1. My name is R[redacted] Clapp. I have been asked to prepare a report in this matter by Linda Laurent, at the law firm of Reich and Binstock. I received my undergraduate degree in Biology from Dartmouth College (1967), and I received an MPH degree from the Harvard School of Public Health (1974) and a D.Sc degree in Epidemiology from Boston University (1989). From 1980 to 1989 I served as Director of the Massachusetts Cancer Registry in the Massachusetts Department of Public Health, Boston, Massachusetts. I have worked as a consultant and academic in environmental health and epidemiology since 1989. I am a member of the American College of Epidemiology, the Society for Epidemiologic Research, the International Society for Environmental Epidemiology, and numerous other scientific and professional organizations. Since 2001, I have been Professor at the Boston University School of Public Health and have published additional scientific articles. A current copy of my curriculum vitae is attached to this affidavit.

2. This report is my summary of the evidence that specific chlorinated pesticides cause or contribute to non-Hodgkin's lymphoma in humans. In arriving at this conclusion, I have followed the same approach I use in all my work as a professional epidemiologist. I begin with a description of the methods I have followed in arriving at my general causation opinion in this matter.

3. Whether in an epidemiologic or medical-legal context, I use the same guidelines, perspectives and tools for analysis to evaluate whether or not there exists a causal connection between workplace or environmental exposures, and latent disease. In both contexts, I use and apply the well established, widely recognized and useful "Hill guidelines," which Professor Austin Bradford Hill offered in 1965, as set forth and explained herein.

4. Epidemiologists concerned with the causes that contribute to human cancer risk routinely use the Hill guidelines or "viewpoints" as a set of useful tools for drawing scientific inferences and deductions about causation from all the available relevant principles, data,

information, and observations. [1] I present here a discussion of how epidemiologists inquire into the contributions of the environment, particularly the workplace environment, to causing cancer.

1. The Experimental and Observational Sciences Are Used to Provide Data to Make Judgments About Causality

Scientific practice is taken up with more than exploring questions of causation, but this is a central question in many tort cases. What does "A causes B" mean to a scientist? Apart from philosophical aspects of scientific causality, most scientists have adopted a pragmatic approach whose formal articulation goes back at least to John Stuart Mill's famous "Method of Difference."[2] Briefly, Mill's Method holds that A causes B if, all else being held constant, a change in A is accompanied by a subsequent change in B. The formal method to detect such an occurrence is the Experiment, whereby:

- all things are held constant except A and B,
- A is varied, and
- B observed.

Not all sciences can utilize a strictly experimental method, however. Some scientists must be content to make observations of the real world and deduce scientific fact by applying reasoning and principles from experimental sciences or logic and mathematics. Astronomy, geology, and epidemiology are such sciences. Astronomers cannot manipulate distant stars and planets experimentally, but they can apply the methods and results of terrestrial physics along with mathematical theories like quantum mechanics or relativity theory to make inferences about the interiors of stars or the structure of other galaxies. Another observational science is geology. In one of its sub-disciplines, seismology, scientists *observe* earthquakes; they certainly do not stage city-

---

[1] The historical context of these guidelines is of interest: Sir Bradford Hill proposed his viewpoints in 1965, well before the International Agency for Research on Cancer (IARC) or U.S. agencies such as the EPA or OSHA had begun promulgating lists and categories or carcinogens. Further, Dr. Bradford Hill's own commentary on the use of his guidelines was most instructive: **they are not meant to replace common sense and judgment but to aid them.**

[2] John Stuart Mill, System of Logic: Ratiocinative and Inductive. (Longman, Green, & Co. 1906.

sized experiments on the factors that cause earthquakes. The inability of geology or astronomy to conduct full-scale experiments does not connote an inability to do good science, or that the science involved is inherently more "error prone" or less reliable than a branch of science that can conduct full-scale experiments.

