industry: the greatest boost to his research was a consequence of the 1929 crisis, when his working week was reduced to three days, allowing him to concentrate on his research on beriberi in his home garage.)

The insistence on the infective origin of diseases, up to the first decades of the XX century and further, reflected a bias of the scientific community (a systematic distortion) which had consequences on research and obviously derived from the huge successes achieved by microbiology during the second half of the previous century. On the other hand, the role of *a priori* expectations in the whole of scientific research and in the interpretation of medical practice is essential: it is well known that in the XVII century vomit and diarrhea were considered signs of the effectiveness of the treatment of heart disease with foxglove (digitalis), while now we know they are symptoms of overdosing. The reason is obvious: in the XVII century the theoretical model at the heart of medical practice – especially in dropsy or cardiac insufficiency – was based on the idea of eliminating excess liquids.

Returning to Goldberger, he suspected that the disease did not have infectious origins on the basis of reports coming from Italy from which it turned out that although no precautions were taken to "avoid its propagation" there was no risk of infection from the sick to the healthy. In a vast "pellagra colony" near Venice, where between 300 and 500 patients were housed at any time, none of the staff, mostly nuns, had ever contracted the disease. With an analogous investigation on the treatment centers in the USA, Goldberger confirmed that the staff did not fall ill, and formulated a diet-related hypothesis for the first time. But there was a problem: although in some institutions patients and personnel were fed different food, in others this did not happen. The explanation put forward by Goldberger was that the nuns kept for themselves the best foods and served themselves bigger portions, but it was a weak explanation. He then started a series of experiments on monkeys, inoculating them with patients' tissues and secretions, but always obtained negative results.

During an epidemiological research in a hospital for the chronically ill in Georgia, Goldberger studied 996 patients admitted during 1910. After a year 418 patients were still there (i.e. those who had not died or been discharged); 32 of these (7.7%) had developed pellagra during the year of observation. If the 293 staff members who had been in close contact with the patients had been exposed to the same risk of con-

tracting the disease, 22 cases of the disease (7.7% of 293) could be expected, while none occurred. On the basis of this epidemiological evidence, Goldberger decided to start a trial in two institutions in Jackson, Missouri. The trial consisted simply in supplying the patients with food supplements (especially pulses and milk): in the first institution none of the 99 guests kept under observation during the following year developed the disease, while in the second institution one case occurred on 105 guests. The convincing aspect of these observations is based on the fact that pellagra is an illness with a short period of latency: that is, unlike what happens with diseases predominant today, such as cancer, the onset occurs shortly after being exposed to the causal agent (in this case a nutritional deficiency) and regresses just as quickly.

Goldberger had high evidence standards (but, as we will see, low ethical standards), and was not satisfied with what he had achieved up to then. In the Mississippi state penitentiary he implemented the inverse trial: 11 volunteers were fed a simplified diet, poor in protein, while 108 prisoners were kept on a normally rich diet. All other living conditions were made highly comparable between the two groups, and possibilities of transmitting infective agents were minimized (actually the 11 volunteers subjected to dietary restrictions received better treatment). The "experimental" group received nutrition essentially based on wheat, corn, rice, cane sugar, potatoes and pork fat, for a total of 2.500–3.500 calories a day (against 3.500–4.500 in the control group). It should be noticed that the dietary restrictions were not applied immediately, but after a few months of observation of both groups to verify the absence of differences between them which could influence the results. The outcome of the trial was disconcerting: six individuals on 11 belonging to the experimental group and none of the control group developed pellagra. It is obvious that today no Ethics Committee would approve Goldberger's conduct, and it may be doubted whether the volunteers were really such. But the tireless researcher was not yet satisfied: starting from 1916 he conducted another series of trials to falsify the infective hypothesis. Adopting a "heroic" attitude which today would hardly be approved of from an ethical viewpoint, he persuaded 16 volunteers from the public Health Service (including his wife) to ingest and inject various materials (urine, faeces, blood) from patients suffering from pellagra. Nobody developed the disease.

