1994-06-01

Applying Weight of Evidence: Issues and Practice. A Report on a Workshop Held October 24, 1993, Windsor, Ontario

International Joint Commission

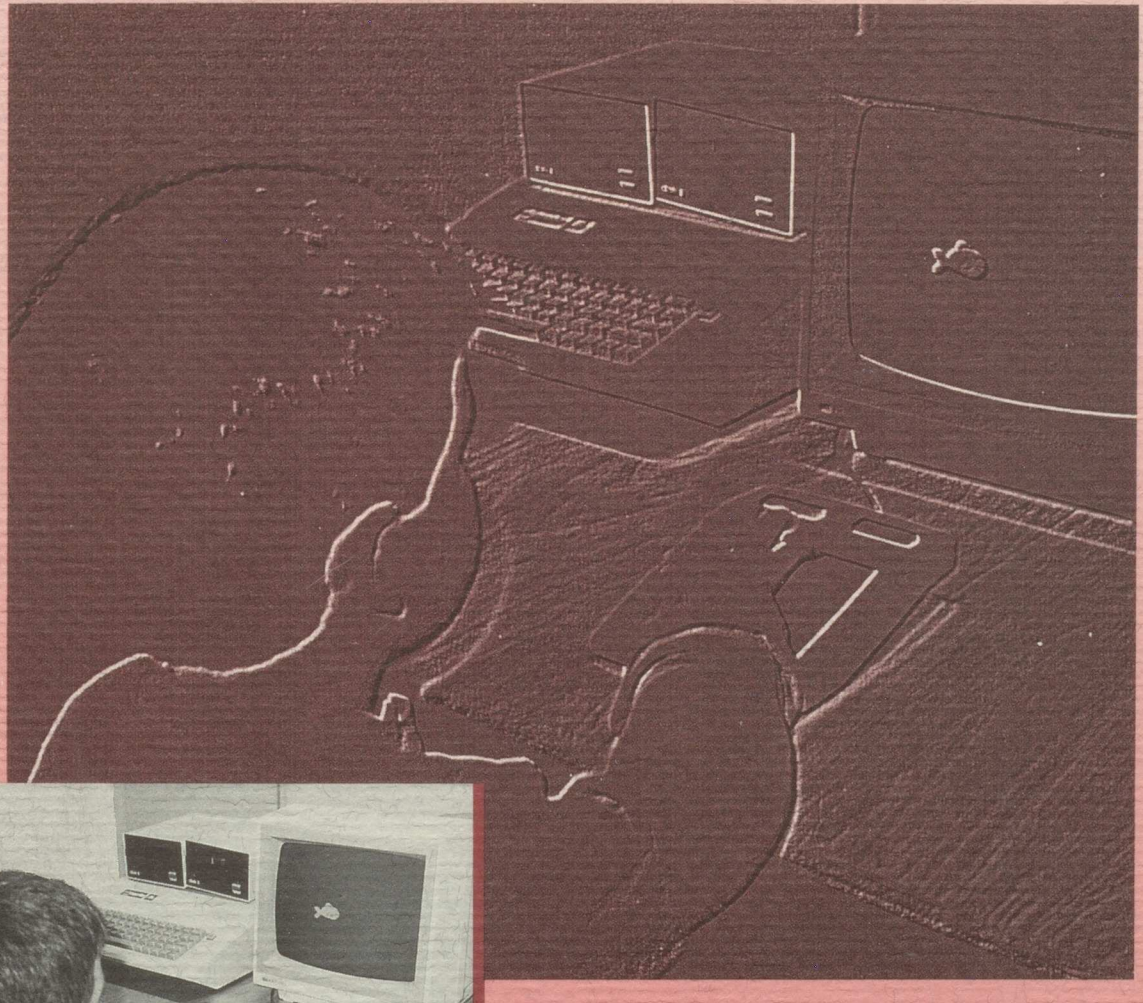
Michael Gilbertson

Sally Cole-Misch

Applying Weight of Evidence:

Issues and Practice

A Report on a Workshop held October 24, 1993



International Joint Commission
Commission mixte internationale

APPLYING WEIGHT OF EVIDENCE:

Issues and Practice

A Report on a Workshop held October 24, 1993

Windsor, Ontario

Compiled and Edited by:
Michael Gilbertson
and Sally Cole-Misch

International Joint Commission

ISBN 1-895085-81-0

June 1994

Cover photo:
Testing intellectual and cognitive development in 4-year olds, part of the studies undertaken by Drs. Joseph and Sandra Jacobson

Printed in Canada on Recycled Paper

TABLE OF CONTENTS

Acknowledgements iv
Disclaimer iv
Preface v

INTRODUCTORY REMARKS
Mr. Gordon Durnil, United States Chairman, International Joint Commission 1

SCIENTIFIC PRINCIPLES
Mr. Glen Fox, Canadian Wildlife Service 2

IMPLICATIONS OF THE DAUBERT CASE
Professor Margaret Berger, Brooklyn Law School 6

CHILDREN OF GREAT LAKES FISH CONSUMERS
Dr. Joseph Jacobson, Wayne State University 9

BASIS FOR REMOVING BIOLOGICALLY ACTIVE PERSISTENT TOXIC SUBSTANCES
Dr. William Owens, Procter and Gamble Company 15

SCIENTIFIC INFERENCE AND THE PRECAUTIONARY PRINCIPLE
Mr. Jack Weinberg and Mr. Joe Thornton, Greenpeace 20

WEIGHT OF EVIDENCE VERSUS PROOF OF CAUSATION
Dr. Rosalie Bertell, International Institute of Concern for Public Health 27

DISCUSSION 32

TABLE

1. McCarthy Memory Scale scores by maternal milk PCB level and duration of nursing 13

ACKNOWLEDGEMENTS

The workshop proceedings were transcribed/processed by Mary Ann Morin and Karen Jeffrey, and produced by Bruce Jamieson.

DISCLAIMER

This report is a compilation of presentations, and subsequent discussions, at a workshop held October 24, 1993 during the 1993 Biennial Meeting of the International Joint Commission, in Windsor, Ontario. While the Commission supported this workshop, the views expressed do not necessarily represent the views of the International Joint Commission.

PREFACE

The International Joint Commission's 1993 Biennial Meeting on Great Lakes Water Quality, included a Workshop on Weight of Evidence. The Commission, in writing its Sixth Biennial Report on Great Lakes Water Quality in 1992, stated that unequivocal evidence has been presented to confirm cause-effect linkages between specific persistent toxic substances and specific adverse impacts in fish, birds, turtles and various mammals.

The Commission recognized in its report that scientific data are open to interpretation and that, notwithstanding the confirmed cause-effect links, unequivocal conclusions may be difficult to reach, especially if individual studies are considered in isolation. Unequivocal evidence of injury to humans caused by exposures to persistent toxic substances may be difficult or impossible to obtain because of the subtlety of the effects associated with low contaminant concentrations and the potentially confounding factors.

The Commission noted that critics have attempted to find flaws with individual studies in order to discredit findings and conclusions about persistent toxic substances. While limitations to study design may exist, the Commission concluded that these did not necessarily invalidate the findings and conclusions when considered in a weight of evidence context. It therefore adopted a "weight of evidence" approach: when evidence from the many studies that indicate injury or the likelihood of injury is taken together, it provides sufficient justification for the virtual elimination of the discharges of persistent toxic substances to the Great Lakes.

The Commission also advocated the use of this weight of evidence approach by the Parties to identify other substances suspected of being persistent and toxic and that therefore should be subject to the policy contained in the Great Lakes Water Quality Agreement concerning their virtual elimination.

The Commission continues work to better define what is meant by weight of evidence. For example, when should a cause and effect approach be used rather than a weight of evidence approach, and what are the specific differences? Do these methodologies relate to prospective as well as retrospective situations? How do these methodologies differ from risk assessment and what are the advantages and disadvantages of each approach?

At the workshop, representatives from industry and public interest groups participated with members of the scientific and legal professions to contribute their points of view. There is no doubt that the Commission's chlorine recommendation in the Sixth Biennial Report -- that the Parties, in consultation with industry and other affected interests, develop timetables to sunset the use of chlorine and chlorine-containing compounds as industrial feedstocks -- precipitated a vigorous debate on the evidence and on the recommended solution. These proceedings of the Weight of Evidence Workshop at the 1993 Biennial Meeting are designed to help resolve this debate by clarifying the terms and methodologies, and the circumstances under which they should be used.

INTRODUCTORY REMARKS

Chairman Gordon Durnil
International Joint Commission
Washington, D.C.

I want to welcome you. I have been looking forward to this session with a great deal of interest. We have a very interesting panel. As you know, the International Joint Commission has recommended in our Sixth Biennial Report that the Governments of the United States and Canada adopt a "weight of evidence" approach. Both governments, and the Government of Ontario, have now accepted that recommendation, so we are going to try and figure out what it means that our governments have accepted. Is it Perry Mason terminology? Is it understood only by lawyers? Is it the measuring of scientific reports -- that the tallest stack wins? Can it be identified by applying percentages? Fifty-one percent of the scientists agree that brown is blue, therefore the weight of evidence must be that brown is blue. I don't think it's any of that, at least in my view.

As a lawyer, I want you to pass out of your minds the way lawyers use words. Don't think about a preponderance of the evidence in civil trials, and beyond a shadow of a doubt in criminal trials and so forth. Forget all that. In our use of weight of evidence, we really are not looking for absolute truth, but rather for the potential for adverse effects on the environment, including humans. And I think the word "potential" is critical to this debate and this definition we are looking for. Most of the accepted definitions of risk relate to the potential for adverse effects. Now the questions that are on my mind center around such things as, and I hope you will all resolve them here today, how do we know when there is sufficient "evidence" or accumulated "knowledge" or enough "potential for harm" so that we should expect a reasonable person to assume that scientists should sound the warning and policymakers should act? Do we look at each scientific study and weigh the nonquestionable conclusions with those that are questionable? Do we reject all conclusions in a study if one or some of the conclusions cannot be proven beyond some level of doubt? Even though it's possible to pick holes in every study, especially on methods, does not a definitive time come when there is enough evidence upon which to act?

> *... if the scientific community has grave concerns about potential threats from a substance being introduced into the environment, but they don't have sufficient data, especially human data, to definitively declare a substance harmful, but they suspect they may have such proof when the next generation reaches puberty, should they not come forward with evidence they now have, even if such evidence is more suspicion than fact?*

On a broader scale, do we just deal with scientific studies in applying a weight of evidence approach or do we also consider the perceptions of lay people in a specific community? Last week I attended a **RAP (Remedial Action Plan)** review meeting in Presque Isle, in Erie, Pennsylvania. We were out at the bay with a number of governmental people from both countries, looking at the waters, standing on the edge, and a couple of old codgers came up and they were bearded and sort of grizzly looking, like they may have been the grandfathers of some of the guys playing for the Phillies in the World Series (against that other team). One of them had an orange hunting cap and a fatigue jacket and the other one had on one of those camouflage hats and camouflage coats, so I couldn't tell what he

looked like. But they walked up and said, "You guys are standing in our fishing spot." So we moved, and we said, "You been fishing here long?" And they said, "Yes, since the 1930s, been here every day." And we said, "Anything different?" And they said, "Yes, everything is different. There used to be a lot of fish, we used to get all of our meals out of here, now we don't catch very many fish. There's not much out there, the ones we catch have sores all over them, and you can't eat them. The sea gulls are not here, we saw just one today and nobody cares." When we asked what had caused the problems, they pointed out a coking plant and a couple of factories across the bay.

So the only reason I'm telling that story is I think those guys are part of the weight of evidence, and I think we have to consider their perceptions and their reactions. There is a very interesting article in *Environment Magazine* last month, which makes the point that quite often the public sees the problem, sees what they think is the cause and ties the two together in their minds very quickly. And then a whole body of scientists from industries, from government, from elsewhere comes along and tells them that can't be the case and maybe 15 or 20 years later we find out it is the case. So I don't think we can leave the perceptions of the public out of the weight of evidence, but we'll hear what all of you have to say.

For a final introductory thought, if the scientific community has grave concerns about potential threats from a substance being introduced into the environment, but they don't have sufficient data, especially human data, to definitively declare a substance harmful, but they suspect they may have such proof when the next generation reaches puberty, should they not come forward with evidence they now have, even if such evidence is more suspicion than fact? Is that something that we should talk about here today? If not, then how do we deal with preventive measures in the face of scientific uncertainty?