Scientists may, however, extrapolate from laboratory scale experiments to make scientifically defensible statements about the origins of a "black hole" in space or the causes of earthquakes on our planet. There may be disagreement among experts as to the aptness of a *particular* extrapolation or inference, but generally there is no disagreement that the process of applying events or principles observed on the scale of the laboratory bench to events occurring on the scale of a geographic region is scientifically defensible, and indeed something similar is the norm in virtually all observational sciences.

In the biological sciences, in general, and in the public health field, in particular, inferences for one group of humans are regularly drawn from epidemiological studies from another group of humans. Significantly, inferences about humans are also made on the basis of observations of, or test-tube experimentation, on animals. Indeed, the scientific reasonableness of drawing inferences from animals to humans provides the principal justification for the decision of National Institutes of Health to devote hundreds of millions of dollars funds to animal research. Any *particular* inference may be arguable, and certainly may be the basis of a dispute between the parties in a lawsuit, but the *method and reasoning* are not subject to debate.

In general there are three sources of information on the effects of toxic exposures in human beings: (a) case reports, (b) toxicological research (including both animal studies and chemical/structural research), and (c) epidemiological studies.

(a) The use of case reports regarding the effects of toxic exposures in human beings – A case report, *i.e.*,, a report in the medical or scientific literature of a single case or series of cases, are one of the most important sources of information scientists have on effects of toxic substances, and often the only source of information. Detailed reports of cases of accidental poisonings or suicides provide information, such as autopsy data, biopsies and detailed clinical data, not obtainable

by any other route. Moreover they constitute important and obvious "natural experiments," experiments where the relationship between the exposure and effect is usually clear. The use of case reports in medicine is longstanding and important, as evidenced by the continued appearance of such reports in the literature[3]. Indeed the *logic* of a case report is similar to that of a more formal case-control or cross-sectional study.

(b) **The use of toxicological research reports to understand the effects of toxic exposures in human beings** – Toxicological research (including both animal studies and chemical/structural correlations), along with epidemiology, is one of the two other sources of information provides much of the basis for scientific judgments relating toxic exposures to health effects.

Toxicology is an *experimental* science, while epidemiology is an *observational* science.[4] The advantages of being able to conduct an experiment are obvious. Because John Stuart Mill's famous Method of Difference depends upon observing the result on B of a change in A, *other factors must be held constant.* The essence of an Experiment is the control of all factors, except for A and B. This kind of control allows the scientist to ask quite precise questions about explicitly defined A's and B's, and get relatively unambiguous answers.[5]

(c) **The use of epidemiological studies regarding the effects of toxic exposures in human beings** – Epidemiological studies are observations of "natural experiments" that are occurring in the real world. The idea is to find situations which are almost like laboratory experiments, observe them, obtain as much information as possible from them, and then interpret the results. The essence of the natural experiment in epidemiology is almost always a comparison between groups, for example, a group exposed to a chemical and one not exposed. The ideal situation would be to have the groups in the real world the same in all relevant respects (*i.e.*,

---

[3] Ibid at 373-426.

[4] *The Lancet*, for example, one of the world's leading medical journals, contains a Case Report *every week*

[5] Whether complete control is practically possible varies, of course, but the principle should be clear.

comparable) except for the variable under study. Unfortunately such natural groupings are rarely comparable, and techniques must be used to account for known differences. However, not all sources of non-comparability are known. If not a necessary accompaniment of the variable being investigated, these residual factors fall by chance in the two groups being compared. The result is that there are usually differences solely attributable to the random way these factors are distributed between groups in the particular study. The "chance" fluctuations in apparently otherwise similar populations require an epidemiologist to use statistical tools to evaluate the role of "noise" that might be obscuring an underlying "signal."

Observing some unintended or "natural" experiment in the real world, which is the essence of observational sciences like **epidemiology**, has the enormous advantage that it involves human beings living under conditions similar to ones found by plaintiffs in a personal injury suit.

