Pure Appl. Chem., Vol. 75, Nos. 11–12, pp. 2521–2529, 2003.

Workshop 5.3

Why epidemiology of endocrine disruptors warrants the precautionary principle*

Michael Gochfeld‡

Environmental and Community Medicine, UMDNJ-Robert Wood Johnson Medical School and Environmental and Occupational Health Sciences Institute and Consortium for Risk Evaluation with Stakeholder Participation, 170 Frelinghuysen Road, Piscataway, NJ 08854, USA

Abstract: The precautionary principle is controversial, and critics invoke the need to wait for "sound science" before taking "costly" regulatory action. For human health effects, epidemiologic results are often considered more valuable than toxicologic studies in animals. Direct human evidence on the endocrine effects of environmental chemicals has been slow to accumulate because of inherent sample size limitations of exposed populations and overconservative hypothesis testing approaches. Moreover, human health outcomes may take decades to emerge. Indeed, even huge population-based studies of hormone replacement therapy have been inconclusive regarding both benefits and risks. This paper argues that certain intrinsic standard epidemiologic methods are stacked to avoid making a type I error. Moreover, these combine with extrinsic limitations (long latency, high cost), leading me to conclude that reliance solely on epidemiology to provide definitive answers, will almost inevitably delay the discovery of meaningful associations warranting timely action for protection of public health. There are several ways in which the inherent conservatism of epidemiology is arrayed against preventative regulatory steps, hence a precautionary approach is warranted while awaiting the results of additional studies which for various reasons may be very long delayed or even impossible.

INTRODUCTION

Although formally established in the literature only during the 1990s, the precautionary principle is firmly entrenched in society, dating back to antiquity. Proverbs such as "better safe than sorry" and " an ounce of prevention is worth a pound of cure" are articulations of the precautionary principle. The precautionary principle emerged mainly in the context of new technologies such as bovine growth hormone and genetically modified organisms, but can also be extended to regulation of environmental pollutants including endocrine active substances. Applied to new technologies with uncertain adverse outcomes, it behooves those who will profit from a proposed technology to demonstrate conclusively its safety or that its benefits will outweigh its harm. In its most frequent incarnation, it means that the existence of uncertainties, unresolved facts, or the need for more research shall not be invoked to delay controls or regulatory action, particularly when the consequences of inaction or delay are serious or irreversible.

*Report from a SCOPE/IUPAC project: Implication of Endocrine Active Substances for Human and Wildlife (J. Miyamoto and J. Burger, editors). Other reports are published in this issue, *Pure Appl. Chem.* **75**, 1617–2615 (2003).

‡Tel.: 732 445 0123, ext. 627; Fax: 732-445-0130; E-mail: <gochfeld@eohsi.rutgers.edu>

The most frequent definition (Rio definition of 1992) is "Nations shall use the precautionary approach to protect the environment…. Where there are threats of serious or irreversible damage, scientific uncertainty shall not be used to postpone cost-effective measures to prevent environmental degradation." Or in simplest terms a potential hazard is not innocent until proven guilty. It is commonly stated that the precautionary principle has been adopted in Europe but abjured in the United States. Actually, a number of U.S. environmental statutes embody precautionary language. For example, the Toxic Substances Control Act (1977): "among the many chemical substances and mixtures which are constantly being developed and produced, there are some whose manufacture, processing, distribution in commerce, use, or disposal may present an unreasonable risk of injury to health or the environment…"

Agencies responsible for regulating environmental pollution on both the national and international scale set emissions standards or contaminant levels based, in part, on the results of scientific studies published in the peer-reviewed literature. Yet, even science-based standard setting is often affected by economic and technical considerations. Most agencies, the general public, and particularly scientists, like to think that standards are based primarily on science and that as new scientific information becomes available, standards will be revised. This, however, is not always (or even often) the case. In the United States, for example, the Occupational Safety and Health Standards are mostly based on pre-1970s science—attempts to update them having been stayed in court. Moreover, for some kinds of contaminants, particularly where effects are subtle or occur at very low levels, or where funding is not directed, scientific data are slow to accumulate.