So, those are some of my questions and we have a panel here to give us their expert view and their subject matter. The panel includes Mr. Glen Fox of the Canadian Wildlife Service, discussing scientific principles; Dr. Joseph Jacobson of Wayne State University, who will give results of his research on children of Great Lakes fish consumers; Professor Margaret A. Berger, Brooklyn Law School, will discuss the implications of the *Daubert* case; Dr. William Owens of Procter and Gamble Company will present research on the basis for removing biologically active persistent toxic substances; Mr. Jack Weinberg of Greenpeace will advocate the precautionary inference; and Dr. Rosalie Bertell of the International Institute of Concern for Public Health will talk on weight of evidence versus proof of causation.

2

SCIENTIFIC PRINCIPLES

Mr. Glen Fox
Canadian Wildlife Service
Ottawa, Ontario

We are all, with all the other biota in the Great Lakes ecosystem, unwitting subjects in an unknown number of natural experiments. They are unknowingly initiated, and have not been through the statistician's office for his input on design, they are not screened or approved by an animal care or a medical ethics committee. At some later date, you or I, or somebody we know will make some observation or an event will occur that will bring our plight to our attention.

Then we will initiate studies based on group, or on population characteristics and compare the effects on different populations or groups in the hope of relating the observed differences to differences in the local environment, or lifestyle of these individual populations. Such ecological correlations provide clues to causal or ecological hypotheses that may be tested in individuals.

Observations of the apparent effects of contaminants on free-living fish and wildlife and human health are always correlational. That's what we have to work with. Potential causal agents considered are those we measure or observe and they are probably only a subset of those present. We start out with imperfect knowledge at best. Free-living organisms are exposed to a number of contaminants and stressors and the effects we observe or measure are the organism's integrated biological response to that suite of stressors. We must not fall prey to what is known to epidemiologists as the "ecological fallacy," the idea that occurrence of an effect in conjunction with a plausible environmental factor proves that the factor is the cause.

These observations are often all we can work with. We need to draw together the disparate threads of evidence and make them into some sort of coherent whole so that we can scientifically and ethically make socially defensible regulatory decisions.

Lilienfeld and Lilienfeld, two very prominent epidemiologists, suggest that in medicine and in public health it would appear reasonable to adopt a rather pragmatic concept of causality. They wrote as follows: "A causal relationship would be recognized to exist whenever evidence indicates that the factors form part of a complex of circumstances that increase the probability of the occurrence of the disease and that a diminution of one or more of these factors decreases the frequency of that disease."

In disease prevention, it is initially only necessary to identify an association between exposure to a critical factor and the incidence of disease without necessarily identifying the ultimate cause of the disease. For example, we can protect people from lung cancer by persuading them to stop smoking, long before we can figure out all the factors that are in the tobacco smoke that cause the problem. In 1964, the U.S. Surgeon General's Advisory Committee on Smoking and Health made the first attempt to address the relationship between tobacco smoking and lung cancer using epidemiological evidence. The committee concluded that statistical methods alone cannot establish proof of a causal relationship; the causal significance of an association is a matter of judgment, which goes beyond any statement of statistical probability.

For example, we can protect people from lung cancer by persuading them to stop smoking, long before we can figure out all the factors that are in the tobacco smoke that cause the problem.

We must, therefore, have a basis for deciding whether a statistical association derived from an observational study represents a cause and effect association. To do this we systematically evaluate the evidence using the criteria that have been established by epidemiologists. These criteria provide one means of objectively evaluating the relationship between a suspected cause and associated effect. It's a process and a framework upon which we can build a balanced judgment. These criteria are of greater assistance in rejecting causal hypotheses than in confirming them. We can use them to deal with the quality of the evidence that we have; not necessarily to measure the quantity of evidence. And they will provide us with the means of deciding what evidence is admissible, so to speak, before we start to weigh the quantity of evidence.

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 a 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? One of the fathers of epidemiology, Sir Austin Bradford Hills, wrote that all scientific work is incomplete, whether it is observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us the freedom to ignore the knowledge that we already have, or to postpone the action it appears to demand at any given time. Complete logical certainty is not available in science. The best we can do is reach the most reasonable explanation based on the evidence at hand.

We have tried to apply these criteria to case studies on populations of fish, wildlife and human health in the Great Lakes basin. In 1989 in Chicago, I presented these criteria and a number of people then applied them to their data. These case studies and the criteria were published in the August 1991 issue of the *Journal of Toxicology and Environmental Health* as the proceedings of the First Cause-Effect Linkages Workshop. The Second Cause-Effect Linkages Workshop was held in association with the 1991 Biennial Meeting at Traverse City in Michigan where more data was presented. The proceedings of that workshop were published in the December 1993 issue of the *Journal of Great Lakes Research.*

More recently the Chlorine Institute contracted CANTOX to look into the issue of whether or not there is a toxicological problem in the Great Lakes that might be related to chlorine, and to elucidate the scientific principles for evaluating the potential for adverse effects of chlorinated organic chemicals. CANTOX, a very well-recognized toxicological contract group, used these same criteria and proposed their use as an adequate way of approaching this question. So I think we are talking about something that has been tried and tested.

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 fix the situation. Instead we should use appropriate strategies to prevent the damage from occurring in the first place. In recent political discussions in our two countries and at this 1993 IJC Biennial Meeting, we have repeatedly heard our neighbours, our children, our constituents, our taxpayers, our board members and our employees tell us that there must be a fundamental change in thinking of industry, government and society. I think that ethical issues are too often subservient to legal and economic issues. To protect human health we need to consider what is ethical rather than what is legal or least expensive. We've talked a lot about protecting human health, but we also have to protect biodiversity and the planet Earth and to do that we need to consider what is ethical instead of what is permissible.

This change requires that humankind recognizes its true place in relation to this world. We are part of Nature, inseparable constituents of the ecosphere and that is a truth that cannot be denied. One of my favourite thinkers in the field of modern ecology is Stan Rowe from the University of Saskatchewan, who has written a wonderful book called *Home Place* in which he has had a hard look at where the world is going and what it is we need to do to realign our thinking to be more compatible with our continued existence on this biosphere. Stan puts it this way, "Nature is where we come from and where we belong in our earthly existence. Nature, (i.e. the ecosphere) is home, with responsibilities for care and affection and aesthetic concern that the word 'home' implies. To be at home means asking ourselves about our intentions of staying on, about care of the furnishings and their maintenance, about sympathy for the other occupants and their welfare. These are all matters with powers to initiate fundamental revolution in the practice of our arts and sciences and in time becoming our second nature as we prepare to minister to the natural home place."

One of the fathers of epidemiology, Sir Austin Bradford Hills, wrote that all scientific work is incomplete, whether it is observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us the freedom to ignore the knowledge that we already have, or to postpone the action it appears to demand at any given time.

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.

In environmental decisionmaking, it is preferable to have lots of data, but in the end we have to use experience. Scientists have traditionally been obsessed with not being wrong in reporting that some phenomenon was occurring or that it was caused by some factor. Contrary to present administrative practice, in environmental decisionmaking 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 A. Berger
Brooklyn Law School
Brooklyn, New York

I'm here, I assume, to speak about the legal standard of causation and the impact, if any, of the Supreme Court's recent decision in *Daubert v. Merrell Dow Pharmaceuticals* which was decided on June 28, the very last day of the Supreme Court's '92-'93 term. This is the first time that the Supreme Court has ever considered what the standard should be for the admissibility of scientific evidence. 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 Bendectin 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. Interestingly enough by the time this case came to the Supreme Court, only one plaintiff had ultimately managed to win and even that plaintiff has never been paid. In the remaining cases, juries either found for the manufacturer or the trial court set aside jury verdicts for the plaintiff, or even if the trial judge allowed a jury verdict for the plaintiff to stand, the appellate court had set aside the verdict.

By the time the *Daubert* case -- which was a case coming out of the Ninth Circuit in California -- had to be decided, there were all these other cases rejecting proof of causation. The trial judge in *Daubert* granted summary judgment for the plaintiff. In other words, on the basis of affidavits submitted by experts for the plaintiffs and the defendant, the trial judge found that the defendant would not be able to succeed as a matter of law. The case went up to the Ninth Circuit, the three-judge appellate court, which in a very summary opinion by Judge Kozinski said: first, the appropriate standard to apply was the "general acceptance" standard of what the scientific community agreed on, the so-called *Frye* standard that came from a 1920 case dealing with lie detectors, and second, pursuant to that standard, plaintiffs' expert testimony wouldn't have been admissible because it hadn't been peer reviewed. Therefore the Ninth Circuit concluded that it would throw the case out and affirm the grant of summary judgment.

This went to the Supreme Court, probably in part because what the judge said in affirming was so much broader than what needed to be said to dispose of the case. By suggesting that no scientific evidence would ever be admissible unless it had been peer reviewed and by also suggesting that the only test was "general acceptance," the judge was perhaps overly broad in his analysis. Anyway, the Supreme Court took *certiorari* (a writ to call up the records of an inferior court) and the case was argued before the Supreme Court. There was enormous interest in the case and 22 *amicus* briefs were filed. Groups with an interest in science, as well as members of the corporate bar, the plaintiff's bar, and persons interested in issues of state versus federal law all somehow managed to find a basis for writing a brief in *Daubert*. I also wrote an *amicus* brief on behalf of the Carnegie Commission on Science, Technology and Government.

Interestingly enough, by the time the case was actually argued in the Supreme Court, and I was at the oral argument, neither side was arguing exactly those issues that the Supreme Court had certified for review. Neither side was really saying that all evidence that is to be admitted has to be peer reviewed and neither side had a good word to say for the *Frye* "general acceptance" test. What it really boiled down to between the plaintiffs and the defendant was that the plaintiffs were saying: we have qualified experts. There is no contention in this case that the experts were not qualified. They have perfectly valid degrees, they have terrific CVs (curricula vitae), they all had experience in the fields in which they purport to be experts. The plaintiffs were saying that once you have an expert like that, an expert with credentials, the court has to allow such an expert to testify. The defendants were saying that's not enough -- there has to be a foundational inquiry before a court will allow an expert witness to testify, an inquiry as to whether the expert has a theory that has been sufficiently validated to be of assistance to the court in this case.

The result in *Daubert* was that the Supreme Court reversed the grant of summary judgment, meaning that the plaintiffs get another chance, but I do not think that the plaintiffs will ultimately succeed. The Supreme Court said that there is a "gatekeeping" function for the federal judge -- that the judge must make a determination before the judge allows a qualified expert to testify. Now that, of course, brings us to the crux of *Daubert*. What will that determination consist of? What is it that the expert has to be able to say? Well this is also where the court's opinion gets a little vague and *Daubert* certainly is not the end of all discussion on how a court makes this determination on the admissibility of scientific expert testimony. You can't have a magic formula for this kind of a case.

What I think is important are some of the things that the court acknowledged in the course of reaching its decision. One was that it recognized that there are questions for judges, and other questions for scientists, and that the judge is not to just look at the scientific product and say, do I agree with the results? Do I now agree that Bendectin does or does not cause birth defects? Rather, the judge should seek to ascertain whether the expert employed a proper methodology in reaching his or her conclusion. Is what the expert did or relied on in reaching his or her opinion in this case consonant with a scientific method? A court must at least be able to see, if the expert is claiming that tests reveal such and such, whether there really were tests? How were those tests done? What was the rate of error in the tests? Those are the kinds of issues that a court must look at.

This went to the Supreme Court, probably in part because what the judge said in affirming was so much broader than what needed to be said to dispose of the case. By suggesting that no scientific evidence would ever be admissible unless it had been peer reviewed and by also suggesting that the only test was "general acceptance," the judge was perhaps overly broad in his analysis.

One of the other Bendectin cases, the *DeLuca* case, was decided in a different circuit; the Third. That circuit had also reversed a grant of summary judgment for defendant and had said to the trial judge, we don't have enough of an explanation here as to why you think there is something wrong with the expert proof. When the trial judge went back and brought in the experts, and had an evidentiary hearing, some very interesting matters turned up. The plaintiff's principal expert could not account for some of the figures that he had said he was relying on. It turned out, for instance, that he had plugged in numbers from intermediate studies by authors who subsequently had finalized their epidemiological studies and corrected their original numbers. He had calculated risk ratios for studies that hadn't put down any risk ratios and he could not explain to the court how he had gotten to those risk ratios now that he was being asked specific questions. In addition, it turned out that the plaintiff's second expert, who did some reanalyses of data dependent on the numbers from the first expert, had never independently verified his numbers at all. So the trial judge in *DeLuca*, this other Bendectin case, granted summary judgement again after he held this evidentiary hearing when the case was sent back to him, and this time the circuit affirmed.

The Supreme Court in *Daubert* seems to me to have had a case like *DeLuca* in mind. The court is really saying to the trial judge, it's not that you have to be a scientist and understand what this result is, but you can at least ensure that questions get asked about how this scientific work was actually done, before you allow an expert to express an opinion. If the methodology was flawed the expert proof must be excluded.