Nonetheless, questions inevitably arise about the biological/scientific comparability (and thus the legal relevance or "fit") of the people and exposures and diseases studied in one place and time and other people at other places and times. For example, questions such as whether the comparison of the cases and controls was truly comparing "like with like," which is precisely the kind of problem that can be and generally is avoided in a tightly controlled experimental study. Thus, as my colleague Dr. David Ozonoff explained in detail in a 1994 peer reviewed article, toxicological experiments and epidemiological studies each have characteristic strengths and weaknesses. [6]

In view of the fact that different scientific disciplines have disparate strengths and weaknesses, and the propensity of scientists to disagree, the key question for scientists – and courts – becomes determining how scientists decide which studies and data and experiments and articles to use and rely and for what purposes, *i.e.*, how do they *interpret and apply* the results of scientific studies?

---

[6] Ozonoff, D "Conceptions and Misconceptions about human Health Impact Analysis" *Environment Impact Assessment Review,* 14, 499-516, 1994.

2. Interpreting Scientific Studies: How Scientists Make Judgments About Causality About the Collective Information They Observe And Review

It is well know that when different scientists interpret the same studies they do not always reach the same conclusion. How and why do scientists interpret the "same" basic facts, the same set of numbers, the same research report, in different ways?

Two aspects and tools of scientific interpretation are relevant to this discussion. In the literature of scientific methodologies they are commonly (but not invariably) referred to as internal and external validity.

Internal validity refers to a judgment about the extent to which the experiment or study produces valid information on its own terms, *i.e.*, the extent to which it is *internally* valid. Thus, for internal validity the crucial question to be answered is not, "If benzene causes cancer in C57BL/6 mice does it also do so in Wistar rats or humans?" but rather "Did the experiment validly show that benzene caused cancer in C57BL/6 mice?"

External validity, on the other hand, refers to a judgment about the extent to which the internally valid results of an experiment or study can be generalized to other situations, and to which ones. Thus, for external validity the crucial question to be answered is not "did the experimental evidence adequately demonstrate that benzene causes cancer in Wistar rats?" but rather "If benzene *does* cause cancer in Wistar rats, does it *also* do so in humans?"

a. Internal validity: How good is the study in its own terms?

At the heart of a case report, a toxicological experiment, or an epidemiological study lies a comparison.

* Case reports are usually made to call the attention of the medical community to an "interesting" observation (compared to "the usual"), such as a rare disease in the context of an unusual exposure (e.g., sudden death in a worker exposed to chlorinated solvents). The comparison is with previous or usual experience.

* In an experiment, the comparison is between the different states of B, when A is varied.

* In an epidemiological study, it is the analogous comparison in the "natural" or unintended experiment that is being observed.

Once an unusual event is observed or an unexpected experimental result is obtained it remains to explain or interpret the observation or result, whether the result is a difference or a lack of a difference in the expected or compared entities.

Take as an example a study comparing the health outcome of two distinct groups of human beings, one group comprised of those workers in a factory who were exposed to a chemical used in the production process, and the other group consisting of all members of the general population, most (but perhaps not all) of whom were not exposed to the chemical. Suppose the workers have more disease than the general population (the following analysis works just as well in the case of no increased disease). There are three generic reasons such a difference (or lack of difference) might be observed, referred to as "bias," "chance," and real effect. They are conceptually independent, but not mutually exclusive forces, *i.e.*,, all, some, or none can operate simultaneously. Each must be evaluated to extract a valid message ("the real picture" or the "true signal") from the study.

(1) **Understanding the Role of "Bias" in Evaluating the Internal Validity of a Research Study**: Another term for "bias" is "systematic error." This differs somewhat from the common usage of the word (*e.g.*, lack of objectivity, independence, and/or neutrality), and in epidemiology the word has been refined and qualified to encompass a wide variety of sources of systematic error, each given a name.[7]

For example, epidemiologists talk of various kinds of "information bias" ("recall bias," "observation bias," "differential or non-differential misclassification bias," etc.) or types of "selection bias," "confounding bias," etc. All biases have as their underlying mechanisms factors that make the compared groups different in more ways than just the variable being studied. Because the object of an experiment or study is to isolate one element (exposure to the chemical in my example), one must estimate the effect of the uncontrolled differences on the comparison.