Therefore, the precautionary principle is increasingly invoked to take regulatory action in the absence of incontrovertible scientific data demonstrating harm. The precautionary principle is considered controversial, and indeed those who oppose or caution about its application usually refer to it in the extreme sense of complete banning of a new technology or a toxic material. However, in practice precaution represents a spectrum of actions ranging from those that can be clearly and confidently based on adequate and sound data to interim actions taken while additional data are pending. Moreover, precautionary actions are themselves not irreversible. Taking preliminary remediation steps or delaying the introduction of new technologies may incur costs, but can be reversed when information on safety or risk become available. To extricate "precaution" from controversy, advocates now refer to "precautionary approaches" rather than a "principle".

Ideally, environmental health policy ought to reflect scientific consensus and risk assessment and management. Risk assessment is never completely free of uncertainties, and often it is necessary to articulate policies when science and risk assessment have yet to provide clear answers. I view risk assessment and precaution as complementary, rather than opposite approaches. There are many environmental health issues where uncertainties abound, and the role of endocrine active substances (EASs) and their consequences are fraught with uncertainties. The consequences of uncontrolled exposure can be significant, at least to individuals, yet the difficulty of demonstrating or validating connections between chemical and consequences, remains challenging, particularly in epidemiologic studies of humans. Since Colburn et al. [1] popularized the issue of "endocrine disruptors", opponents have argued that the much higher dose effects of intrinsic hormones such as estradiol (E2) or pharmacologic hormones, would mask or vitiate any possible effects of the environmental estrogens. This very point, however, strengthens the argument for a precautionary approach. The controversy over epidemiologic studies of hormone replacement therapy—where studies have involved many thousands of women—highlight the difficulty of arriving at scientific consensus. The main issue involves resolving the uncertain benefits vs. the uncertain harms [2,3].

EPIDEMIOLOGIC STUDIES OF ENVIRONMENTAL EASs

Although there are abundant epidemiologic studies of endogenous hormones and pharmacologic use of hormones (e.g., diethylstilbestrol and hormone replacement therapy), epidemiologic studies of endocrine affects in humans from environmental contaminants are relatively limited [4]. The organochlo-

rine pesticides were probably the first environmental EASs recognized through their effects on avian reproduction. In humans, exposure to the pesticide DDT appeared to impair lactation in both U.S. and Mexican women [5]. Exposures to polychlorinated diphenyls and dibenzofurans in the Yu-Cheng (Taiwan 1978–79) disaster, was associated with a variety of developmental deficits including shortened penises [6], a finding reminiscent of Florida alligators exposed to organochlorines [7]. Bertazzi et al. [8] reported a low rate of breast and endometrial cancer in women exposed to dioxin after the Seveso (Italy) explosion, raising the question of an antiestrogenic effect of dioxin. The putative thyroid effects following an accidental exposure to polybrominated biphenyl compounds proved elusive [9]. Recently, Goh et al. [10] reported a complex relationship between occupational exposure to trichloroethylene and insulin levels, with no pattern for steroid levels. Probably the greatest recent controversy has been over the proposed relation of organochlorines to breast cancer, summarized by Arab (Topic 3.2).

PREMISE

In the face of controversy over the meaning of animal and human studies, the delay in reaching consensus, and the possibility of irreversible harm, the precautionary approach is viewed as a way of reducing exposure to EASs. The alternative wait-and-see approach, anticipates reassuring results when epidemiologic studies on human health impacts are finally completed decades from now. This paper examines the limits of epidemiology and why even apparently reassuring results (i.e., statistically negative or inconclusive) cannot be the sole factor influencing risk management. I arrive at the same conclusion as Tukker (2002), that epidemiology has "a limited role in leading to preventive action". I identify almost a dozen reasons for this limitation, only a few of which will be discussed here.

For the purpose of this paper I will assume that in the absence of a precautionary approach, no regulatory action would be taken without convincing epidemiologic evidence of causation, usually in the form of multiple, statistically significant studies in relevant populations.

Table 1 Aspects of the epidemiology literature which delay consensus on causation OR "Why epidemiology stacks the deck".

INTRINSIC TO METHODOLOGY
Alpha of 0.05 is too restrictive (avoiding type I error seems more important than avoiding type II error).
Focus on confidence limits retains the 0.05 level.
Willingness to question statistically significant results if sample size small.
Misclassification biases toward the null hypothesis, making type II error more likely.
Participation bias reduces power.
Multiple comparison corrections, over-correct, making type II error more likely.
Tendency to use two-tailed test, when one-tailed test is warranted makes type II error more likely.
Confounders are always available to take the blame and vitiate an association or delay its acceptance.
EXTRINSIC TO METHODOLOGY
Population size imposes a limit on power.
Opposition to meta-analysis.
Human health effects may have long latency.
Epidemiologic studies take several to many years to plan, execute, and publish.
Large-scale epidemiologic studies are prohibitively expensive.