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."

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. . ."

That is the consequence, the court is simply 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, such as for instance, those shown by Glen Fox in the previous talk, 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.

Finally, there have been many, many panels on *Daubert* since the opinion came out. In speaking to judges, the main impression I get is that they feel that they need to know a lot more about the scientific method. I think they will be turning to the scientific commu-

nity to find out, for instance, what are the hallmarks of a properly conducted epidemiological study? What are problems with animal studies? How should one deal with the interrelationship between an animal study and an epidemiological study? Issues of causation are, of course, not going to go away.

At bottom, the Bendectin litigation probably was a relatively easy case because none of the evidence pointed to causation in a very meaningful way. In addition to which, and I think that this is what ultimately impressed jurors and judges, the defendant started doing some very sophisticated work showing that in particular communities the rate of birth defects remained the same before Bendectin was on the market. Of course, it's also true that there are lots of other substances out there that perhaps can cause birth defects, and none of these studies showed the courts exactly what else was on the market at the same time that Bendectin was not and Bendectin was. But that is the nature of the problem in these kinds of cases. So I would hope that those of you who are scientists might have suggestions for the legal community on how to translate some of the attempts you are making to distinguish between good science and bad science into criteria that the courts can utilize.

Finally, there have been many, many panels on Daubert since the opinion came out. In speaking to judges, the main impression I get is that they feel that they need to know a lot more about the scientific method. I think they will be turning to the scientific community to find out, for instance, what are the hallmarks of a properly conducted epidemiological study? What are problems with animal studies? How should one deal with the interrelationship between an animal study and an epidemiological study? Issues of causation are, of course, not going to go away.

CHILDREN OF GREAT LAKES FISH CONSUMERS

Dr. Joseph Jacobson
Wayne State University
Detroit, Michigan

For several years now my wife Sandra Jacobson and I have been doing research on a cohort of children whose mothers ate relatively large quantities of Lake Michigan fish during the 1970s. The children were born in 1980 and 1981 and have been studied prospectively from birth. We are now completing an 11-year infant followup, but the only data that are complete and that I can talk about today are for the infant to four-year followup phases. Because we cannot experiment on these children and randomly assign them to different exposure levels, the studies of course are correlational by definition. The key feature of the correlational method that was used is the control for potential confounding variables, that is, to control for as many influences on these developmental outcomes as possible. The objective was to determine the degree to which the prenatal exposure to **PCBs (polychlorinated biphenyls)**, which was the major contaminant in Lake Michigan fish at the time, may have affected developmental outcomes. We were looking primarily at intellectual and cognitive development, although we also looked at behavioral development and physical growth.

In the 1970s, Harold Humphrey of the Michigan Public Health Department had found elevated PCB levels in blood sampled from Lake Michigan fishermen and a moderate correlation between the amount of Lake Michigan fish that the fisherman ate and the

level of PCBs in their blood. It is clear that consumption of these fish was a major source of PCBs for the fishermen. On the other hand, there was no evidence of health effects or physical anomalies. Our study was undertaken in response to an initiative of Dr. Wayland Swain, who was director of the U.S. EPA (U.S. Environmental Protection Agency) Large Lakes Research Station at Grosse Ile at the time. Swain surmised that, even if the adult fishermen were not affected, infants and children might be more vulnerable to this kind of exposure, especially if the exposure was to the fetus *in utero*.

The study began in July 1980 and lasted over a 16-month period. Over 8,000 women were interviewed in four major maternity hospitals located near Lake Michigan. They were asked in detail about the Lake Michigan fish they had eaten during the prior year, species by species. If they had eaten more fish in the past, they were asked about that period as well. We came up with a summary measure of contaminated fish consumption, which we defined as the annual Lake Michigan fish consumption, in the present or past, whichever was greater. The final sample included 313 infants, of whom 242 were from families where the mother reported elevated levels of Lake Michigan fish consumption, and 71 were from families near Lake Michigan where the mothers had not eaten Great Lakes fish.

Even though we recruited both fish eaters and non-fish eaters in the sample, I have to emphasize that the study was not designed as a comparison between an exposed group and a control group because everyone in western industrial countries, such as ours, is going to have some levels of PCBs in his or her body. What we were measuring in the studies was prenatal PCB exposure, and you do not have to eat Lake Michigan fish to become exposed to PCBs. In fact, we found elevated PCB levels in some instances in mothers who ate no Lake Michigan fish. Although it is clear that eating Lake Michigan fish during the 1970s increased your risk for high levels of PCB exposure, there were, and continue to be, other routes of PCB exposure.

The primary index of prenatal exposure was from umbilical cord blood samples from a subset of the children. Because PCBs have a long half-life in biological tissue, the cord-serum can give you a record of *in utero* exposure. We were also interested in postnatal exposure through breast feeding because PCBs are lipophilic and concentrate in breast milk. We wanted to assess the degree to which exposure from breast feeding might put the child at risk. The level of PCBs in the breast milk and the amount of breast milk consumed by the infants were used to estimate how much of the contaminated milk the infant consumed. In our four-year followup study, we also got levels of the child PCB body burden from blood serum samples.

The blood and breast milk samples were analyzed by packed column gas chromatography at the Michigan Department of Public Health. We were using the Webb-McCall method based on total PCBs. We did not have the technology at that time to evaluate specific PCB congeners, and the data that I will be reporting today are based on total PCBs. We know that individual PCB congeners differ considerably in terms of their toxicity and the kinds of effects they will have on biological systems. Unfortunately, we have no basis for speculating about which congeners may have been responsible for the effects that we saw. The children who were exposed to higher levels of PCBs were also exposed to higher levels of dioxin and dibenzofurans, and it could be actually those contaminants, which co-occur with PCBs in the environment, that were responsible for the effects that we saw. I am going to talk about PCB exposure, but we probably should consider our PCB measure as a marker for an environmental exposure, since the precise chemical composition is not yet known.

Infants were assessed at birth, at seven months and at four years. As I said in my introduction, the biggest problem in this kind of human correlational study, or any human exposure study, is the risk of spurious correlation. Because subjects cannot be randomly assigned to control for potential confounding influences, the strategy was to measure as many other factors as possible, known or suspected to affect the outcomes being studied and then to control for those other influences statistically. Twenty-four potentially confounding in-

fluences were included as control variables that were measured in connection with the four-year followup study. These included prenatal exposure to alcohol, maternal smoking during pregnancy, mother's age, sex of infant, perinatal medical complications, mother's IQ, HOME Inventory, which assesses the quality of intellectual stimulation provided by the parent, familial stress, and so forth. Since all of these could impact on the intellectual and behavioral development of the children, they were all measured.

Our statistical strategy is based on the premise that a third variable, a possible confounder, cannot be the true cause of an observed deficit unless it is related both to the exposure and to the outcome. We selected control variables based on those known or suspected to affect the outcome. We then controlled statistically for all the potential confounders that related, even weakly, to exposure using a criterion of $p > 0.10$ so that any third variable that was even weakly related to exposure was controlled statistically in all the analyses. In all the results that I will be reviewing with you today, a toxic effect was inferred only if the exposure was associated with the outcome after controlling for any potential confounder weakly related to the exposure. In addition, in this research all of the individuals involved in testing the infants and children were "blind," that is they were unaware of the fish consumption and biological measure of exposure for the infant.

When we looked at prenatal PCB exposure, there were only three control variables that were relevant. The first two of these were very weakly correlated with prenatal exposure: mother's age and gravidity. Where the mothers were older, they had accumulated more PCBs in their body, and therefore passed on more to their fetuses. Gravidity is like maternal age: the women who had been pregnant more times were older and had accumulated more PCBs. There was also a correlation with examiner, which we have to assume was due to chance. These three variables were controlled for in all analyses of the effects of prenatal PCB exposure.

With four-year serum PCB levels, there was actually a positive correlation with **socio-economic status** (SES). The higher SES children have higher PCB levels at age four. That makes PCBs very unusual because most risk factors, such as lead and alcohol, in our society are more concentrated in the lower social classes. The higher SES mothers breast fed longer and passed more PCBs to their children postnatally through the breast milk. It was the four-year-old children from the more highly educated mothers, who had breast fed longer, who had the heavier PCB body burdens.

Even though we recruited both fish eaters and non-fish eaters in the sample, I have to emphasize that the study was not designed as a comparison between an exposed group and a control group because everyone in western industrial countries, such as ours, is going to have some levels of PCBs in his or her body. What we were measuring in the studies was prenatal PCB exposure, and you do not have to eat Lake Michigan fish to become exposed to PCBs.

The levels of PCBs in the cord blood were very low, which was not surprising given that PCBs are lipophilic and cord blood is very lean. Unfortunately two-thirds of these samples were below the laboratory's detection limits, which means that we could not get a very reliable assessment of exact quantity on a large proportion of these samples. The effect of being close to the detection limit means that we have an increased risk of a Type II error; that means that there may be some real effects that are difficult to detect because we have not been able to measure the exposure as accurately as might be necessary. And in some cases where we detect effects we may be understating them because of the problems with the reliability of the measure. The levels were considerably higher in the maternal milk and so with the maternal milk PCB level, we have a more reliable assessment. About half of the four-year-old children had detectable serum PCB levels that actually approached those of their mothers. These children were virtually always children who had breast fed, and those who were breast fed over six, 12 months, or some cases 18 months were exposed to quite heavy doses of PCBs. We did an analysis to examine the determinants of the four-year PCB levels. Prenatal exposures were small and, as the child grew, that very small

amount that crossed the placenta became virtually undetectable in the blood. The cord-serum measure was unrelated to the measure of PCB body burden at four years. Instead, it was maternal milk PCB levels and duration of breast feeding which proved to be the principal determinants of the four-year PCB levels.

Turning to the effects on physical growth, we found that both higher cord serum PCB level and consumption of Lake Michigan fish predicted smaller birth weight, smaller head circumference, and reduced gestational age. The relationship was dose dependent. These effects on birth size were consistent with reports from Japan and Taiwan in which children were exposed prenatally to much higher levels of PCBs and related contaminants from maternal consumption of PCB-contaminated rice oil. There have also been studies of occupationally exposed women, working in capacitor plants in the U.S. and Japan, whose infants were reported to be smaller at birth. And there was one general population study in Japan, where female infants were shown to be smaller at birth in more heavily exposed mothers. Although the effects we saw on birth size were statistically significant, we do not think they were clinically significant. The absolute birth size deficits were very small, ranging from 160 to 250 grams, which are similar to those you find with children of mothers who smoke during pregnancy. But the important difference is that infants exposed prenatally by smoking grow faster over the first five or six months and tend to catch up. When we remeasured the children in our study at five months, they were still small, and the prenatal PCB exposure still predicted smaller size at five months. Even at four years, we found a weight deficit, in that the children who were exposed prenatally weighed, on average, 1.8 kilograms less. The evidence of persistent weight deficits is again consistent with the evidence from the Taiwan exposure. In addition, there is a laboratory study of rat pups which showed persistent size deficits associated with prenatal PCB exposure. Similarly, the general population study with female Japanese children, showed persistent weight deficits into childhood. The effect we have seen on physical growth is related only to prenatal exposure. There was no apparent effect on the physical growth of the children exposed to much higher levels of PCBs postnatally by breast feeding.

In terms of cognitive development during infancy, the principal finding was that the more highly exposed infants exhibited poor visual recognition memory at seven months. The test of visual recognition memory was a new test called the Fagan Test, in which the infant is seated on the mother's lap in front of an observation chamber. The observer watches through a peep hole and observes the infant's gaze; whether the infant is looking to the left or to the right. The infant is initially shown two identical photos for 20 seconds to give him/her a chance to encode them in memory. The familiar photo is then paired with a novel one. The normal response is to look longer at the novel photo since the infant has seen the familiar one and will now find it boring. If the infant does look longer at the new one, we can infer that the infant has encoded the initial one in memory, is able to retrieve it from memory, and is able to discriminate between the two photos. Thus, if the infant prefers the new photo, we can infer that there is some very basic aspect of cognitive processing that is intact. Cord-serum PCB and maternal fish consumption levels were both associated with poorer performance on this test. Infants exposed prenatally to higher levels tended not to prefer the new photo. The effect was highly dose dependent. If you look at the highest exposed infants, you can see that their preference for the new picture is at 50%; they show essentially no preference for the new picture. This suggests that there is some aspect of encoding information into short-term memory that they are having trouble with due to prenatal exposure to these contaminants. As with physical growth, postnatal exposure from breast feeding had no effect on cognitive performance in this test.