---

[7] See e.g. Bailey LA, Gordis L, Green M "Reference guide in epidemiology: *Reference Manual on Scientific Evidence*, Federal Judicial Center, 1994, pp. 121-80, esp. p. 132.

A common source of potential bias in an epidemiological study is "confounding," and I illustrate bias with this example. Suppose scientists were comparing cancer rates in two groups. As in all epidemiological studies, this comparison is of the nature of an experiment, but one that is "handed to us" by nature (*i.e.*, circumstance), not one of our own devising. Thus scientists are unable to control everything they might like in this comparison.[8] It might be, for example, that the workers in this hypothetical instance are considerably younger than the general population, and because cancer risks rise with age, they would be expected to have less cancer than the comparison group. If this difference were not somehow accounted for, the observed increase in the number or incidence of cancers in the worker group is actually likely to underestimate the true effect. The same non-comparability could influence a comparison in the opposite way if the workers were on average older than the general population (say, if they were a group of retirees).

The most important means of coping with bias is to recognize it. An important part of the training and practice of an epidemiologist is to recognize and account for the effects of the inevitable non-comparability found in observational studies. Once recognized, an epidemiologist can often gauge the impact of a source of bias on the results and adjust conclusions accordingly. Sometimes the data themselves can be "adjusted" ("controlled") to eliminate the non-comparability in the two groups for certain factors like age or sex.

(2) <u>Understanding the Role of "Bias" in Evaluating the Internal Validity of a Research Study, and Understanding the Insignificance of "Statistical Significance"</u>: Not all sources of non-comparability are known.[9] Providing that they are not a necessary accompaniment of the variable being investigated, these residual factors are distributed by chance between the two groups

---

[8] Epidemiologists depend on Nature to be their "research assistants" and Nature is not usually very tidy or cooperative. Thus it is normal and natural for there to be "loose ends" sticking out of epidemiological studies, loose ends that other epidemiologists usually cannot resist pulling. As with most things, designing an informative study is difficult. Criticizing one is easy.

[9] This is a deterministic view of disease causation. One could also take a probabilistic view, in which case scientists would have to discuss sample error from some assumed super-population of identical study settings. This alternative view does not affect any of the points made

being compared. The result is that there are usually differences solely attributable to the random way these factors are distributed between groups in the particular study. The "chance" fluctuations in apparently otherwise similar populations require an epidemiologist to use special tools to discern the true meaning from the chaos of disparate data – to "see" the true picture amidst a welter of images, or to "hear" the true, underlying "signal" in the midst of the noise produced by these variations. The mathematical tools used for these purpose involve statistical analysis.

The main purpose for statistics in epidemiology, then, is to evaluate the role that random effects ("chance") might have played in the results. Statistical methods *do not* prove that chance is the source of a difference (or lack of difference). These methods only provide information on how likely it is that chance *could* have played a part if there were no bias and no true effect. The meaning of "statistical significance" is that the likelihood that chance *could* have produced the observed results *if there were no bias and no real effect* is less than some arbitrarily predetermined level, such as 5% ("p<.05").[10]

For the reasons stated above, it is absolutely false – and, indeed, a serious interpretive error – to assert that a result that is not "*statistically significant*" means the results must be due to chance and only to chance. And for these reasons, prominent epidemiologists eschew "statistical significance," believing that it is not a *sine qua non* of "good science" and maintaining that "it is neither necessary nor appropriate as a requirement for drawing inferences from epidemiologic data."