EPIDEMIOLOGY "TO THE RESCUE?"

Before embarking on a critique of epidemiology from one who teaches the subject, it is important to recognize its value. In many aspects of environmental health, data obtained in animals are viewed with suspicion or skepticism when applied to humans. In the area of carcinogenesis, an agent that causes can-

cer in two test species, in both sexes, and in more than one organ, can be considered a probable human carcinogen. But only epidemiologic studies can validate definition as a "known" human carcinogen. In risk assessment, for example, there is a large and growing literature on the means and limitations of extrapolating from animal data to humans. And interspecies extrapolation is one of the main areas of uncertainty [11].

Epidemiology affords the opportunity to eliminate this uncertainty. Studies on asbestos and cancer, for example, have mainly been accomplished in humans. Likewise, arsenic and cancer are based on human data, since this is one carcinogen for which no animal model exists [12]. But the expectation that epidemiology can eliminate uncertainty and provide convincing answers for all audiences for all problems is an illusion. Epidemiology has severe limitations imposed by costs, methodology, and available populations [13].

INTRINSIC METHODOLOGIC ISSUES

Hypothesis testing in epidemiology and toxicology

In the mid-1900s, hypothesis testing emerged as a central paradigm in science. Scientists developed an idea or hypothesis, framed it in terms of a null hypothesis (Ho), gathered data observationally or experimentally, analyzed the data, and rejected or failed to reject the null hypothesis. Since the early 20th century, the 0.05 level was arbitrarily used as the criterion for rejecting the null hypothesis, so that rejecting the Ho at the alpha level of 0.05, meant that there was only a 5 % chance or one chance in 20 that the Ho was actually true (or should not have been rejected) [14]. This amounts to a 5 % probability of making a type I error (a false positive or rejecting the Ho when it is true and should not be rejected).

There are many fields where being 95 % percent confident of an experimental result is desirable or even necessary, but it is not a universal necessity. Moreover, this form of hypothesis testing does not allow an a priori definition of the converse probability of making a type II error, failing to reject the null hypothesis when it is false. This is contingent on the power of a study, power being the probability that if the Ho is false, the study will generate data that rejects the Ho. Rarely is power addressed in scientific publications, and often it cannot easily be calculated retrospectively. However, a power of 80 % is often considered desirable, meaning that even in a well-designed study, one has a 20 % chance of making a type II error. Thus, in the evolution of hypothesis testing, avoiding a type I error became more important—approximately four times more important—than avoiding a type II error. This means, in effect, that the deck is stacked against detecting an effect that is real.

Consider an example where drug A is being developed as a cure for disease D. A cautious investor might insist on being 95 % confident that A is indeed effective against D, but a consumer suffering from D would be content with a much lower level of confidence, such as 80 % or perhaps even 51 %. Thus, the traditional alpha of 0.05 is merely a statistical convention, not an absolute threshold for significance.

Hypothesis testing applies equally to toxicology and epidemiology, but in the former, at least, the experimenter has the capability—though not always the resources—to design studies of sufficient power.

Hypothesis testing has dominated the thinking of most scientists alive today. But simple hypothesis testing vis-à-vis the Ho, is already being replaced. Computerized techniques yielding absolute estimates of probability, afford new ways of looking at scientific data and providing confidence intervals around point estimates.

What does statistically significant mean

When we require a study to be statistically significant we look to see whether Ho is rejected at some a priori alpha level, usually the 0.05 level. Where effects are subtle or power is low, lack of statistical sig-

nificance at the 0.05 level is not reassuring. Knowing the power of a study in advance is often valuable, but sobering. Far too many epidemiologic studies lack the power to reject the Ho of no association. Negative studies abound, with p values in the range of 0.06 to 0.20 which are interpreted as "negative" or "not significant", but which are actually potentially positive studies which simply lacked power. Indeed, performing a study that lacks power may be unethical, unless one can clearly indicate the lack of power.

