Applying Weight of the Evidence – Issues and Practice: A Report From the Workshop Held October 24, 1993

Compiled and Edited by:

Michael Gilbertson and Sally Cole-Misch

International Joint Commission

June 1994

… SCIENTIFIC PRINCIPLES

Mr. Glen Fox, Canadian Wildlife Service Ottawa, Ontario

… The following is a brief review of the epidemiologists' **criteria for causality**:

- The first one is **time order. Does the cause precede the effect in time?** This may be difficult to establish in systems with little historical data.

- The second is **strength of the association and asks whether cause and effect coincide in their distribution.** Is the prevalence of the effect in the exposed populations large relative to unexposed populations?

- The third is **specificity of the associations. Could the effect be due to different cause? Could the proposed cause produce other effects?** Can alternate hypotheses be eliminated? In the context of the Great Lakes, where a multiplicity of persistent toxic substances and ecological perturbations are present, specificity may be complicated by chemical interactions, commonality of the mode of action, and interspecific differences in the susceptibility of biota.

- **Consistency of the association is the fourth criterion. Has the association been repeatedly observed in different places, circumstances, times and species, or by other investigators with different research designs?**

- And finally, **coherence of the associations. Is the cause-effect interpretation consistent with our current understanding of biological mechanism(s) underlying the effect?** Is an exposure-response relationship present? Do laboratory studies support

the proposed relationship? Do remedial actions lead to altered frequency and severity of the effects? Only biologically plausible associations can result in biological significance, however, judgments on this basis are bound by our imperfect knowledge at any time.

Weighing the strength of evidence is always required. What is the nature of the evidence that must be ignored to conclude that no causal relationship exists? What alternate explanation will fit our observations and what other differences between our contrasted groups could equally, or better account for the observed incidences?

… **Cause and effect associations which are epidemiologically consistent should be confirmed experimentally, if possible** using extensions of Koch's postulates for proving that the particular pathogen causes a specific disease. **First we would do an experiment with controlled exposures of a susceptible organism to a concentration gradient of that chemical or suspected agent**, be it a complex-effluent or contaminated medium, that is associated with the effect in the field. From those controlled exposures we would expect to find a related gradient in the response. **The second strategy is to show, from analysis of samples from field studies, that the organisms in the field are exposed to the suspected contaminant or agent and that the degree of exposure is consistent with the degree of exposure that causes the effect in a laboratory animal**.

Economically and practically, it is far easier to regulate contaminants at the source of production, than to react after their release into the ecosystem. We should not wait for damage to occur and then try to fuc the situation. Instead we should use appropriate strategies to prevent the damage from occurring in the first place.

… **I think a paradigm shift like this will affect our viewpoint from which we assess the weight of evidence**. **As a society we must decide on the appropriate standards of proof for causality and the existence of adverse effects**.

At the moment we have the cancer population standard, which is one case in one million. There is a public health standard, which is one in 10,000 to one in 100. The doctor's standard is between one in 10, and one in a 100. The legal standard for proof of causality is greater than 50%. The scientific standard is greater than 95%, which is biased towards the prevention of "acceptance errors" rather than "rejection errors." **We must decide whether to use one of these criterion or one that is based on ethics, knowledge, experience and concern for the biosphere**.

…**Contrary to present administrative practice, in environmental decision-making it would be preferable to take action aimed at**

protecting or restoring a resource based on an erroneous causal relationship than to delay the decision for one or two decades and thereby risk losing the entire resource.

…IMPLICATIONS OF THE DAUBERT CASE

Professor Margaret Berger, Brooklyn Law School

… **The court was faced with a case in which the central issue was causation**. The litigation arose out of the use of the drug Bendectin, which for a while was the leading morning sickness remedy for women. **Bendectin WAS approved by the FDA (U.S. Food and Drug Administration), and never lost its approval, although it was eventually taken off the market by the manufacturer because of the more than 2,000 lawsuits that were ultimately brought. These charged that Bendcctin caused birth defects, primarily limb reduction defects.** So about 2,000 cases arose from the more than 20 or 30 million births to mothers who took Bendectin.