[10] The original source of the 5% criterion is lost in time. It apparently came from the original applications of statistical methods to agricultural experiments and expressed a cost-benefit statement about the expense of redoing a large trial involving a whole growing season and plots of various seeds and fertilizers. Its use for public health purposes might thus be questioned. It is interesting to note that in other sciences, notably, physics, another common criterion for "statistical significance" is not 5% but 10%. In any event, virtually every elementary statistics text warns the student of the highly arbitrary nature of the figure.

These views are hardly mine alone. Instead, they are representative of the views of both Sir Austin Bradford Hill, one of the 20th century's preeminent statisticians, and some of most highly regarded epidemiologists in this country, such as Dr. Kenneth Rothman (who is: (a) the co-author of the most widely used textbook on epidemiology; (b) the former Editor-in-Chief of the peer-reviewed journal, EPIDEMIOLOGY; and, not least, (c) my colleague at the Boston University School of Public Health) as well as other epidemiologists, such as Dr. Noel Weiss. Thus, Hill chided those who relied on "significance tests" to prove or disprove causation.

> No formal tests of significance can answer those questions. ("Is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?") Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the 'proof' of our hypothesis." - "I wonder whether the pendulum, has not swung too far -- not only with the attentive pupils, but with the statisticians themselves. - Fortunately I believe we have not yet gone so far as our friends in the USA where, I am told, some editors of journals will return an article because tests of significance have not been applied. . ."[11] .

Similarly, in an amicus brief to the US Supreme Court in the *Daubert* case, Professors Rothman and Weiss, and others including me, stated: "***Significance testing, however, is neither necessary nor appropriate as a requirement for drawing inferences from epidemiologic data.***" [12]

> The amicus brief continued:
>
> *The notion that only when data demonstrate "statistical significance" do epidemiologists draw inferences about observed associations between suspected risk factors and medical conditions is mistaken.* Significance testing is nothing more than a statistical technique that attempts to evaluate what is called "chance" as a possible explanation for a set of observations, and classify the

---

[11] Austin Bradford Hill, *The Environment and Disease - Association or Causation?* Proceedings of the Royal Society of Medicine (1965) 58: 296 at p. 299.

[12] Rothman and Weiss, "Summary of Argument" section of their *amicus* brief in *Daubert.*

> observations "significant" or "not significant" based on the likelihood of observing them if there were no relationship between the suspected cause and effect. *Testing for significance, however, is often mistaken for a sine qua non of scientific inference.* . . . Scientific inference is the practice of evaluating theories. As such, it is a thoughtful process, requiring thoughtful evaluations of possible explanations for what is being observed. Significance testing, on the other hand, is merely a statistical tool that is frequently, but inappropriately, utilized in the process of developing inferences.
>
> * * *

Significance testing, in my opinion and in the view of many respected scientists, places too high a value on a "yes-no" answer to an oversimplified question: Is the probability that the observed association could appear by chance, even if there is no actual relationship, low enough to justify rejection of chance as the explanation of the observed association? The result of using significance testing as the criterion for decision-making is that the focus is changed from the information presented by the observations themselves to conjecture about the role chance could have played in bringing about those observations. Dr. Rothman has stated the issue thus:

> With the focus on statistical significance, if chance seems to be a plausible explanation, then other theories are too readily discarded, regardless of how tenable they may be. As a result, effective new treatments have often been overlooked because their effects were judged to be "not significant," despite an indication of efficacy in the data. Conversely, if "significance" seekers find that the results of a study are calculated as improbable on the basis of chance, then chance is often rejected as an explanation when alternative explanations are even less tenable. [13]

---

[13] Rothman et al., amicus brief in *Daubert*, citing K. Rothman, *Significance Questing*, 105 ANNALS OF INTERNAL MEDICINE 445, 445-46 (1986) (citations omitted). According to the Rothman and Weiss amicus brief:

> A better approach to evaluating the error in scientific measurement is the use of "confidence intervals." A confidence interval is a range of possible values for a parameter that is consistent with the observed data within specified limits. The process of calculating a confidence interval within the chosen limits is know as "interval estimation." See K. Rothman, Significance Testing at 119.
>
> An important advantage of interval estimation is that it: "do[es] not require irrelevant null hypothesis to be set up nor [does it] force a decision about 'significance' to be made -- the estimates can be

The outcomes of statistical tests are strongly influenced by the size of the study population. For small populations, very large observed differences, of substantial *public health* significance, may still not be *statistically* significant[14]. That is to say, a large effect that a scientist would take seriously from the public health point of view cannot be differentiated on its face from chance. Either chance or a real causal influence (or bias) could be responsible for the worrisome effect. Conversely, in large populations, very slight and substantively meaningless differences can be "statistically significant."[15]

Statistical methods are sometimes viewed as standard, agreed-upon, and mechanical procedures. Scientists even allow computers to do them, seemingly without human intervention. But as any statistician knows, there is a great deal of judgment in deciding which tests to use in which circumstances, which tests are valid in those circumstances, and what they do and do not mean. Less well recognized is that statistics itself is, like all active disciplines, a field in ferment and change. Thus not all statisticians will agree on the propriety of even commonly used tests[16]. In his recent book, *Statistical Inference,* Michael Oakes has written:

> presented and evaluated by statistical and other criteria, by the researcher or the reader. In addition the estimates of one investigation can be compared with others. While it is often the case that different measurements or methods of investigation or theoretical approaches lead to 'different' results, this is not a disadvantage; these differences reflect important theoretical differences about the meaning of the research and the conclusions to be drawn from it. And it is precisely those differences which are obscured by simply reporting the significance level of the results.

Rothman, et al., amicus brief in *Daubert*, quoting L. Atkins and D. Jarrett, *The Significance of "Significance Tests,"* in J. Irvine and I. Miles (eds.) *Demystifying Social Statistics* (1979).

[14] A detailed example showing how results can be of public health significance but not statistical significance can be found in Ozonoff, David, "Conceptions and Misconceptions about Human Health Impact Analysis," *Environmental Impact Assessment Review,* 14:499-516, 1994.

[15] For example, a difference of 1/8" in height between east coast children and west coast children will be statistically significant if very large numbers of children on both coasts are measured.

[16] A good example is the Fisher Exact Test, commonly used for small tables frequently encountered in environmental epidemiology. Certain well known statistical programs even force the user to employ this test if several table cells contain expected values of less than five, even though it has been known for years that the test is inappropriate. Cf. D'Agostino R, Chase W, Belanger A, "The appropriateness of some common procedures for testing the

> It is a common complaint of the scientist that his subject is in a state of crisis, but it is comparatively rare to find an appreciation of the fact that the discipline of statistics is similarly strife-torn. The typical reader of statistics textbooks could be forgiven for thinking that the logic and role of statistical inference are unproblematic and that the acquisition of suitable significance-testing recipes is all that is required of him.[17]

When used, statistical methods are meant to help scientists *evaluate* the possible role of chance[18]. Scientists must evaluate the possibility of a concurrent *real* effect separately, as I now discuss.

(3) **Understanding the Relative Role of a Real Effect (or the Absence of a Real Effect) in Evaluating the Internal Validity of a Research Study**: The most important reason for a difference between two groups, however, is an actual effect or influence from the variable being studied (exposure at work in my example), *i.e.*,, that "A *does* cause B." As discussed in greater detail below, scientists recognize that **"causation"** should *not* be regarded as an experimental or epidemiological result, but rather as a *"judgment"* made about the experimental or epidemiological data. See Federal Judicial Center REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (1994) at p. 157 ("causation is a judgment issue for epidemiologists and others interpreting the epidemiological data."). See also the extended discussion of this point in K. Rothman & S. Greenland, *Causation*

---

equality of two independent binomial populations," *Am Statistician* 42:198-202, 1988, and references therein.