1.2. *How much evidence is needed?*

I will not linger on other research – on a grand scale – conducted by Goldberger in different areas of the USA to understand what dietary deficiencies explained the geographical diffusion of pellagra. His last effort consisted in contributing to isolating the vitamin which was called PP (pellagra-preventing). This result was achieved thanks to a series of ingenious inferences beginning from observations and experiments, through a logic by exclusion: first he excluded that the protective agent could be a protein, as diets rich in soy or casein (and consequently rich in protein) could not prevent the onset of the disease, while a diet poor in protein but enriched with yeast was very effective in preventing it. Conducting experiments on dogs and cats, he established that the preventive agent had to be a heat-resistant agent, which was finally isolated by Elvehjem in 1937 (8 years after Goldberger's death) (Terris, 1964). The high standard of evidence used by Goldberger, if worthy of considerable respect from a scientific point of view, would now create many ethical and organizational problems. It is almost certain that today no researcher would be able to achieve what Goldberger managed within his lifetime; to give an example, a Canadian researcher has been trying to overcome the ethical problems arising from a study on lung cancer, based only on patient interviews, for at least a year (J. Siemiatycki, personal communication). One could question whether the evidence accumulated by Goldberger in 1914 was not already sufficient, and whether it was really necessary to organize the trial among convicts and the one among the 16 Health Service volunteers. This question actually raises an extremely intricate issue which has not been properly discussed up to now, related not so much to the "absolute" quantity of evidence we deem necessary, but rather to its "relative" quantity (and nature).

In other words: the question is not so much, in an abstract sense, how much information we must gather to be convinced by a particular hypothesis (a question correctly applied to basic sciences), but rather how much information we need in relation to the required preventive measures and to the consequences of these measures.

In physics and molecular biology it is legitimate to pose the epistemological question of the "evidence" and to compare different philosophies (based

for instance on verifying or falsifying the hypothesis). Medicine is different in that it is "a humanistic discipline founded on a scientific basis". Let us imagine there is a legitimate suspicion that a simple environmental intervention, which is cheap and of modest impact, could prevent an important disease. The example that springs to mind is that of Semmelweiss: simply washing one's hands could have prevented hundreds of deaths by puerperal fever. The intense hostility towards Semmelweiss is not so much rooted in the delay with which an important scientific hypothesis established itself, or in the denigrating campaign which was conducted against the Hungarian physician (of which Semmelweiss himself, affected by persecution delirium and by a self-destructive attitude, was responsible), but in the delay of the preventive measures related to the obtuseness of the academic environment (Nuland, 1989). The pellagra case is different for at least two reasons: on the one hand the suspicion that it be an infectious disease created the possibility that an excessive insistence on dietary origins could lead astray from a correct solution; on the other, suspicions on dietary origins entailed rather radical alterations in the diet of the poor population in the USA, weighing on family budgets. Having taken into consideration these precautions, there certainly is a threshold beyond which waiting for further evidence to adopt preventive measures is morally and practically unjustified; instead it is quite clear that – within certain limits – persisting in raising questions and satisfying scientific curiosity is legitimate from the point of view of the theory of knowledge, as long as systematic doubt does not interfere with preventive action.

1.3. *The epidemiological transition*

Goldberger is an interesting example because he represents the transition from the microbiological era (in which not only the prevailing causes of death were infectious, but infectious agents were sought for all diseases) to an era in which unbalanced diet or the poor quality of water start to be identified as causal agents. Instead of looking for an external, sufficient and necessary cause of disease, researchers like Goldberger focus on altered relationships between humans and their normal environment (e.g. diet). However, they are still within a mechanistic paradigm: niacin is in fact necessary for good health, and below a certain threshold the lack of niacin becomes sufficient to induce pellagra (with the threshold being characterized by individual susceptibility). Only in the 1950s will a probabilistic paradigm emerge, with the study of the effects of tobacco smoking, asbestos, heavy metals (for cancer), or saturated fatty acids for cardiovascular diseases: in these instances, i.e. for the prevailing diseases of the XX century, causes are neither necessary nor sufficient. We can say that most diseases that affect the population today in Western countries (cancer, infarction, stroke, diabetes, hypertension) do not correspond to the traditional paradigms of the infectious agent, of intoxication, or of a severe lack of nutrients. Not only the cause-effect relationships are probabilistic, but such diseases are inherently multifactorial, i.e. it is the combination of several circumstances that leads to a sufficient but not necessary causal complex.