The central issue in all of these cases was causation.

… Second, the conclusion that comes out of *Daubert* is that **the court recognizes that science and law are different endeavours. If scientists are dissatisfied with the amount of data that they have acquired, they can continue to ask questions, they can ask for another research grant, they can continue questioning**. The Supreme Court in the *Daubert* case recognizes that for better or worse, a court, when an issue is legally ready for determination, must decide the question. It has no choice and the court says in *Daubert*,

> "**There are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision**. **Law, on the other hand, must resolve disputes finally and quickly**. The scientific project is advanced by broad and wide ranging consideration of a multitude of hypotheses, for those that are incorrect will eventually be shown to be so, and that in itself is an advance. Conjectures that are probably wrong are of little use, however, in the project of reaching a quick, final and binding legal judgment -- often of great consequence -- about a particular set of events in the past. **We recognize that in practice, a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations**."

That is the consequence; the court is going to have to decide the legal dispute even though it does not as yet have all of the information.

Now where does this leave us with *Daubert*? I think the **judges** have been given a number of messages. One of the messages is that they **cannot duck responsibility in some cases where controversial scientific evidence is being offered. They will have to do their best to at least decide whether factors…were looked at by the experts**. Did they look at the consistency of results? Did they look at rates of errors? Did they have a theory of plausibility? Exactly what is it that they did? And the courts will have to reject marginal evidence at times.

The court also suggests that there will be instances when scientific evidence will be admissible but the court might still have to decide based on legal standards that it is insufficient to prove the plaintiff's position. The courts are obviously going to have to decide what the legal standard is. I don't think it's at all clear at the moment. For example, one of the things that the court could have done in *Daubert* is to have spoken about statistical significance. It chose not to do so. Whether at some point there will be an effort to translate legal standards into statistical terms is at this point not at all clear. Lower courts and the intermediate appellate courts are obviously going to have to deal with that issue. (pp. 6-8)

…SCIENTIFIC INFERENCE AND THE PRECAUTIONARY PRINCIPLE

Mr. Jack Weinberg (Greenpeace, Chicago, Illinois) and Mr. Joe Thornton (Greenpeace, New York, NY) (p. 20)

In 1993, **the Governments of the United States and Canada accepted the International Joint Commission's (IJC) recommendation to use a weight of evidence approach in reaching conclusions about proposals to eliminate persistent toxic substances from the ecosystem**. **The IJC introduced this concept as part of its call for a precautionary set of cnvironmental policies, including the use of the "reverse onus" approach to chemical regulations**.

The IJC and governments must now more fully define the use and meaning of the term **"weight of evidence approach"** as it is used in this context. **We would like to share some thoughts on the use of a "weight of evidence" approach for evaluating scientific information in a precautionary policy setting**... (p. 20)

...**In defining a "weight of evidence" *or* "precautionary" approach to environmental policy,** the proper role of science is to generate theories and evidence, to suggest how these can inform public policy, and to evaluate the validity and relevance of cited scientific information to the policy matter under consideration... (p. 21)

…**Precautionary Inference**

Two of the most important applications of the precautionary principle are zero discharge for persistent toxic substances and reverse onus for synthetic chemicals. Even after these principles are adopted, however, weighing evidence in a precautionary framework is still required. There will be policy decisions to make, and these will be based in part on scientific information that remains, as always, incomplete, inconclusive, or indeterminate. **There must be some method of evaluating evidence that is consistent with a precautionary standard. This method can be termed precautionary inference**. (p. 24)