At age four, the principal test that we used was the McCarthy Scales, which is like an IQ test for preschool children. It is often difficult to elicit cooperation when you are trying to test a four-year-old and, if an uncooperative child gets a low score, you cannot know if that is due to the fact that they are not competent or they are just not in the mood to do the things you are trying to get them to do. We set a criterion to identify the non-cooperative children. Any child who failed to respond to all, or all but one of the items on

any of 17 designated subtests was considered non-cooperative. By this criterion, 7.2% of the children were non-cooperative due to incomplete data, and their data were excluded from the analysis. Looking at the remaining children, what we found was that prenatal PCB exposure was associated with poorer performance on the McCarthy Verbal and Memory Scales, with an effect just short of statistical significance on the Quantitative Scale. The strongest effect was on the memory scale and it was principally on two subtests: verbal memory, which assesses recall for strings of words, sentences and a story, and numerical memory, sometimes called forward and backward digit span, which tests the child's ability to repeat strings of numbers, both in the order dictated and then in reverse order. The effects on both the verbal scale and the memory scale were dose dependent.

We gave another short-term memory test at age four. In this test the child was shown an array of one or three drawings on a computer screen and then asked to remember it. The child was then shown a series of drawings one by one and told to push a button whenever the stimulus on the screen came from the original memory set. Here we found another dose dependent effect of prenatal PCB exposure. The highest exposed children made considerably more errors, giving another indication of some problems in short-term memory processing ability in relation to the prenatal exposure measure.

Given the relatively heavy postnatal exposure from breast feeding, we certainly were anxious to see if there was any relationship between breast feeding exposure and these same outcome measures. When we looked at the McCarthy Scales, we found that there was a negative correlation between maternal milk PCB level and memory performance, for both verbal memory and numerical memory. However, a longer duration of breast feeding was associated with better memory and also better verbal performance. This positive correlation with duration of breast feeding was due to more optimal intellectual stimulation by the mothers who breast fed. Breast feeding was more common in the higher social class, better educated mothers who had higher IQ scores and gave more optimal stimulation as indicated on the HOME Inventory.

One measure of postnatal exposure suggests a deficit; the other measure suggests better performance. To further investigate this, we broke the milk PCB measure down into five levels and breast feeding down into three levels (see Table I). If you look first at the bottom row of the table, you can see that the effect is seen only in the highest exposed children. It is only where the mothers had 1.25 parts per million or more PCBs in their milk that we saw the memory deficit. If you look at the right-hand column, the longer they breast fed the better the children did. The key column is the one labelled 1250-2600. Looking across the rows the children in that column consistently did most poorly. But, as

Table 1 **McCarthy Memory Scale scores (adjusted for potential confounders) by maternal milk PCB level and duration of nursing.**
(from *J. Pediatrics* 1990; 116: 38-45; by permission)

Duration of Nursing (mo.)	Breast Milk PCB Level (ng/ml)					
	185-499	500-749	750-999	1000-1249	1250-2600	Mean
0-3.0	57.9 (2)	49.4 (10)	48.3 (6)	49.1 (4)	42.7 (2)	49.5
3.1-9.0	54.3 (5)	56.4 (10)	51.2 (11)	53.8 (7)	44.6 (3)	52.1
9.1-18.0	52.5 (7)	58.0 (10)	61.7 (8)	55.4 (4)	47.2 (5)	54.9
Mean	54.9	54.6	53.8	52.7	44.8	

you go down the column, even in this highest exposed group, the longer the children breast fed, the better they did. The key here is the fact that the mothers with more PCBs in their breast milk also had more PCBs in their blood and therefore, also transmitted more PCBs to their infants prenatally. So the infants in this column were exposed more, both prenatally and postnatally. The lower scores that we are seeing in this column, are due not to the postnatal exposure that these children are getting, but to the greater prenatal exposure they are getting. The group with a mean score of 47.2 is the group with the highest postnatal exposure. Since they did better than the other groups in that column with less postnatal exposure the deficit seems to be attributable to the fact that all the children in that column got higher prenatal exposure. We conclude that it is not how much contaminated milk the child ingested, but rather how highly exposed the mother was to start with.

Given that the deficit in infancy could have been due to impaired visual discrimination, as well as impaired memory, we also looked at visual discrimination at age four. The test we used is an old test, called the Matching Familiar Figures Test, but we redesigned it to look at both visual discrimination and the speed at which the information is processed. In this test the child was asked to identify which of the two stimuli on the bottom of a picture was identical to the one at the top. There were 24 sets of pictures. We recorded how long it took the child to come up with an answer and then, if the child gave the correct answer, we would ask why the other picture was wrong. These data generated three summary measures: i) how many of the 24 problems the child got right; ii) average time to respond on all the problems; and iii) average time to respond on the problems for which the child got the right answer for the right reason. We took the third of these, our measure of visual discrimination processing speed. On this measure, the strongest effects were with the maternal milk PCB levels. Where the mothers had higher PCB levels in their milk the children responded more slowly, that is, it took them more time to come up with correct answers. The cord blood PCB effect fell short of statistical significance but was in the same direction. Again, duration of breast feeding led to more optimal performance, thus, the effect of maternal milk PCB level appears to be because of the prenatal, rather than postnatal exposure.

We believe the data indicate diminished potential. All of these four-year-olds seem to be performing within the normal range, but the higher exposed children seem not to be doing as well as they otherwise would have in the absence of this exposure.

In conclusion, I would like to emphasize three points about our data. One, the deficits in physical growth and short-term memory that we have found to date were all specifically related to prenatal exposure. Even though much larger quantities of PCBs are transferred postnatally by breast feeding, there appears to be markedly greater vulnerability when the exposure occurs *in utero*. There are several possible mechanisms to explain this phenomenon. We know that migratory cells and cells undergoing mitosis in the prenatal period are particularly sensitive to toxic insult. There is a blood-brain barrier that protects the brain, but it is not formed until shortly before birth. And there are drug metabolizing capacities that do not develop in the prenatal period but will help the infant deal with postnatal toxic exposures.

I would also like to emphasize that the magnitude of the deficits that we saw was modest. There was no evidence of mental retardation or gross impairment, and yet, if you think of Glen Fox's criteria of consistency of the evidence, we were impressed that the effect appears to have been sufficiently robust to disrupt short-term memory in different domains and in different modalities; verbal and quantitative auditory memory on the McCarthy Scale at age four, visual memory for pictures on the Fagan Test in infancy and on the computer test at four years. We believe the data indicate diminished potential. All of these four-year-olds seem to be performing within the normal range, but the higher exposed children seem not to be doing as well as they otherwise would have in the absence of this exposure.

I have been quoted in the press as having said that our findings have no clinical significance. I have said that the physical growth effects that we saw have no apparent clinical significance, but the short-term memory deficits may be quite significant for later cognitive development. Relatively subtle deficits in short-term memory or attention could have a marked impact on the child's ability to master basic reading and arithmetic skills in school. It is possible that subtle deficits in cognitive processing ability could become magnified if the child has trouble acquiring basic skills, becomes labelled as a slow learner, and lags as a result. Alternatively, it is conceivable that in the structure of the school environment the children could outgrow these deficits with increased school experience. That is why we are proceeding on our 11-year followup to try to get a better picture of the longer-term implications for the relatively subtle deficits that we have seen postnatally and at age four.

BASIS FOR REMOVING BIOLOGICALLY ACTIVE PERSISTENT TOXIC SUBSTANCES

Dr. William Owens
Procter and Gamble Company
Cincinnati, Ohio

Ladies and gentlemen, this workshop is devoted to a discussion of how to achieve a weight of the evidence for evaluating environmental risks. This discussion continues to be emotional. For industry, there exists a perception of excessive financial burden; for the environmental community, there exists a perception of hesitation and the burden of inaction. Above all, there is a societal need for sound decisions, not guess work. Today, I will be presenting the results of a three-year environmental study on a Canadian river, where a bleached kraft pulp mill has discharged effluent for the past 18 years. Setting aside the emotions of the past few days, I would like to look at the costs, human resources, time and outcome of this study -- because I believe it exemplifies both the difficulties and the opportunities, gaining broad stakeholder agreement on the meaning of the ecosystem study. It also has aspects of the reverse onus as industry initiated the study, and, as a mill effluent, the assessment of complex mixtures is addressed.

In 1988, after over 15 years of operation of using chlorine gas, the Grande Prairie mill was, like many pulp mills in 1988, faced with the following situation: the 2,3,7,8 congeners of dioxin and furan were present in the mill effluent in parts per quadrillion, and they were also found in fish near the mill in parts per trillion. At that time, the environmental implications for the river ecosystem were unknown.

It was decided to proceed along two paths:

1. The process was to be changed between 1988 and 1992; highly chlorinated organics were to be reduced by removing both chlorine gas and hypochlorite from the mill operation, using increased cooking, pressure diffusers, oxygen and peroxide reinforcement, and 100% chlorine dioxide substitution.

2. A comprehensive environmental assessment of the river system was to be conducted with two objectives: To assess whether the mill's operation was having an adverse effect on the receiving river's biology and to establish a baseline for evaluating future operations.

Analyses of the mill effluent as process changes were implemented showed a steady reduction in the formation of dioxin and furan. The effluent has been non-detect with the change to 100% chlorine dioxide -- with detection limits from 2-5 parts per quadrillion since July 1992. So the process changes achieved their primary goal. Similar non-detects for polychlorinated phenolics have been demonstrated in monthly analysis at detection limits of .01 ppb.

Now let's proceed by reviewing background information on the site and the key concepts of the study design. We were fortunate to have had performed 20 years of benthic macroinvertebrate monitoring; including a preoperational baseline. These data were important to focusing the study effort as we will see in a minute. However, there were data gaps on the environmental transport of effluent compounds and on the health of fish species. Baseline data are often lacking -- and every attempt should be made to find baseline data or estimates of baseline conditions.

The analysis of the total numbers of benthic macroinvertebrates found in specific classes showed that the pollution sensitive E-P-T (Ephemeroptera, Plecoptera, Trichoptera) group is relatively stable at all stations, even below the mill. However, there is a rise in oligochaetes below the municipal discharge and again below the mill. Below Bear Creek, where surface runoff from the city of Grande Prairie enters the river, there are some additional changes. These results are consistent with an organic and nutrient enrichment pattern from several sources and with no evidence for fundamental, adverse impacts -- these data were critical to eliminate the major confounder of eutrophication impacts and allowed us to concentrate on filling in the needed chemistry and the fisheries biology. Addressing confounders is of critical importance in achieving a broadly accepted weight of the evidence upon which regulators will act and the public will accept.

Our overall study design concept was to look for adverse effects and to attempt to correlate any findings with chemical exposure. A major strength of the study is to use multiple parameters to determine both exposure and environmental effects. As you will quickly see, our ecosystem study includes data on the discharge, water and sediments, invertebrate and fish body burdens. We have also tested numerous biomarkers for usefulness, in addition to organismal level and fish population measures in addition to the benthic data. Thus, a comprehensive effort was undertaken to gain a consensus from stakeholders using a weight of the evidence basis.

The mill, located in Grande Prairie, Alberta, is on the Wapiti/Smoky River system. The study area went to the confluence with the Peace River, where the Diashowa mill had just started operation. Previous research indicated that fish species in this environment could be relatively mobile, so the reference area chosen was on the North Saskatchewan River system which does not have a bleached kraft mill and is free of major industrial activity.

There are a variety of important habitat differences in the study area. The Wapiti is fed by snow melt and glaciers in the Rockies -- it falls sharply through the foothills -- which defines one habitat region of the river. Then in the flatter agricultural and forest lands around Grande Prairie and for about 25 kilometers (15.5 miles) below the mill, the Wapiti forms a second habitat region. In the larger Smoky -- which runs nearly 200 kilometers (124 miles) to the Peace -- is a third habitat region, especially in regard to high natural silt loads which affect both benthic and fish populations. Each habitat represents a change which may affect fish populations or biomarkers, and must be recognized for proper interpretation. Sampling sites are 1) above the municipal sewage outfall, 2) between the sewage outfall and the mill, 3) within 5 km (3.1 miles) of the discharge, 4) from 5 km (3.1 miles) downstream of the discharge to the Smoky and 5) sites near Watino and the confluence with the Peace. Two other sites are the fish spawning areas studied: first, the longnose sucker spawns in smaller side streams in the spring. Big Mountain Creek is a confirmed spawning site. Second, the first confirmed mountain whitefish spawning area to be studied -- at Wapiti Gardens -- the mountain whitefish is a fall broadcast spawner.

The ecosystem itself presents a challenge to fish species, in addition to major temperature fluctuations with the ice cover during the winter and the peak flow during summer floods. A flood in 1990 was particularly extreme, but major fluctuations occur annually. This required a multi-seasonal sampling program -- with some focus on the fall low-flow events when exposure would be high -- again trying to anticipate what times were most important to achieving a sound and accepted weight of the evidence for this site.