2. On the basis of the precautionary principle

The definition and practical use of the precautionary principle imply a reference to scientific knowledge; however, this principle is the expression of an eminently political goal that concerns the defense of the environment and the protection of citizens from technological innovations.

In the last 10 years the precautionary principle has become the central notion at the international level for the prevention of environmental risks, based on the argument that where there are threats of serious and irreversible damage, the lack of complete scientific certainty should not be used as a reason to postpone the undertaking of cost-effective measures to prevent environmental degradation. The basic idea is to require precautionary measures also when the proof of a causal link between a potentially hazardous situation and negative consequences for the environment and public health is lacking, or when scientific knowledge is incomplete. There are two prerequisites that justify an appeal to the precautionary principle: (I) scientific uncertainty: the acknowledgement that often we lack certainty about long-term effects of many ecological processes; this is the consequence of an epistemic assumption concerning biological complexity, in contrast with a reductionist approach; (II) variability in the interpretation of data: the

acknowledgment that often there is disagreement between scientific opinions, as a reflection of the complexity of science, in contrast with a simplistic, neutral and objective image of science.

A large proportion of the conflicts and misunderstandings that characterize the precautionary principle derive from the vagueness of the concept itself that, being a general principle, has induced many interpretations. Two are the most extreme and frequently evoked interpretations, although at least 14 have been proposed (Foster, 2000). The first point of view corresponds to an "absolute" and rigid interpretation of the precautionary principle and transforms it into a metaphysical and paralyzing concept: not a pragmatic guide for decisions, but an abstract ethical concept. Instead, the second point of view uses the traditional scheme of risk assessment – i.e. a quantitative estimate of potential effects to health and the environment balanced with an estimate of benefits –, thus understating the novelties and specificities of the principle.

The precautionary principle has been introduced just because in many cases a balance of risks and benefits is extremely difficult to determine, mainly for lack of information: the circumstances in which it is not possible to invoke scientific certainty to avoid a prudential attitude are more and more frequent, particularly if the potential consequences are vast. The Italian philosopher Mariachiara Tallacchini has noted that the traditional approach of evaluation of environmental impacts – on which risk assessment is based – has a neutral attitude towards uncertainty (that is assimilated to a calculable risk), taking for granted that an objective result is always achievable; instead, the precautionary principle does not aim to be neutral, but rather expresses an orientation in favor of safety. Therefore, to appeal to the precautionary principle becomes relevant just when in spite of scientific uncertainty we need to make a decision whose consequences could deeply affect the community.

We are thus in front of a metaphysical use of the principle as an abstract ethical-political tool, loaded with strong symbolic meanings, on one side; or, as in the recent book by Lomborg (2001) (very dubious on scientific grounds) of a simplistic use of the principle, as a merely economic comparison of different alternatives that a technological choice allows, on the other side. These two extreme interpretations express, in fact, a potential ethical conflict, caused by the tension between the two attitudes that Max Weber called "ethics of responsibility" and "ethics of conviction". The first consists in evaluating, case by case, in a responsible and consequentialist way, the practical implications of our decisions; in the example of environmental choices this could be the approach based on risk assessment. The ethics of conviction is, instead, aimed at deep values and beliefs, and thus tends to be unbalanced and to lack flexibility. Weber's argument is that without an internal tension between the two types of ethics we fall into opportunistic attitudes (without principles) or into an idealistic and sterile attitude.

2.1. *Biological complexity*

A scientific justification of precaution as a principle was born first of all as a consequence of the crisis of traditional toxicology, that underlies risk assessment. In the light of recent studies, the investigation methods of traditional toxicology, in fact, are inadequate to predict the reactions of organisms characterized by a high level of complexity. For a correct interpretation of risks, to the concept of a single, necessary and sufficient cause, it is necessary to substitute the concept of a Lomborg (2001) plurality of causes, taking into account the delicate balance between the history, sequence, duration of exposures, the environment in which they occur, and the organisms that are exposed.