…**Precautionary inference provides a method for making scientific judgments based on incomplete, inconclusive or indeterminate data in a field in which significant harm may occur from a false negative judgment**. **Unlike the current scientific and policy framework, this approach reverses the burden of proof, framing the question with the null hypothesis: "What evidence must we IGNORE to conclude that a causal relationship does not exist?"** (p. 25)

… ***Shifting the burden of proof from society to those who advocate the production and use of chemicals not only changes the standard for policy decisions but has implications for the method by which evidence is weighed. Precautionary inference requires a holistic consideration of an integrated body of direct and circumstantial evidence. The central question of precautionary inference is, "What information must be ignored to conclude that there is no danger to health and the environment?"***
(p. 26)

WEIGHT OF EVIDENCE VERSUS PROOF OF CAUSATION

Dr. Rosalie Bertell, International Institute of Concern for Public Health
Toronto, Ontario

(pp. 27-31)

"…**We need to expand Hill's criteria and note the power of the test. The power**

of the test measures the type 11 error. I think that a lot of poor science has gone on, producing a very large number of studies that show nothing. Just because a study shows nothing does not mean there is nothing happening. I would tell you I know lots of ways to design studies so that no relationship between exposure and illness shows. Anybody can do that. It takes a little more skill to design a study where some relationship does show. What you need to know is the power of the test or the probability that you will accept that null hypothesis as true when it is wrong. Every study should report its power. It is rarely reported. By being more demanding that a type I error not occur we increase the risk of making a type 11 error.

I think the other problem that we have is that **the Hill criteria were based on a linear system, not an ecosystem approach. When you have competing causes of death you cannot expect a linear dose-response.**

…You are not going to get the same dose-response when you have competing causes of death. **You have to have a wider and broader approach to health than a particular criterion expecting a dose-response, which is always responsive to the same degree under all circumstances**.

I think there are other problems with Hill's criteria, which are brought up nicely in the Jacobson study (p. 9-15), in which **the dose-response factor can also depend on the point in the life cycle at which the exposure occurs**. You might not get a dose response with the breast milk but you do get the dose response with in utero exposure. **You have to know the point at the life cycle that the exposure elicits a biological response.**
(p. 28)

…Hill was primarily concerned with severe observable health damage in an exposed person. As a medical researcher, I am concerned not about choosing severe end-points like cancer death, but rather **I am anxious to identify biomarkers at the point where the situation is reversible**. That means a radical change in research orientation. It means looking at biological end-points that are less dramatic than cancer or genetic damage.

… One of the things that we have to do is to **start looking at earlier bioindicators of deteriorating physical well-being and of early signs of deteriorating vigour in the species which might serve as early warnings of trouble**. We have done some work on this approach and it is possible. However **such an approach demands that one not wait for definitive confirmation of causality.** It is better to demonstrate probable causality by an intervention to improve health. ***Weight of evidence* calls for intervention when causality is expected to**

be confirmed, if the deteriorating situation is allowed to go to its logical conclusion.
(p. 29)

… **With respect to the *hazard* I think the burden of proof, at least on many important questions, needs to be a reverse onus. There should be a need to prove something is not damaging before it is used**, and the burden of proof should not be on the victim to say a toxicant is connected with a health problem. I think there are some very good models for testing of pharmaceuticals that could be used in this respect to screen chemicals before they are put into the environment. I would also recommend establishing a health review board that would be at arm's length from industry and government, that would review new projects. Our environmental assessments do not include human health. They are very superficial in that regard and I would call for a health assessment of every major new project.

…**I would also recommend that we move from the relative risk statistic to a little more sophisticated one which is called the "attributable proportion." It is a derivative statistic**. There has been a lot of development of this statistic within the last five to seven years. It was first proposed about 1970. The attributable proportion is a statistical quantity which would let you estimate, for example, what proportion of lung cancers are due to a particular exposure. You might say 17% are due to smoking and 2% are due to radon gas, and so on. You can begin to attribute proportions. That gives you an upper limit for the possibility of improvement."
(p. 30)