A variety of parameters were used to document the habitat types and to define river regions and our reference site. Ultimately, the data showed that the reference site fit the upstream portion of Wapiti more closely than the downstream portion; a recognition necessary for proper interpretation.

Chemical analyses were performed during the study with emphasis on chlorinated organics. The abiotic and the biotic compartments tested included the water column, deposited and suspended sediments, benthic invertebrates, and both fish muscle and fish bile. The fish measurements were largely on individuals, not composites -- this is necessary to test for dose correlations between chemical body burden and the biological observations. This was done to eliminate a prime deficiency in many studies: lack of exposure and dose data, usually due to the high analytical costs involved. However, exposure is one of the most vital aspects of field research, as any toxicologist knows, "the dose makes the poison."

On the biological side, various parameters were measured at the population level for the fish community, especially for two target species: the mountain whitefish and the longnose sucker. As we will see, the mountain whitefish have the greatest exposure to potentially bioaccumulating compounds. Further, the population level is quite important, as there is common agreement that adverse effects, when present, can be clearly measured at the population level.

The fish measurements were largely on individuals, not composites -- this is necessary to test for dose correlations between chemical body burden and the biological observations. This was done to eliminate a prime deficiency in many studies: lack of exposure and dose data, usually due to the high analytical costs involved. However, exposure is one of the most vital aspects of field research, as any toxicologist knows, "the dose makes the poison."

At the individual level, various measures were taken for several fish species, but again concentrating particularly on the longnose sucker and the mountain whitefish. Several parameters such as histology are also widely accepted. Measurements of reproductive capacity and success should be central to initial environmental assessments. The biomarker tests employed during the study included a relatively large set of measurements -- again, most often on the longnose sucker and the mountain whitefish -- but various tests were also conducted on other species such as burbot and walleye. In this class of measures, it should be noted that there is far less scientific and regulatory consensus on what constitutes an adverse effect. Therefore, interpretation and use are far less clear.

The results, beginning at the population level and working downward, show representative fisheries abundance data from fishing efforts in 1991. Target species, the longnose sucker and the mountain whitefish, are numerous enough for adequate sample sizes. At the population level, the Wapiti has a diverse fisheries population indicating a lack of adverse effects from the mill effluent. This is consistent with the historic benthic and the habitat data. Obviously, the reference site, consistent with the habitat observations, is more of a mountain whitefish stream, and is a less hospitable habitat for suckers.

A key measure of population health is growth. Therefore, scales or bone structures were taken for aging with weight and length to calculate growth. The mountain whitefish in the Wapiti grows at the same overall rate as the reference stream. Like the population data, this is an accepted indication of no adverse effect from the discharge.

When one compares the lipid adjusted dioxin body burdens across fish species in 1990-91, there is a very startling finding. Longnose sucker, a bottom feeder, and burbot, an omnivore, and walleye, a predator, had comparable TCDD body burden ranges. The mountain whitefish, however, had uniquely elevated levels on a lipid adjusted basis. Simple water column bioaccumulation or sediment contact could not account for the values in this species.

This startling difference in lipid adjusted body burdens was followed with an analyses of dietary food chain niches based on stomach contents. These results lead to a schematic model indicating that suspended sediments from the mill are the apparent transport mechanism for TCDD. Benthic organisms filter-feeding on these suspended sediments are the primary link to consumers higher in the food chain. As consumption of filter feeders is uneven between species, mountain whitefish consumption elevates body levels. This model was a key step in establishing a weight of the evidence -- providing the scientific basis for the most highly exposed species -- and as a salmonid, the mountain whitefish is also presumed to be a very sensitive species. This also provides a general model for hydrophobic compounds from the mill to enter the food web.

A standard template required that the field crew record a complete data set for each individual fish captured, which included the gross pathology field record for the two target species. We found that external and internal parasite loads and parasite types were similar between the Wapiti and the reference. External secondary sexual characteristics were similar between sites for both species during spawning runs, contrasting with adverse findings at Jackfish Bay in eastern Canada. External lesions such as fin rot were rarely found, and were not elevated in the mill population. Net, no gross physical deformities were found in Wapiti fish.

Histology was performed on liver, kidney, spleen and gonads and there was no evidence for major organ pathologies between the Wapiti fish and the reference samples. No evidence for tumors, neoplasia or preneoplasia were found. Occasional local areas of liver damage -- hepatic focal necrosis -- were observed in conjunction with liver parasites, but at similar frequencies between exposed and reference fish. We also observed some bile duct proliferation in the suckers (and not the more heavily exposed mountain whitefish). To resolve the issue, fillet and bile burdens of mill compounds were tested against bile proliferation. There was no apparent relationship with mill exposure. Thus, the individual measures of exposure showed their worth and value.

As noted, a primary focus of our research was fish reproduction. Reproduction is vital to a species and represents a complex biochemistry, susceptible to chemical toxicants; it is a critical endpoint to evaluate for a sound weight of the evidence conclusion. Initial observations were on gonad size -- which is not significantly different in either species for either sex -- and also the age at which fish become sexually mature. Here again, we have no statistically significant differences, but slight trends to earlier maturity at the exposed site, in contrast to findings of delayed maturity at some mill sites in Scandinavia and eastern Canada.

Finally, fish reproductive hormones were measured from blood serum using radioimmunoassays. At several eastern Canadian sites, there has been some evidence the fish near both bleached and unbleached mills have lower hormone levels. We have analyzed for estradiol, testosterone and 17,20-dihydroxyprogesterone, all critical in the control of the reproductive cycle and spawning activity. To date, we see no differences between Wapiti and reference suckers immediately before, during, or after the spawning run. However, an early blizzard and river freeze up prevented the capture of mountain whitefish during an attempt to evaluate their spawning run; so, climate and seasonal events often hamper data collection and may confound interpretation, and all stakeholders have to appreciate the variability and difficulty of working in the field.

Only one of numerous biomarkers showed a consistent difference during the study: an inducible liver detoxification enzyme, EROD. This is one of a large family of P450 en-

zymes which biotransform and metabolize various hydrophobic molecules. Mountain whitefish EROD is highly induced in a spatial relationship to the mill discharge. In contrast, the longnose sucker induction above background was minor.

We have examined the induction both from a chemical exposure relationship and for any correlation to adverse biological effects. There is with the mountain whitefish an association between induction and fillet dioxin levels: as dioxin levels have fallen in fish with the mill process changes, so has the degree of induction. However, no associations have been found with other biological responses such as liver somatic index or with other parameters such as sex steroids. Hence, EROD induction appears to be a marker of exposure and not of adverse biological effects.

This work is being continued to build upon the design and its findings. Dioxin and P4501A trends are being monitored in mountain whitefish in spring and fall samplings every year. Additional sampling has taken place to monitor reproductive cycles with sex steroid analyses. This fall suspended sediment and fish bile samples were taken to assess other aspects of changes in exposure. Samples have been archived for either histology or blood serum analyses if these should become necessary. Again, this is an effort to have both exposure and biological response data for an adequate assessment to derive a weight of the evidence.

This weight of the evidence approach has had very important results for the mill:

- After three years of one-year permit extensions, the mill received a five-year operating permit in 1992.
- The regulators dropped a proposal to require oxygen delignification as there was a sufficient demonstration of a lack of adverse effects.
- Chlorine dioxide use was accepted.
- Discussions have now begun on lifting fish consumption advisories, as we have demonstrated a fall in dioxin body burdens below regulatory levels.

In summary, a weight of the evidence approach is not a simple task. Considerable effort must be spent on solid, thoroughly reviewed study designs to achieve a weight of the evidence. Exposure validation and concurrent measurement of biological responses are part of the comprehensive measurements needed to satisfy stakeholders.

Now let's carefully look at this effort from the standpoint of what were the critical steps:

First, the design and the ongoing results were thoroughly reviewed **before** execution with provincial and federal regulators and subjected to scientific peer review among both U.S. and Canadian scientists. This careful review was critical to gain acceptance of the study results and to ensure that important weight of the evidence factors were not omitted.

Second, a multi-disciplinary team was necessary including analytical chemists, mill personnel, several fisheries biologists and the associated field team, fish endocrinologists, a pathologist, a climate chemist, and a biochemist for liver analyses. Besides scientific services, all team members contributed towards both the design and the data interpretation.

Third, to truly achieve a weight of the evidence, both chemical and biological measures of exposures had to be thoroughly investigated. Without the chemical data on fillet and bile and the biological data on P4501A to provide an exposure assessment, few would have accepted the overall biological conclusions that the effluent was having no adverse impacts on the ecosystem.

Fourth, there was a comprehensive assessment of the biological endpoints with built-in redundancy -- in many cases, parameters were deliberately designed to reiterate other tests -- so that weight of the evidence conclusion on reproduction had the complete support of several measurements.

Finally, note the time -- three years of intensive study -- and the costs -- approaching $3 million -- which are necessary to do a thorough job, just at one site. Plus the fact that there is a continuing monitoring effort at the site. However, these costs are modest in comparison to capital costs associated with further mill process changes, or the social and economic costs associated with a possible mill closure were it based on assumptions that the discharges were dangerous or causing environmental impacts.

In summary, a weight of the evidence approach is not a simple task. Considerable effort must be spent on solid, thoroughly reviewed study designs to achieve a weight of the evidence. Exposure validation and concurrent measurement of biological responses are part of the comprehensive measurements needed to satisfy stakeholders. Finally, execution in field studies is fraught with difficulty. We have encountered floods, blizzards, equipment breakdowns, and other problems. Expectations of time, money and personnel have to be realistic as to what can be achieved in a given time period.

SCIENTIFIC INFERENCE AND THE PRECAUTIONARY PRINCIPLE

Mr. Jack Weinberg and Mr. Joe Thornton
Greenpeace Greenpeace
Chicago, Illinois New York, NY

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 environmental 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.

Science and Policy

Few scientists would claim that science can establish final or ultimate truths. Rather, science is a method and practice for seeking truth through an iterative process of formulating, testing and revising theories and hypotheses. In a scientific setting, a practitioner seeks evidence in order to strengthen or disprove a hypothesis she or he is actively testing. This effort is part of a larger exercise in the construction of a body of human knowledge.

Science and the knowledge it produces should inform public policy. On the other hand, only in highly authoritarian societies do decisionmakers claim that public policy can or should be derived entirely from science; in those societies, these claims serve primarily to mystify and conceal. In democratic societies, we acknowledge that policy incorporates not only scientific inputs but also considerations of ethics, values and opinions, as well as the

interplay of conflicting interests and perspectives.

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.

When good science informs policy, it increases the likelihood of a match between the policy's stated goal and the actual outcome that occurs when the policy is put into practice. Conversely, when policy consistently fails to achieve its stated goals, this calls into question the policy's intellectual and scientific underpinnings.

Current contamination of the Great Lakes suggests a failure in past environmental policy, a failure that was aided and abetted by limitations or failures in the science that informed that policy. The time has come to re-evaluate theories and concepts such as "assimilative capacity" and "safe threshold levels," particularly as applied to toxic substances that persist and/or bioaccumulate in the environment.

It is also time to re-evaluate policymaking methodologies that are based on these conceptions of "assimilable capacity" and "acceptable harm" -- particularly risk assessment and risk/benefit analysis. As currently practised, these exercises never provide a meaningful prediction of real risks or real benefits. The simplified, narrow models used to "quantify" health and environmental threats bear little resemblance to the complex and unpredictable phenomena that occur when chemical mixtures enter integrated natural systems.

Where data is sparse or harms unanticipated, risk assessments are blind; potential injuries that are poorly understood, difficult to quantify, or simply excluded from the model, never appear in the results. A lack of data serves as evidence of safety. On the "cost" side, the availability of alternatives and the broad social and economic benefits of protective action receive inadequate attention.

Where data is sparse or harms unanticipated, risk assessments are blind; potential injuries that are poorly understood, difficult to quantify, or simply excluded from the model, never appear in the results. A lack of data serves as evidence of safety.

Risk assessments are constructed with a set of narrowing assumptions and choices and are thus highly subjective exercises. Their purported objectivity, however, serves to mask the intellectual and political influences that determine those choices. Thus, exercises in risk assessment can become pseudoscientific artifacts that are manipulated to justify predetermined policy decisions. Even when undertaken in good faith, there still appears to be a systematic tendency to understate total risk relative to total benefit. And risk assessment, with its highly technical language and its pretension to purely quantitative decisionmaking, precludes both democratic participation and adequate consideration of non-quantitative ecological, ethical and political issues.

Legal Proof

In the American system of jurisprudence, a very strong weight of evidence is required to convict a person who has been accused of a crime. That person is considered innocent until proven guilty beyond any reasonable doubt and to a moral certainty as determined by a jury of peers.