Traditional toxicology is essentially characterized by an analytical approach (each chemical substance is evaluated in isolation) and based on strong theoretical premises (in particular a threshold of toxicity). There are some persuasive examples of how an approach based on a case by case evaluation of exposures, that excludes the overall study of interactions among environmental exposures, and relies upon strong toxicological assumptions, is deemed to be misleading: (1) two substances, aniline and norharman, taken in isolation, are not carcinogenic, but administered to rodents combine to form a powerful carcinogen; (2) in experiments conducted in heavy smokers, who received vitamins for the prevention of lung cancer, results that were opposite to those expected were observed (an excess of cancers), most likely because the intake of vitamins has a different value in subjects who already have a mutation and in

subjects who do not; (3) the theory of receptors suggests that dioxin – a chemical that apparently does not cause damage to DNA – exerts its effects above a threshold, since it binds to the Ah receptor: this theory, however, is disconfirmed by epidemiologic observations in workers exposed to dioxin, who show a dose-response relationship even stronger at low levels of exposure than at high levels, with no evidence of a threshold.

We find many expressions of the characteristics of traditional toxicology in journals that are sponsored by the industry (Berry, 2001). Biological reductionism essentially consists in two strong statements: (1) environmental exposures that really count are few and act only at high levels of dose, (2) the control of a limited number of fundamental mechanisms can bring the solution for the cure and prevention of disease.

On the basis of the first point, there would be ''thresholds'' for the toxic action of chemicals, due to the existence of defense mechanisms, which would be effective only up to a certain concentration of the chemical. This is basically the biological model of ''intoxication''. One can, however, presume that cancer and other chronic degenerative diseases like diabetes and cardiovascular diseases – the prevailing diseases nowadays in Western countries – do not correspond to the classical model of intoxication, that is the type of damage that follows the action of a high-dose substance, that (1) binds to a receptor overcoming the metabolic abilities of the organism and causes a gross damage; or (2) grossly alters the functioning of an enzyme. This model is correct for arsenic or lead intoxication, not for chronic degenerative diseases. For the latter a model in which the concepts of timing of exposure, of biological balance and gene-environment interactions are essential is clearly appropriate.

The second point above, that is at the roots of many distorsions in the communication on drug effectiveness (for example therapies against cancer), derives from the century-old theory of ''magic bullets'', in turn related to the theory of receptors: for many biological phenomena there is a specific receptor, to which, for example, external chemicals bind. From the pharmacological viewpoint, acting on the receptor would allow the most effective solution to the problem. In other cases the magic bullet is targeted to a protein other than a receptor or, in the recent years, to a gene, but the substance of the reasoning does not change. We can consider as an example all the efforts that were devoted to find a cure for cancer in the p53 gene/protein.

A very clear example of the role of complexity is represented by the ''X syndrome'', the association between diabetes, hypertension, obesity and cardiovascular diseases (to which one can likely add colon cancer), so widespread in the western world. In all likelihood, the syndrome derives from the convergence of a substantial genetic stability of humankind (99% of our genes are the same as in the Pleistocene humans) and gross modifications in dietary life-style. Our metabolic genes are the same that characterized the hunter-gatherer, whose diet was irregular and prevalently based on vegetables. The lack of regularity in purchasing food selected the so called thrifty genotype, aimed at storing all the energy not immediately consumed. An example of how rapid changes in dietary habits can lead to dramatic consequences for health is represented by migrant populations or populations, like the Pima Indians in Southern California, who have drastically changed their lifestyle. In fact, Pima Indians, who traditionally ate chiefly vegetables and cereals and now have a typical ''American'' diet, rich in meat and refined foods, currently show the highest rates in the world for several of the diseases that compound the ''X-syndrome'': 37% of men and 54% of women have diabetes, and the frequency of obesity is extremely high.

We have mentioned colon cancer as a possible component of the X syndrome; in fact, this type of tumour, of which we knew so little until very recently, seems to be due to a mechanism similar to the one characterizing the X syndrome, the so called peripheral resistance to insulin. It is plausible to think that the accumulation of insulin as a consequence of an increased resistance of peripheral tissues becomes a proliferative stimulus that allows mutated cells to be selected leading to a neoplastic clone. Therefore, a hormone that is not usually mentioned among the carcinogenic risks (insulin) can represent, under given circumstances, a carcinogenic factor.