Alternative alphas

Perhaps the easiest way to deal with the inherent statistical bias favoring type II errors in epidemiology is to set alternative alpha values or ignore alpha altogether. With small sample sizes and low power, an alpha of 0.05 is an invitation to a false-negative study. Setting alpha at 0.10 has been shown as a reasonable basis when a one-tailed test should have been invoked. It is also reasonable when a study lacks power to reject at the 0.05 level. In any case, readers should look with suspicion on "negative" studies particularly where p values are in the range of 0.05 to 0.10. Dispensing with alpha altogether is also reasonable, since modern statistical programs can calculate an absolute P value for any statistic. It is simple enough to publish the result and the P value and let the readers bring their own interpretation to the data.

CONFIDENCE LIMITS

Confidence limits have been extensively used in biometrics for nearly 20 years, with almost all epidemiologic publications on relative risks or odds ratios, including the 95 % confidence limits. If the confidence limits include 1 (i.e., the lower 95^{th} % CL < 1), then this is interpreted as not statistically significant at the 0.05 level. Taking a hypothetical point estimate of 2.0 (95 % CL 0.5, 8.0), it is equally likely that the true value is 8.0 as it is that it is 0.5, and it is more likely than not that the true point estimate is above 1.0. Therefore, slavish reliance on confidence limits can be just as deceiving as simply rejecting or failing to reject the Ho.

Possible decision rule for those who insist on 0.05

- If Ho is rejected at the 0.05 level, then regulate or prevent exposure.
- If Ho is not rejected at the 0.05 level but the power is not adequate, then regulate or prevent exposure.
- If Ho is not rejected at the 0.05 level and power is adequate, then no regulation or prevention needed at this time [15].

Biological significance vs. statistical significance

One commonly hears that a particular study yields a statistically significant result, "but is it biologically significant?" That is, even in the face of statistical significance, readers may question the meaning of a result. It is important that the converse question gets equal billing—"even if it isn't statistically significant, does that mean that it's not biologically significant?" One of the commonest reasons for not achieving statistical significance is lack of power in the study design, usually because of small sample size. If a study achieves statistical significance (rejects the Ho), then it cannot, by definition lack power, even though the results may remain open to question and confirmation by additional studies.

Multiple comparisons

Not only does an alpha of 0.05 carry with it an unacceptably high probability of making a type II error, but there is another barrier imposed. Once one undertakes a study, particularly an expensive epidemiologic study, it is desirable to investigate as many endpoints or as many exposures as possible. Thus, if one is studying the effect of soy milk formulas on infant development comparing soy- to cow-based or breast-fed babies, a series of endpoints might be considered including height attainment, weight for height, bone ossification, cognitive development, and gonadal and genital development. If for example, only the latter were statistically significant at the 0.05 level, one might interpret this as evidence for a phytoestrogen effect. But skeptics would argue, "no", when you have six endpoints that are being tested, that increases the chance that any one of them might emerge as "significant" by chance alone, therefore you have to correct for multiple comparisons. Statisticians apply post hoc tests for multiple comparisons, for example, the Bonferroni test.

However, if you consider that each of the six hypotheses is an independent null hypothesis, any one of which might be biologically as well as statistically significant, it is apparent, that a multiple comparison test merely increases the probability of a type II error and is inherently fallacious. Rothman [16] provides a detailed statistical treatise on why one should not correct for multiple comparisons in most epidemiologic studies.

Confounders

Unlike multiple comparisons, confounding is a real complication in interpreting epidemiologic studies. There are many sources of confounding. When the study population is large and resources unlimited, it is often possible to stratify samples and correct for confounding. In reality, however, confounding often cannot be eliminated in a single study, and critics are quick to seize upon this. Confounding arises when both the independent variable (e.g., soy milk) and the dependent variable (growth rate) are correlated with a third variable that is not tested or testable. Socioeconomic status (SES) is a frequent confounder. If people with higher SES are more likely to go to pediatricians who recommend soy milk (more expensive) and also enjoy better nutrition and medical care, then a relation between soy and growth rate would be confounded by SES, and this might obscure a negative relationship between soy and growth, if SES had been corrected for.

Authors, editors, referees, or agencies, invoke confounding as a way of explaining or undermining positive results. In its famous study of dioxin exposure and cancer in 10 industries, the National Institute for Occupational Safety and Health qualified its positive study by noting that "we cannot exclude the possible contribution of factors such as smoking and occupational exposure to other chemicals," even though the nature of the study design made confounding highly unlikely [17]. More recent papers [18] attempt to quantify the possible impact that confounders might have on a relative risk (RR) or odds ratio (OR).