This high standard is a good example of "a precautionary approach." In the evolution of our democratic society, an important value judgement was made. We decided to place an extremely high value on preventing the incarceration or execution of persons who are innocent. It was well understood that the decision to highly value the protection of the inno-

cent comes at a cost to society: namely, many instances where individuals who perpetrate criminal acts will go free. We decided, however, that the overriding public interest is to protect the law-abiding citizen from civil authority and thereby prevent the abuse of power, corruption, arbitrary action, and even honest judgment errors.

On the other hand, our society uses a different standard of proof in judging, for example, the outcome of a lawsuit involving conflicting interpretations of the implications of a contract between two equal parties. If there is a difference of opinion on the facts, determination is made by a "preponderance of evidence." Neither side has a special burden of proof to overcome. There is parity between the parties and the decision favors the evidence that is most persuasive.

The societal decision to establish a particular standard of proof in some sphere of concern reflects a societal value judgment about that sphere of concern. The standard chosen reflects a judgment about the appropriate way to make decisions that impact that sphere of concern under circumstances when the data is incomplete and there is uncertainty.

When the data base is rich and the level of certainty about cause and effect linkages is high, virtually any standard of proof will yield the same result. The greater the uncertainty, however, the greater is the likelihood that a mistaken inference will occur. Under such conditions, the actual outcome is as likely to be influenced by the standard of proof in use, as by the data and the evidence. A precautionary standard reflects a societal decision to tilt the balance toward mistakes of one type over those of another.

Precautionary Standard

Precautionary standards are a normal part of everyday life. Common sense dictates that there must always be a relationship between the amount of caution to be exercised, the magnitude of potential for harm, and the degree of uncertainty in predicting outcomes.

Consider, for example, a parent attempting to determine how much freedom of action to give a child. The starting point is the potential for significant harm. If the child is balancing on the ledge of a tenth story window, the prudent parent will take preventive action, even before concluding that the child is certain to fall.

Consider, for example, a parent attempting to determine how much freedom of action to give a child. The starting point is the potential for significant harm. If the child is balancing on the ledge of a tenth story window, the prudent parent will take preventive action, even before concluding that the child is certain to fall. If, on the other hand, the child is playing on a similar ledge four feet off the ground, a more relaxed attitude may be appropriate and the exercise might serve as a learning experience for both parent and child.

A loving parent will take action if there is potential for the child to be killed, but can be much more relaxed if the likely danger is a bruise or a scratch. In neither case does the parent want to see the child hurt -- but the potential for significant harm is key to determining the amount of caution and therefore, the appropriate course of action.

Another example was suggested by a friend who teaches medicine. A patient checks into the hospital on Friday night with symptoms of pneumonia. Based on an examination of the symptoms, the physician reaches a professional judgment that there is an 85% chance the disease is pneumonia and a 15% chance that the patient is suffering from Legionnaires' disease.

The physician must now decide which medicine to prescribe. Medicine A is very effective for treating pneumonia but is quite ineffective in treating Legionnaires' disease. If, however, medicine A is prescribed and the correct diagnosis is Legionnaires' disease, by Monday morning the patient will probably be dead.

Medicine B, on the other hand, is fairly effective in treating common pneumonia, but it is not as effective as A. Medicine B, however, also works for Legionnaires' disease, and lacks significant side effects.

Simply weighing the evidence might tell the physician to prescribe medicine A. Eighty-five times out of 100, this would be the right choice. A physician who does so, however, makes an error and will lose the patient 15% of the time. The consequences of a wrong choice are not identical for each outcome. Good medical practice thus requires precautionary decisionmaking.

Weighing evidence in order to decide upon a course of action under circumstances of uncertainty is not a value-neutral exercise. The loving parent does not conclude, "Odds are that the kid won't fall." The prudent physician does not decide, "Statistical considerations favor a diagnosis of pneumonia."

Precaution must be built into the rules of inference. The goal is not to determine which description of the world is most probably correct. The goal, rather, is to make inferences that can inform a course of action that will minimize the likelihood of significant harm. When the harm is large, the uncertainty is great, and our ability to predict the future is limited, we adopt a precautionary standard to judgment and inference.

Reverse Onus

In a criminal law case, as expressed above, a defendant is presumed innocent, the burden of proof is on the state, and the jury is instructed to reach a guilty verdict only if, after weighing all the evidence, it concludes the defendant is guilty beyond a reasonable doubt and to a moral certainty.

By confused logic, North American policymakers have extended these civil liberties from people to chemicals. Without thoughtful consideration, society has taken upon itself the burden to prove that a particular chemical, a class of chemicals or pollution from a particular industrial process harms health or the environment. In the absence of such definitive proof, the rights of chemicals to continue polluting have been protected.

In a criminal law case, a defendant is presumed innocent, the burden of proof is on the state, and the jury is instructed to reach a guilty verdict only if, after weighing all the evidence, it concludes the defendant is guilty beyond a reasonable doubt and to a moral certainty . . . by confused logic, North American policymakers have extended these civil liberties from people to chemicals.

Such a framework tilts the balance of justice in the wrong direction. It is a policy of precaution that favors the interests of synthetic chemical manufacturers and users over and above the interests of public health and the environment. Somehow, society has decided that it prefers to err on the side of pollution and disease rather than to err on the side of a clean environment and health. This principle, however, derives neither from scientific principles nor from some thoughtful consideration of public ethics and morality. It originated at a time when the potential for toxic pollution to harm public health and the environment was still poorly understood. That this policy still continues is testimony to the considerable wealth, power and clout of the chemical manufacturing industry.

The IJC proposes to change this situation with the principles of reverse onus. This means that when applying the weight of evidence approach in deciding when to act, the burden of proof should not be on society but rather, on the producers and users of synthetic chemicals. Such a policy protects society from abuse of power by chemical companies, and also from corruption, arbitrary action, and even honest judgment errors.

The Precautionary Principle

In recent years, there has been growing recognition of the potential size, scope and duration of damage to ecosystems and health that can be caused by the production, use and discharge of synthetic chemicals into the ecosystem. We are learning that:

1. Environmental damage can be widespread and severe before the injury and its complex of causes have been clearly identified;

2. Even after injurious practices are discontinued, environmental damage can persist for long periods and even continue to intensify;

3. The potential for harm is unbounded and can threaten even the integrity of the human species and its ability to reproduce.

As a result of this growing understanding of the significance and unpredictability of the injury that synthetic chemicals may cause to the ecosystem, "precaution" has become a byword of environmental policy. This concept was first introduced into international law in the "Ministerial Declaration of the Second International Conference on the Protection of the North Sea" in 1987.

The Ministers of the Contracting Parties had agreed to address "polluting emissions of substances that are persistent, toxic and liable to bioaccumulate, at source." Their approach, often called the "precautionary principle," states that action should be taken:

> "When there is reason to assume that certain damage or harmful effects on the living resources of the sea are likely to be caused by such substances, even where there is no scientific evidence to prove a causal link between emissions and effects."

Some argue that the IJC's "weight of evidence approach" is weaker than the "precautionary principle." This interpretation is false, however, and in sharp conflict with the IJC's usage. The weight of evidence approach does not simply involve weighing positive against negative or inconclusive evidence according to traditional standards of proof. The Commission, rather, has called precaution the "basic underpinning" of their strategy.

Some argue that the IJC's "weight of evidence approach" is weaker than the "precautionary principle." This interpretation is false, however, and in sharp conflict with the IJC's usage. The weight of evidence approach does not simply involve weighing positive against negative or inconclusive evidence according to traditional standards of proof. The Commission, rather, has called precaution the "basic underpinning" of their strategy. The use of a precautionary context changes both the purpose and the practice of weighing evidence. The issue now being explored is the development of a methodology for weighing evidence in a precautionary framework -- or what might be called "precautionary inference."

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.

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?"

For example, policymakers must rely on scientific evidence to guide decisions concerning which chemicals and/or classes of chemicals should be classified as persistent toxic substances under the terms of the Great Lakes Water Quality Agreement and thus subject to virtual elimination and zero discharge.(1)

The starting points for such an evaluation are the definitions of toxicity and persistence as established by the Great Lakes Water Quality Agreement and the International Joint Commission:

> "Toxic Substance" is defined as "a substance which can cause death, disease, behavioural abnormalities, cancer, genetic mutations, physiological or reproductive malfunctions or physical deformities in any organism or its offspring, or which can become poisonous after concentration in the food chain in combination with other substances."
>
> "Persistence" is defined as a measure of the long-term fate of the substance in any environmental medium.

Based on these definitions, scientific evidence must be considered to determine whether it is plausible to assume that a particular chemical and/or class of chemicals fits this definition and, therefore, should be considered persistent toxic substances under the terms of the Agreement. Since little or no data are available for the majority of the 80,000 synthetic chemicals now in commerce, precautionary inference is necessary to decide which chemicals may reasonably be presumed to be persistent toxic substances. For ethical, practical, engineering and ecological reasons, the IJC has concluded that attempts to regulate chemicals one-by-one are doomed to failure, so the focus for environmental policy has appropriately shifted to classes of chemicals. Thus, one role for precautionary inference is to determine whether members of a given class of chemicals can plausibly be presumed to be persistent toxic substances.

Since little or no data are available for the majority of the 80,000 synthetic chemicals now in commerce, precautionary inference is necessary to decide which chemicals may reasonably be presumed to be persistent toxic substances. For ethical, practical, engineering and ecological reasons, the IJC has concluded that attempts to regulate chemicals one-by-one are doomed to failure . . .

In this effort, consideration should be given to the environmental behavior of the members of the class that have been studied, theoretical understanding of the chemical and physical properties of the class, and the presence, absence and/or function of these and related compounds in nature, if they exist. If this information suggests that it is plausible to presume that members of a chemical class may be persistent toxic substances, the onus can be reversed. Specific exceptions may be made if it can be shown that a given compound is not a persistent toxic substance.

A second issue for precautionary policy is the identification of industrial processes and/or other anthropogenic activities that are the sources of persistent toxic substances that have been identified for elimination.

Many or most chemical-based industrial processes involve the production and release of hundreds or thousands of compounds, the majority of which are unidentified. With such

uncertainty, evidence must be evaluated from a precautionary stance in deciding whether a given process or activity is a plausible source of persistent toxic substances. Relevant evidence includes the properties of those compounds that have been identified in releases from the process, the nature of the feedstocks and the process environment, and the emissions from other processes with similar feedstocks and/or process environments. Again, the burden of proof rests with those who would engage in industrial activities to demonstrate that their processes are not sources of persistent toxic substances.

Finally, society may wish to consider causal linkages between environmental contamination and damage to health and the environment that has already occurred, an issue above and beyond the implementation of the precautionary principle, reverse onus and zero discharge. Again, such an inquiry requires weighing the evidence in a precautionary framework. No epidemiological study can control for all confounding variables, and the webs of cause and effect that connect chemical mixtures to global health effects are too complex to be fully illumined by the tools and models currently available to the health sciences.

Precautionary inference in this field relies on a holistic evaluation of an integrated body of evidence from laboratory experiments, wildlife studies and epidemiological investigations. The focus shifts from whether or not causal relationships have been definitively proven to considering whether a body of direct and/or circumstantial evidence suggests a plausible hypothesis that harm has occurred.

Conclusion

Precautionary inference is a method for evaluating scientific evidence within a precaution-based policy framework. It is a system for considering scientific evidence when a "false negative" judgment would result in significant harm and when there is uncertainty in our predictive ability.

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?"

The burden of proof rests on the producer and/or user of the chemical(s). 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?"

This approach is particularly useful for identifying industrial processes that are likely sources of substances that may cause harm to health and the environment, for prioritizing classes of chemicals for phaseout, and for evaluating causal linkages between existing environmental contamination and health problems in humans and other species.

The emerging evidence on the effects of persistent toxic substances in the Great Lakes and worldwide -- evaluated using precautionary inference -- demonstrates the great harm that occurs when the precautionary principle is not followed in synthetic chemical policy.

WEIGHT OF EVIDENCE VERSUS PROOF OF CAUSATION

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

My task is to try to pull some ideas on weight of evidence together. I would say in starting that if you think about the fish and pollution presentation (p. 15-20), which was very good, you begin to realize what it would take to simulate such an expensive high-tech study for people. Yet, we expect just that of the ordinary citizen whose only information source is vital statistics, the first cause of death. A person's whole medical history is telescoped into first cause of death. Did they die of an accident, or did they die of pneumonia or did they die of cancer? The mobility of our population poses even more problems yet we expect the local people to raise an alarm: "Say, there is a pollution problem," and then government and industry can fund the high-tech expensive studies to dispute or elucidate the concern. Moreover, the blame for any possible mistakes or false calls falls on the citizen or the scientist who may have alarmed somebody. It is this disproportionate situation and the fact that little or no resources are routinely put into looking at human health in communities at risk even though the stakes are so high that has brought us to today's discussion. Underlying the weight of evidence debate is the burden of proof or reverse onus debate, and disparity in resources.