This is the context in which the problem of food policy in Western countries should be addressed. It is not a problem of ''intoxication'', but of a balance between genetic characteristics and dietary habits. From this point of view the food companies are not very health-oriented: they mainly use low-cost

dietary components like fat, refined sugar and salt, i.e. nutrients that contrast with our genetic background.

2.2. *The precautionary principle*

The precautionary principle was born in a context of environmental protection and of public health, and only in that context can it be correctly evaluated. It is not a question of "pure science", as some seem to believe: it is not to satisfy scientific curiosity that we investigate the speed at which animal species disappear, or the temperature of the planet increases, or what is the shape of the dose-response relationship for carcinogens. Sticking to the latter question, if the problems were just scientific, the uncertainties that surround it (is the relationship linear or non-linear at low doses?) would interest a limited portion of scientists; instead, the formulation of hypotheses on the shape of the relationship implies establishing whether the pollution levels to which we all are exposed as citizens do or do not have effects on health. Therefore the precautionary principle is not only a scientific problem, although it rests on scientific knowledge.

The principle is characterized by two elements: (a) an inversion of the burden of proof, from those who incur in a damage and have to show the cause-effect relationship, to those who can cause a damage and must show that a causal relationship is unlikely; (b) the establishment of a standard of proof: the decision about what and how far must be proved to consider the cause-effect relationship as likely (Bates, 1994).

The real philosophical problem is the second point: if applied literally, in fact, the precautionary principle can become trivial and paralyzing. The reason why it was formulated in that way is not related only to the requests of environmentalist movements, but also to the internal developments of biology, toxicology, and scientific ecology. As from the study of ecosystems we have learnt that they are extremely delicate and complex structures, similarly toxicologists should abandon a simplistic paradigm according to which damage is exerted only beyond a threshold for a gross interference with the cellular metabolism. In brief, it is not sufficient not to "poison" animal species or humans, since a damage can be done also by disturbing systems in a dynamic equilibrium, as in peripheral resistance to insulin.

If we understand the historical context in which the precautionary principle was born (in the last decades ecosystems have been damaged more than in all previous history) and we share the premises, then we can also commit ourselves to making it more effective and realistic. On scientific grounds it is possible – but not granted – that proteomics can introduce more effective methods to study toxic effects at low doses (by investigating the reaction norm of cells to toxic agents, through enhanced or reduced gene expression and protein profiles). On political grounds, people have correctly stressed that the precautionary principle cannot be unilaterally applied without considering alternatives. Sometimes to ban a chemical has overall effects that are more harmful than keeping it under social restrictions: the example of DDT and malaria is clear (Wynder, 1999). We cannot think nor hope that the introduction of the precautionary principle substitutes risk assessment, that is the quantitative evaluation of the balance between risks and benefits. Rather, the principle is introduced to reinforce the idea that scientific uncertainty cannot be used to avoid protective measures, especially when the consequences are vast. This is so also because there is no limit to uncertainty: science is naturally based on doubt and criticism. However, scientific uncertainty cannot be an obstacle to preventive action (Editorial, 2000).

References

Bates, D. V.: 1994, *Environmental Health Risks and Public Policy*, UBC Press.

Berry, C.: 2001, "The role of toxicology in evaluating priorities in environmental health", *Trends in Pharmacological Sciences* **22**, 277–280.

Editorial 2000, 'Caution required with the precautionary principle', *The Lancet* **356**, 265, 330–332.

Foster, K., Vecchia, P., Repacholi M.: (2000), 'Science and the precautionary principle', *Science* **288**, 979–981.

Lomborg, B.: 2001, *The Skeptical Environmentalist. Measuring the Real State of the World*, Cambridge: Cambridge University Press.

Nuland, S. B.: 1989, *Doctors. The Biography of Medicine*. Vintage Books, New York.

Terris, M.: 1964, *Goldberger on Pellagra*. Baton Rouge: Louisiana State University Press.

Wynder, E. L.: 1994, 'Studies in mechanism and prevention: striking a proper balance', *American Journal of Epidemiology* **139**(6): 547–549.

Dipartimento di Scienze Biomediche ed Oncologia Umana
Università di Torino,
Via Santena 7,
10126 Torino, Italy,
E-mail: paolo.vineis@unito.it

Fondazione ISI
Villa Gualino
Via Settimio Severo 65
10133 Torino, Italy