[17] Oakes, Michael, *Statistical Inference*, Epidemiology Resources Inc., Chestnut Hill, MA, 1990. Oakes then goes on to quote a review (by Dusoir) of a statistics text in a technical journal:

> "A more fundamental criticism is that the book, as almost all other elementary statistics texts, presents statistics as if it were a body of coherent technical knowledge, like the principles of oscilloscope operation. In fact statistics is a collection of warring factions, with deep disagreements over fundamentals, and it seems dishonest not to point this out."

[18] As expressed by the epidemiologist Kenneth Rothman in his *Daubert amicus* brief: "The result of using significance testing as a criterion for decision making is that the focus is changed from the information presented by the observations themselves to conjecture about the role chance *could* have played in bringing about those observations." [emphasis in original]. Quoted by Berger M, cited above (op. cit., note 8). Rothman is the author of a standard text, *Modern Epidemiology* (see next note), and formerEditor in Chief of the journal *Epidemiology*.

and Causal Inference," in: K. Rothman and S. Greenland, MODERN EPIDEMIOLOGY (Second ed. 1997) at pp. 7-28[19].

It is apparently not always appreciated that this is true. There is a tendency to believe that somehow "causation" is not a subjective judgment or interpretation but an actual, real, objective, discoverable, and measurable property of a relationship that can be demonstrated empirically, as if some associations had readable labels on them that said 'causal' and all that scientists need is the right instrument to read the label.[20] In sum, although some scientists may be loathe to admit it, and although many lawyers and judges may not believe it, there is simply no magic formula or easy checklist for making scientific judgments[21].

---

[19] As professors Rothman and Greenland explain, at p. 22 of their textbook:

> Perhaps the most important common thread that emerges from the debated philosophies [of scientific causation] is Hume's legacy that proof is impossible in empiric science. This simple fact is especially important to epidemiologists, who often face the criticism that proof is impossible in epidemiology, with the implication that it is possible in other scientific disciplines. Such criticism may stem from a view that experiments are the definitive source of scientific knowledge. Such a view is mistaken□Even the most careful and detailed mechanistic dissection of individual events cannot provide more than associations□ .

[20] Thus Judge Kosinski, in the *Daubert* remand, writes of the plaintiff's case that it does not "attempt to show causation directly; instead, they rely on experts who present circumstantial proof of causation." Of course there is no such thing as a "direct" proof of causation.

[21] Professors Rothman and Greenland are not alone in their view that judgment – not a checklist -- is a scientist's most useful tool in inferring causation. Indeed, that perspective is shared by a number of the nation's leading epidemiologists and other scientists, historians of science, and philosophers of science. Thus, an amicus brief tendered to the US. Supreme Court in the *Daubert* case by Harvard professors Stephen Jay Gould (Zoology, Geology, and History of Science), Gerald Holton (Physics and History of Science), Everett Mendelsohn (History of Science), and Kathleen Joy Propert (Biostatistics), Columbia University professor Ronald Bayer (Sociomedical Sciences), and NYU professor Dorothy Nelkin (Sociology and Law) explained that "'[c]onclusiveness in inferring causality -- in epidemiology as with the study of all free-living human beings – is a desire more often than an accomplishment.'" Amicus Brief of Bayer, Gould, etc., quoting Mervyn Susser, *Rules for Inference in Epidemiology*, 6 REGULATORY TOXICOLOGY AND PHARMACOLOGY 116, 127 (1986). These scholars went on to observe that "[a]s a consequence, those who seek in science the immutable truth they find lacking in the law are apt to be disappointed." (*Ibid.*) Furthermore, "One notable similarity [between law and epidemiology] is the dependence of both fields upon ***subjective judgments***. In the end, a quality which lawyers should understand -- judiciousness – matters more than any. Scientists use both deductive and inductive inference to sustain the momentum of a continuing process of research. The courts of