Misclassification bias

In any epidemiologic study where an exposed and referent population are compared, it is assumed that most or all of those in the exposed population were actually exposed while most or all of the referent population were unexposed. It is apparent that this is rarely likely to be the case. In an industrial-based study, not all workers have the same exposure. In a study of fish-eating populations exposed to mercury in the Seychelles and Faroes, not all families ate the same large quantities of fish. Where possible epidemiologists try to rank exposures, creating at least a crude dose–response relationship. In addition to exposure misclassification, there can be outcome misclassification. Some people with the disease may be overlooked. Other people may be ascribed a disease that they did not have. This is particularly a problem with death certificate studies of mortality causes. Any random misclassification in exposure or

outcome biases toward the null hypothesis [19], making a type II error more likely. Since no study can absolutely ascertain exposure, this is a nearly universal conservative bias.

Subject protection and institutional review boards

Some historical epidemiologic studies, most notably the Tuskeegee syphillis study, showed that subjects in epidemiologic studies might be placed in jeopardy by the study itself. Even in controlled clinical trials, subjects on placebos may be denied promising therapies. Therefore, following the World Medical Association's Declaration of Helsinki [20], human subjects protection and institutional review boards (IRBs) emerged, and their review is now required for every study involving human subjects. Increasingly, IRBs are tightening the standards for epidemiologic studies. In some cases, full disclosure of a study's purpose may create additional biases by influencing whether subjects participate or how they respond. There is an increasing concern in the public health community that IRBs are jeopardizing the future of important areas of epidemiologic research, and it may become impossible to expect epidemiology to ever address the questions now posed of it.

Participation bias

One consequence of increased requirements posed by IRBs is a reduction in participation rates. In toxicology, experimental rodents are not given a choice about participating in studies. Most epidemiologic studies require consent of the subjects, at least a willingness to answer questions or provide blood samples. This is now being extended to review of records of past illnesses, where no physical harm to subjects occurs. Some subjects, particularly African Americans, are suspicious of any epidemiologic studies and refuse to participate, thereby leading to the under-representation in all studies and the lack of ethnic-specific information. Participation bias stacks the deck in two ways. It reduces the sample size, thereby reducing the power of the study. It is always a source of suspicion in interpreting the results (whether negative or positive). Thus, a positive study with a high refusal rate may be dismissed on the grounds that the results were influenced by selection or participation bias.

EXTRINSIC EPIDEMIOLOGY ISSUES

Population limitations

A major limitation of epidemiology is that for any give type of exposure, the population that can be studied is not infinite. For example, only a few thousand people are occupationally exposed to putative xenoestrogens. The tobacco industry took delight in claiming that the cancers attributable to smoking were probably due to alcohol consumption since the two exposures were highly correlated. Finding smokers who consumed little or no alcohol was challenging, and finding populations with neither, even more so. If there are only a few hundred people exposed to a particular chemical, then power is limited by the available universe, even if all participated.

Uneven receptor populations

Risk assessment and standard setting often take into account that among humans a subset (usually ill-defined) may be at substantially greater than average risk. A common default value of 10 is used, assuming that susceptible people are 10 times more vulnerable than average. Epidemiologic studies may or may not incorporate susceptible individuals, and the proportion of susceptibles in a study may vary greatly. In the absence of specific investigations (such as genotyping) this proportion is unknown. If the proportion is low, the impact of an exposure will be underestimated; if the proportion of susceptibles is high the impact may be overestimated. This is another type of misclassification.

Even when exposure is correctly classified, the inclusion of a high percentage of nonsusceptibles in the exposed population militates against detecting an effect, even when the same percentage is present in the controls. For example, a study of breast cancer that includes a large number of Asian women would be biased toward the null [18].

Meta-analysis and the precautionary principle

Meta-analysis offers several ways of combining studies, each of which alone lacks power, but which in combination may generate sufficient power to allow confidence in the outcome. Despite initial skepticism [22], meta-analyses are performed with increasing frequency and can be used to demonstrate or refute a proposed association. Often, however, meta-analysis does not provide a definitive answer as Hebert-Croteau [23] concluded in an analysis of estrogen and colon cancer: "Inadequate assessment of exposure, poor control of confounding factors, and changing patterns of use over time might have contributed to the slow emergence of this association postulated almost two decades ago. Additional large studies are needed to replicate this finding and explain the exact mechanism of this putative protective effect."