I would like to consider first the talk by Glen Fox (p. 2-5), and the Hill criteria he used for causality. Much of what is now attempted in toxicology and environmental health is an imitation of the success with infectious disease. The human body posed a similar problem when the germs were discovered, when we started recognizing bacteria and viruses. The human body is full of many different micro-organisms. When someone gets scarlet fever or polio, how does one pick out which one of those bacteria or viruses are really causing the disease? Infectious disease studies set up certain criteria whereby a researcher can identify which one was the culprit. In some ways, it was easier than identifying toxics. There was not an industry out there saying "Well, it wasn't my bacteria, somebody else put that bacteria in there." What I am implying is that it was a less political struggle than toxicology. The infectious disease criteria proved to be useful and a lot of things were put into our medical system to facilitate research. What helped was not only the kind of research that went on, but also such routine things as pathological examination of tissue removed in surgery. One could take the tissue and look at the pathology or the pathogens present. We don't now do this for toxicology. There is no routine toxicological study of tissue, although millions of tissues are being removed from human bodies every day in North America. We are not even progressed to the point of a support system to help in sorting out which toxicants are causing which diseases. Every question raised requires an expensive special collection of data for an epidemiological survey.

Now with this background, Sir Austin Hill, between 1965 and 1967 came out with the list of criteria that is used by epidemiologists for determining causality. It includes various associated research studies such as demonstrating the effect of the toxicant in animal studies. There is a list of requirements for demonstrating causality and they are fairly good. Subsequently, during the 1970s in the U.S., a committee was pulled together of about 200 people to research the questions we should be asking if we want to understand environmental health problems. It is again a matter of establishing the baseline, the "normal." How do

you know when the community health is different, i.e. "not normal"? What do you compare with? This research was undertaken during the 1970s and completed in about 1978. It was supposed to become part of the 1980 U.S. census. It would have been a sub-study and it would have been administered randomly, covering the whole United States and providing a common basis for comparison. We would have had some way to compare regionally and it would have been very useful. No one really knew about this report except for the 200 people working on it. It was voted down in Congress and the booklet, which is still good and useful, is totally out of print. I have a copy and if anyone wants it, I will photocopy it. It is useful, it is helpful, and we need to start asking some of these questions routinely. Systematically collecting data would be one way of applying Hill's criteria to the complicated reality of the 1990s.

However, I think we also have to remember that we need to improve on Hill's criteria of causality. It was a first cut. It is not the last answer and I think that, given our experience of the last 20-30 years, we need to add some criteria to it.

One of the Hill's criteria for causality refers to statistical significance of the finding. Here you are trying to prevent rejecting the null hypothesis when it is true, called a type I error. The null hypothesis is that there is no connection between the toxicant and the illness. Scientists protect the null hypothesis at a 5% level or a 1% level. That means one accepts the null hypothesis unless the outcome was so unusual that it could not have happened by chance more than 5% or 1% of the time. We need to expand Hill's criteria and note the power of the test. The power of the test measures the type II 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 II error.

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 II 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. One of the most obvious examples of this is looking for dose-response with respect to cancer deaths in an area where you have low socio-economic status or a third world situation, where the person is more than likely to die during the pre-cancerous, infectious disease phase than of cancer. 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. There are, for example, exposures which affect the thyroid gland. A fetal thyroid gland develops around the fifth month, so you find a difference in fetal exposure before the fifth month and after the fifth month. The same is true with any other organ system that is forming. So timing in the life cycle is important.

Sometimes the toxic effect is in the offspring of the exposed person. I think we are becoming more and more aware of the effect which Einstein, who was one of the most forthright proponents of nuclear technology, pointed out and that is the subtle intergenerational loss of intelligence in the community exposed to radiochemical pollution. If we start damaging brains, we are going to have reduction in IQ, general reduction in population intelligence, and that moves me to what the famous geneticist, Muller, pointed out: namely, the loss of vigour in the species. When the species starts losing vigour, you are on a species death path or route. We have to pay more attention, not only to the long-term effect in the individual, but to the long-term effects on the species. I think that as we move into more and more subtle damage to the living system, it is going to be the intergenerational effects that will become prominent.

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. I would just point out here that once you start an intergenerational loss of vigour, you are in an irreversible pattern. The same thing relates to our fixation with looking at cancer death, which is certainly a severe end-point. However, if there is excess cancer death it means that you have been doing the wrong thing for some 30-40 years, and the process at that point is irreversible. 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.

Sometimes the toxic effect is in the offspring of the exposed person. I think we are becoming more and more aware of the effect which Einstein, who was one of the most forthright proponents of nuclear technology, pointed out and that is the subtle intergenerational loss of intelligence in the community exposed to radiochemical pollution. If we start damaging brains, we are going to have reduction in IQ, general reduction in population intelligence, and that moves me to what the famous geneticist, Muller, pointed out: namely, the loss of vigour in the species. When the species starts losing vigour, you are on a species death path or route.

Another problem with the Hill criteria, given our present level of pollution, is that it basically assumes that you have a normal healthy population with which to begin. They are exposed to something toxic and there is an ill effect. We have been exposed to a growing number of toxic radionuclides and toxic chemical materials at an escalating rate for the last 40-50 years and I think we have developed highly susceptible sub-populations. I am thinking of some of the multiple chemically-sensitive people. There are also other problems in our society which demonstrate a worsening of the host response. Whenever you have a hazard, you have pathways to people and then you have the response of the person.

The responses of people have also changed. One can't just look at the hazard and the pathways and think of the responses as automatic. The population is not homogeneous. I am thinking of some of the problems which bother me very much, like AIDS. AIDS is a virus which, by its evolutionary composition, is rather an old virus. It has been around a long time. Formerly it was observed as a terminal disease in those over 70. What is different in our time is that it is showing up in people in the 20s and 30s and that was not seen before. But what it says to me is that something has changed in the host response. I think there are other examples of this and we need to look more closely at host response variants.

I think there are ethical questions underlying decisions with respect to pollution. What to do about these problems is not yet clear and sometimes you have to make a judgment call. I would see some differences, for example, where the citizens have a choice. If you have a choice of avoiding a hazard it seems to me different from a hazard which is in your air and which you really can't refuse to breathe. You might not be able to move your residence. There is an element you have to look at when making judgments and that is the individual's ability to avoid the exposure.

I would like to make some suggestions for further reflection. I really think we have to flesh out more clearly what we mean by a weight of evidence approach: how broad it should be; and what it needs to consider. I have recommendations in three categories to be studied. One is with respect to the hazards. The second one is with respect to the pathways, and the third is with respect to the host response that I think could form the basis of a new approach.

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 think we can also recognize science advocacy as legitimate. Scientists are always trying to say that they are purely objective, but it is not really true. It is impossible to avoid choices such as what to research, how to design a study, what related research is "credible," etc. I think we should be more honest and forthright. I would recommend two ways of dealing with this: one would be some type of a science court where there could be at least a clarification of the issues. I also participated in a good system the Germans thought up when they were trying to deal with the Kalkar breeder reactor which was on the border between Holland and Germany. If there was an accident it would be an international affair. They were trying to make an estimate of the extent of nine accident scenarios. What they did is put out calls for a grant proposal for estimating the health effects of these nine accidents and they gave out two contracts, one to people who were proponents of the reactor, and one to people who were opponents of the reactor. Both groups were given exactly the same baseline data, they were given access to the same computer programs and software, and they were told to come up with the estimates of the number of health effects for each of these nine accidents. The study was mandated by the Bundestag. It was an excellent process and clarified a lot. The predicted numbers of casualties ended up different, but we could explain exactly why they were different, where the decisions had been made, what things were scientific and what estimates were judgment calls. I think more of that type of assessment would help.

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. In other words, if only 20% of the cases are connected with an exposure, then your massive program to reduce that exposure can at best give you 20% improvement in a health statistic. It tells you where to put your public health effort, for one thing. Attributable proportion can be estimated now in stratified samples, for example age specific. It is quite a sophisticated technology which is available to us and which I think we should start using.

I would also move into such things as proportional compensation. This addresses some of the legal issues. Compensation for injury for workers, or a law suit for the public, is usually all or nothing. You win or lose in this situation. I think we could begin to deal with it in a much more sensible fashion if we used attributable proportion. If we said 20% of the cases are due to this exposure, then 20% of all health cost for this illness would be covered. There will be resistance to this on the part of the public, but I think we need to move out of the deadlock situation and find new ways of dealing with compensation. Lifestyle choices could enter into the funding of medical care. If 17% of the lung cancers were due to smoking, and you chose to smoke, maybe you should pick up 17% of your health related costs. There are possibilities here. I am not saying those are perfect answers but I am trying to open up a future where we can dialogue and we can find a better way to deal with the problems than in the past.

I have found it particularly hard to deal in the legal framework where basically you have to double the incidence of disease to meet the legal standards of probable cause. You can say it is more probable that the disease is caused by the exposure, than that it was caused by something else if the disease rate is more than doubled. To fulfill this requirement in law, an industry has to suddenly double the occurrence of some disease through its pollution or there is no compensation. That is an irrational kind of criteria and puts a scientist in a terrible position. It also implies you can keep increasing gradually the levels, say of cancer or birth defects, and would never be legally responsible for causing the problems. It is a difficult area. We certainly need an interdisciplinary approach and we need some creative ideas on how to handle decisions. I would look forward to working with people over the next two years and try to get some very clear criteria for decisionmaking.

With respect to the second area, namely *pathways*, we need to investigate biochemical changes after the pollutant is released. For example, cobalt 60 was ignored as a milk contaminant during the nuclear fallout period because the body has a short residency period for inorganic cobalt. In the field, however, inorganic cobalt was incorporated into Vitamin B12 in the cow's rumen. This has a much longer residency period in the body and is stored in liver. Incorporation into the food chain may also be a slow process. The United Nations estimates that carbon 14 will have its maximum public health impact 150 years after release to the environment. Sometimes the toxic material is not released but its precursor is released.

I have found it particularly hard to deal in the legal framework where basically you have to double the incidence of disease to meet the legal standards of probable cause. You can say it is more probable that the disease is caused by the exposure, than that it was caused by something else if the disease rate is more than doubled. To fulfill this requirement in law, an industry has to suddenly double the occurrence of some disease through its pollution or there is no compensation. That is an irrational kind of criteria and puts a scientist in a terrible position.

The third area which needs broadening in the weight of evidence approach has to do with *host response*. This might include past health history of a community, other toxic exposures, differential protection for pregnant women or persons with multiple chemical sensitivities. Protection may relate to age, sex, life cycle, occupation, ethnic background or other pertinent factors.

Hopefully these complex issues, which are of serious import, can be discussed in an open and constructive dialogue involving industry, scientists, government, human rights proponents, ethicists and the interested public.

DISCUSSION

Daniel Green: In about a year Dr. Needleman will be testifying in a United States court in a product liability suit involving exposures of plaintiffs to lead. The evidence concerning PCB discharges to the Great Lakes could result in similar product liability suits against Monsanto as the company that produced PCBs in North America. If you, Dr. Jacobson, were asked by a defense lawyer whether prenatal exposure to PCBs of a particular child in your study had caused a diminishment of intellectual potential, would you answer "yes" or "no"?

Joseph Jacobson: We are in a different position from Dr. Needleman with his findings on lead. The lead literature contains multiple studies that confirm Dr. Needleman's studies.

Unidentified: How did the Commission use the concept of weight of evidence, pertaining to an individual chemical, to lead to the decision to recommend sunsetting a class of chemicals?

Chairman Durnil: I think that the recommendation had more to do with the commitments that the two governments made on virtual elimination and their inability to achieve that end through regulation alone. We believed that ratcheting down the allowable discharges, for example of the 11 critical substances listed by the Water Quality Board, would never get you to virtual elimination and that for these unnatural compounds there has to be zero human input.

Unidentified: How long does it take for the actions to stop the emissions of dioxins to show up as changes in the environment?

Chairman Durnil: I think that the recommendation had more to do with the commitments that the two governments made on virtual elimination and their inability to achieve that end through regulation alone. We believed that ratcheting down the allowable discharges, for example of the 11 critical substances listed by the Water Quality Board, would never get you to virtual elimination and that for these unnatural compounds there has to be zero human input.

Willie Owens: In terms of the levels of dioxins in fish, dioxins have a half-life of about six months. As we started to ratchet down the effluent we would start to see changes the following spring and they would continue to decline. Pulp and paper effluents are extremely complex mixtures of chemicals and the first question to ask is whether there were any problems with the new process or the product. Toxicity is a biological response to something, so that is why I personally advocate that both Environment Canada and the U.S. Environmental Protection Agency undertake assessment of the health of organisms on a regional basis, both for point and nonpoint sources in watersheds. In the past it has taken 15-20 years to determine that something has gone wrong. But it takes resources and well developed measurement systems to move forward as a society.

Unidentified: The International Joint Commission and several other bodies in the U.S. are calling for a definition of weight of evidence. Industry has said that it is committed to being a part of the process of defining weight of evidence and how it should be used. What should we recommend about how to proceed?

Chairman Durnil: The Commission does not have the resources to do that well enough but we have made it one of our priorities for the next two-year cycle, and based on these comments we can think of how to proceed. Obviously if industry is not a part of it, it will not work.

Saulius Simoliunas: Professor Berger said that now the court will recognize that there are two proofs, one scientific proof, one legal proof. I hope that there will be some Supreme Court judge to explain that, because to me it does not make too much sense.

Wayne Schmidt: I work for the National Wildlife Federation and we are one of the groups that place great reliance on Joe and Sandra Jacobson's research because of its importance in the policy-making arena. What is your reaction to the criticism of your research, particularly among the health care professionals?

Joseph Jacobson: Well, I am perplexed by some of the criticism. A lot of the criticism focusses on our alleged failure to control for things that we did control for, particularly maternal drinking during pregnancy. My own caution about our findings has to do with the magnitude of the deficits which are modest, and the longer-term implications of the deficits are as yet unknown. What we have is a clear preliminary indication, from a single prospective study, that there is some damage, but the extent of the damage and the practical significance of that damage are unfortunately not all that clear as yet.

Rosalie Bertell: These things take a long time. From the experience of 30 years in public health, I can say that everything is shaded and controversial, and you are going against conventional wisdom. For example, by the 1930s the relationship between smoking and lung cancer were relatively definitive. In 1919, insurance companies would not insure asbestos workers. One hundred years after the uranium mining disaster in Czechoslovakia, it was repeated in the southwest United States. Public health consistently resists the flow of the evidence.

Glen Fox: More than 20 years have elapsed since the time that we knew that we had this serious toxicological problem in the Great Lakes. Public health should be about taking the proverbial handle off the pump and stopping the cholera outbreak. There will be costs, there will be risks and possibly mistakes and dollars to industry, but we are talking about the health of future generations and about the ecosphere. For many substances we cannot do the kinds of studies being advocated by Willie Owens.

There is an evolution of knowledge. When PCBs were first produced in the 1930s, no one knew about persistence or bioaccumulation. It was only in the mid-1960s and the early 1970s, when Jensen was seeing these chemicals in the seals in the Baltic, that people understood the implications of this select set of compounds. If you do not have a mechanism, you only have an observation and you are not much further forward.

Willie Owens: There is an evolution of knowledge. When PCBs were first produced in the 1930s, no one knew about persistence or bioaccumulation. It was only in the mid-1960s and the early 1970s, when Jensen was seeing these chemicals in the seals in the Baltic, that people understood the implications of this select set of compounds. If you do not have a mechanism, you only have an observation and you are not much further forward.

Glen Fox: I find that rather scary. For substances such as PCBs we now have a fairly complete picture which includes biological mechanisms that make biological sense. But this took a very long period of time and we still have not got PCBs under control. What will happen when the next kind of lesion or syndrome occurs? Will we be able to respond any quicker to investigate it or to control the substance that caused it?

Willie Owens: That is where I come back to biological monitoring of our ecosystems. What was needed at the time was a system network that could indicate whether or not the waters and bird colonies were all right. There seemed to be insufficient resources or an information network to build on the initial observations in colonies in the Toronto and

Hamilton area. There was an insufficient mass of evidence to get peoples' attention.

Jim Macaulay: Could Dr. Jacobson not find a group that was not exposed to PCBs?

Joseph Jacobson: We started with the premise that everyone is exposed and so the design of our research was to investigate whether the more highly exposed infants or children consistently performed more poorly.

Unidentified: If PCBs are only an indicator of other exposures, this has a lot of policy implications, for instance if it is dioxins. Do animal studies indicate that there are a lot of compounds that may be responsible for the kinds of things that you are seeing?

Joseph Jacobson: Research on animals exposed to PCBs and specific PCB congeners has shown the same kinds of behavioral effects that we have seen in our cohort, but there is very little work on other compounds in this regard.

Unidentified: At the beginning of this long and interesting discussion, somebody used the word ethics, before we skated off into some other fascinating material.

Ann Mahan: It is important to monitor the ecosystem, but we cannot keep putting things into the ecosystem and then monitoring to find out what is happening. Using reverse onus, we need to assume that it can cause harm until we know that it does not.

Unidentified: That is what has been happening to the farm workers in California. When one pesticide is finally forced off the market because of the injury to the health of farm workers, a new one is substituted. I am also concerned about the ethics of researchers, educators and funding organizations. For example, there are scientists who apply for grants using the jargon of the funding agency, but instead undertake the studies that are of interest to them. Politics are used to influence what gets funded and how the information is released and used, and there are examples of intellectual dishonesty where statistical data are massaged to obtain the politically correct answer for publication in an Ivy League journal.

What will happen when the next kind of lesion or syndrome occurs? Will we be able to respond any quicker to investigate it or to control the substance that caused it?

• • •

That is where I come back to biological monitoring of our ecosystems. What was needed at the time was a system network that could indicate whether or not the waters and bird colonies were all right. There seemed to be insufficient resources or an information network to build on the initial observations in colonies in the Toronto and Hamilton area. There was an insufficient mass of evidence to get peoples' attention.

Karey Shinn: As a member of the public, I want to know how do officials make practical preventive decisions about public health under conditions of crisis management? Most information is unusable under these conditions since it is frequently designed only to be usable a few generations from now. In practice, a mayor or a schoolteacher, who may not be a scientist, makes decisions with far-reaching consequences based on very little information but based on the opinions of those who are available.

Robert Schubring: What was the basis for the Commission's decision to advocate a sunset on all chlorine manufacture? Was it based on the fact that there were quantities of DDT and certain other chlorinated pesticides in the Great Lakes for which the Commission had a mandate under the Treaty to achieve virtual elimination from the Great Lakes?

Gordon Durnil: On June 7, 1990, the Commission set out a series of priorities that included an examination of the terminology of the policy contained in the Great Lakes Water Quality Agreement concerning virtual elimination of persistent toxic substances. This

has been the policy of the Canadian and United States Governments since the signing of the revised Agreement in 1978. Through the Virtual Elimination Task Force and a series of roundtable discussions involving industry, environmental groups, and scientists and regulators, we arrived at the conclusion that the policy was unattainable through regulation alone. We recommended that for those substances that were so onerus that society cannot tolerate them, there must be some date, whether it is five years, 10 years, or even 50 years, when the substances will no longer be brought into existence. We reviewed the list of 11 critical pollutants set out by the Water Quality Board. The majority of them are chlorinated organics, which then raised the question of how do you deal with chlorinated organics, where the evidence indicates they are harmful, without dealing with chlorine itself.

Robert Schubring: Chlorine is essential to the manufacture of items critical for national defense, such as silicone for micro chips and titanium for aircraft turbine blades, and for the manufacture of platinum catalytic converters for air pollution control on automobiles. Our concern is how do you get from 11 toxic substances in the Great Lakes, that your body has a mandate to oversee, to something that has absolutely no relevance to that whatsoever?

Gordon Durnil: That is your conclusion, not mine. We recommended that industry had to be involved in setting a timetable, so that there was no social or economic disruption.

John Mahan: A large part of society believes that if we can get enough science and technology we can solve the problem. Science is a system of inquiry. It is not a system of answers or of decisionmaking. No matter how much science we have, there is always more science we will want and need and we will never have all the answers, but decisionmaking comes through judgment, wisdom and ethics. Science is a tool, not a solution. And so we need to use the best science we can, but we've got to go beyond that and be guided by ethics. That takes us to reverse onus and the precautionary principle.

A large part of society believes that if we can get enough science and technology we can solve the problem. Science is a system of inquiry. It is not a system of answers or of decisionmaking. No matter how much science we have, there is always more science we will want and need and we will never have all the answers, but decisionmaking comes through judgment, wisdom and ethics. Science is a tool, not a solution. And so we need to use the best science we can, but we've got to go beyond that and be guided by ethics.

• • •

It seems that there are no easy answers in these complex issues, since it is difficult to make policy decisions to protect an ecosystem or a sensitive species when there will be effects on the socioeconomic structure of our society.

Thomas Hoerman: I am an employee of the BASF Corporation. I seem to recall that "situational ethics" is a principle whereby you look for the greatest good for the greatest number. It seems that there are no easy answers in these complex issues, since it is difficult to make policy decisions to protect an ecosystem or a sensitive species when there will be effects on the socioeconomic structure of our society.

Valerie Denney: One practical point of view relating to the precautionary principle is that there is not enough money, either in industry or in government, to finance all the studies that all of us would like to see done to ensure a high degree of certainty about a lot of these chemicals. There are just too many that have been accumulating for too long and that may have synergistic effects. We need both to be cost effective and to protect public health. In a time that the public is overtaxed and resisting spending more money on regulation, the precautionary principle must be combined with a transitional program that meets the needs of workers affected by these decisions.

Ann Jarrell: I think that we need to focus on how policy makers make decisions since this is not studied enough. As a scientist working for the Health Standards Division of the Occupational Safety and Health Association, I developed a scientific record which was reviewed, but the decision was taken out of my hands and made at the political level. Whatever exposure level was set would not necessarily coincide with my recommendation.

Gordon Durnil: The scientific community often forgets that communication from one level to the next level is a critical element. Whether scientists are trying to communicate to a congressman or member of parliament or a CEO, there needs to be clear communication even though this can be a very difficult thing to do effectively with technically complex material.

Unidentified: We need a way of determining the economic feasibility of many of these chemicals and of their alternatives. Companies benefit from the products, but it is the citizens who must find the resources to fight these chemicals. I prefer a reverse onus model in which the 15,000 organochlorine chemicals would be banned and, if a company wants one particular organochlorine substance, let the company undertake the studies to prove the safety of the chemical. When the results are completed, they should be made available to citizens who may wish to argue about the safety in an open forum.

Glen Fox: I am concerned that the economists are not developing the science of economics to deal with environmental issues, or if they have, it is not widely used. As a regulator, I have sometimes come to the conclusion that a product was ecologically dangerous. But in preparing a risk-benefit or a cost-benefit analysis, the environment always ends up looking as though it has no value.

Rosalie Bertell: In the preparation of the Ontario Hydro 25-year plan, we investigated the human health costs of each of the different ways of producing electricity. In essence, because the province ends up paying all the health costs, these do not enter into either the assessment hearing or to the decisionmaking. We prepared a six-volume submission that included estimates of the externalized costs to the province or to society.

We need a way of determining the economic feasibility of many of these chemicals and of their alternatives. Companies benefit from the products, but it is the citizens who must find the resources to fight these chemicals. I prefer a reverse onus model in which the 15,000 organochlorine chemicals would be banned and, if a company wants one particular organochlorine substance, let the company undertake the studies to prove the safety of the chemical.

Gordon Durnil: Yesterday, David Crombie commented on changing our ways of thinking and the interdependency of economic health with environmental health. For example, when we talk about relative risk, are we going to accept that it does not apply to minorities who need a free source of food such as the catfish from the Detroit River? We watch the dissatisfaction of voters in our two countries primarily electing people they do not want, because they do not want who they have.

Unidentified: There seems to be an analogy between how some people are supersensitive to chemicals because of the general degradation of our general health, and the virulence of zebra mussels and other exotic species in already weakened ecosystems. I should like to see more emphasis on the teaching of the scientific method. This would produce more independent thinkers willing to explore alternative hypotheses and might inspire better solutions from less authoritarian types of personality structures.

Jack Weinberg: As we approach the next millenium, humankind is facing issues that we never faced before because, in the past 50 years, we have obtained the capacity to disrupt ecosystems on a global scale rather than, as previously, on a local or regional scale. The conservation ethic has helped us to start putting a value on species and on the value of natural beauty. But if we do not note what is happening to nature as a result of human

action and callously place no value on what we are losing, then we are jeopardizing our own survival as a species.

Gordon Durnil: In the next two years, the Commission will be wrestling with the subject of weight of evidence as one of our priorities. I would like Brad Leinhart, Jack Weinberg, Rosalie Bertell, and Glen Fox to write what they think we should be doing on this topic in as productive a way as possible, and send it to Mike Gilbertson at the Regional Office. I want to thank you all for coming to this workshop. It has been most enlightening to me and I really appreciate it.