CONCLUSION

It is this last conclusion that highlights the need for precautionary approaches to many environmental health issues, including endocrine active substances. Epidemiologic studies are slow, often weak, and seldom conclusive, and decades may be required to reach consensus. Several standard defaults in epidemiology stack the deck against detecting associations, particularly weak associations, making it much more likely that a false Ho will be accepted (type II error) than that a true Ho will be rejected (type I error). The failure to report power, the setting of alpha at the artificially high default of 0.05, correcting unnecessarily for multiple comparisons, or blaming positive effects on hidden confounders are all ways of making detection of biologically significant associations quite unlikely. Faced with this problem, it is not only reasonable, but desirable to take effective precautionary actions to mitigate exposures to putatively hazardous materials, while awaiting additional scientific evidence, particularly when the consequences of inaction or laissez-faire, may be serious or irreversible.

ACKNOWLEDGMENTS

Joanna Burger, Russellyn Carruth, Bernard Goldstein, and many other colleagues have provided valuable insights on the precautionary principle. This project was partially funded by National Institute of Environmental Health Sciences (ESO 5022), the Department of Energy (AI # DE-FC01-95EW55084, DE-FG 26-00NT 40938) grant to the Consortium for Risk Evaluation with Stakeholder Participation (CRESP), and the conference hosts SCOPE and IUPAC.

REFERENCES

1. T. Colburn, D. Dumanoski, J. P. Myers. *Our Stolen Future*, Dutton, New York (1996).
2. C. Schairer, H.-O. Adami, R. Hoover, I. Persson. *Epidemiology* **8**, 59–65 (1997).
3. U.S. Preventive Services Task Force. *Ann. Intern Med.* **137**, 834–839 (2002).
4. National Research Council. *Hormonally Active Agents in the Environment,* National Academy Press, Washington DC (1999).
5. B. C. Gladden and W. J. Rogan. *J. Pediatrics* **119** (Part 1), 58–563 (1991).
6. Y. L. Guo, T. J. Lai, S. H. Ju, Y. C. Chen, C. C. Hsu. *Organohalogen Compd.* **14**, 235–238 (1995).
7. L. J. Guillette, Jr., D. B. Pickford, D. A. Crain, A. A. Rooney, H. F. Percival. *Gen. Comp. Endocrinol.* **101**, 32–34 (1996).

8. P. A. Bertazzi, A. C. Pesatori, D. Consonni, A. Tironi, M. T. Landi, C. Zocchetti. *Epidemiology* **4**, 398–406 (1993).
9. K. Kreiss, C. Robert, H. E. Humphreys. *Arch. Environ. Health* **37**, 141–147 (1982).
10. V. H.-H. Goh, S.-E. Chia, C.-N. Ong. *Environ. Health Perspect.* **106**, 41–44 (1998).
11. H. R. Pohl and H. G. Abadin. *Regul. Toxicol. Pharmacol*. **22**, 180–188 (1995).
12. National Research Council. *Arsenic in Drinking Water*, National Academy Press, Washington DC (1999).
13. A. Tukker. *J. Epidemiol. Commun. Health* **56**, 883–884 (2002).
14. R. A. Fisher. *Statistical Methods for Research Workers,* Oliver & Boyd, Edinburgh (1925).
15. R. M. Peterman and M. M'Gonigle. *Mar. Pollut. Bull*. **35**, 231–234 (1992).
16. K. J. Rothman. *Epidemiology* **1**, 43–46 (1990).
17. M. A. Fingerhut, W. E. Halperin, D. A. Marlow, L. A. Piacitelli, P. A. Honchar, M. H. Sweeney, A. L. Greife, P. A. Dill, K. Steenland, A. J. Suruda. *N. Engl. J. Med.* **324**, 212–218 (1991).
18. M. Goodman, M. Kelsh, K. Ebi, J. Ianuzzi, B. Langholz. *Epidemiology* **13**, 50–58 (2002).
19. A. H. Wu. *Environ. Health Perspect.* **107** (Suppl. 6), 873–877 (1999).
20. World Medical Association Declaration of Helsinki. World Medical Association (1964) (as amended) see <http://www.wma.net/e/policy/17-c_e.html>.
21. S. L. Chow. *J. Psychol*. **121**, 259–271 (1987).
22. N. Hebert-Croteau. *Cancer Epidemiol. Biomarkers Prev.* **7**, 653–659 (1998).