(4) **Understanding the Relative Significance -- and Insignificance -- of "Negative" Studies**: Understanding the operation of bias and chance is especially important in interpreting so called "negative studies" (studies where no differences are apparent, or where the differences are not "statistically significant"). Differences produced by real effects can easily be masked by poor exposure classifications (misclassification bias), chance can appear as a possible explanation merely by virtue of a small population available for study (poor statistical power), and potential risks can be undetectable by observing the exposed population for too short a time (bias produced by failure to account for adequate latency), to name just a few factors complicating interpretation of such outcomes. On the other hand, factors that can produce spurious increases in exposed groups in occupational studies are much less common, as most forces operate to *lower* the observed risks, not raise them[22].

(5) **Understanding the Relative Importance -- and Unimportance -- of Studies that Show a Relative Risk of 2.0 or More, *i.e.*, a "Doubling Dose"**: One concluding statement is needed about how to interpret results of epidemiological studies with respect to causation. One of the most common "measures of effect" used in such studies is something called the Relative Risk (RR), or its close approximation, the Odds Ratio (OR). The RR is the risk in the exposed population divided by the risk in the unexposed population. Thus a RR = 2.0 means that the risk in the exposed population is double the risk in the unexposed.

In my experience as an epidemiologist who does participate in the legal process as an expert, some attorneys maintain and some courts believe that a Relative Risk of two is needed before one can conclude from an epidemiological study that the exposure is "more likely than not" the

---

law, and the courts of application, use inference to reach decisions about what action to take. Those decisions cannot rest on certitudes, most especially when population risks are converted into individual risks." (*Ibid.*, quoting Susser, *op. cit.*, at p. 128 (my italics)).

[22] Of these, the most important are non-differential exposure misclassification and small sample size. See Ozonoff, D. "Assessing the effects of exposure misclassification in hazardous waste studies," paper presented at the annual meeting of the International Society of Environmental Epidemiology, Research Triangle Park, NC, November 1994; also Ozonoff D, et al., "Health Problems Reported by Residents of a Neighborhood Contaminated by a Hazardous Waste Facility." *Amer. J. Ind. Med.* 11:581-597, 1987.

outcome was due to the exposure. The arithmetic basis of this proposition would seem quite transparent,[23] but like many things in this subtle and complex science there are sound and accepted reasons why this argument is not valid. The reasons are both technical and ethical.

(a) Understanding Certain Pertinent Technical Issues Regarding "Relative Risks," "Odds Ratios," and "Attributed Risks"

i. Understanding that the "Relative Risk" (RR) or "Odds Ratio" (OR), as estimated in a particular study, represents a range of likely values, not a single fixed or definite number: The relative risk (RR) or its equivalent (the odds ratio (OR) as an estimate of the relative risk) is itself an estimate from the data of an underlying reality, the "real" risk. RRs or ORs, like other statistics used to summarize data, have some margin of uncertainty associated with the fact that the data are in some sense just one realization of an idealized, very large set of possible realizations, just as the results of flipping a fair coin ten times varies from one realization (set of ten flips) to the next. Thus the RR or OR has a "confidence interval" around it that expresses how "stable" the estimate is in repeated trials. A 95% confidence interval is the range of numbers that would include the "real" risk 95 times out of 100 if the same study were to be done over and over again, allowing for random fluctuations of the data inherent in the selection of subjects. Thus, a relative risk of 1.8 (less than two) with a confidence interval of 1.3 to 2.9 could very likely represent a true RR of greater than two (as high as 2.9 in 95 out of 100 repeated trials).

ii. Understanding that the RR or OR, as estimated in a particular study, is an average value summarizing a broad range of attributes of underlying study subjects: A RR = 1.9 is a summary of the overall risk to a population that is usually heterogeneous with respect to important risk factors. Thus it might include smokers, alcoholics, people who are obese, the elderly, persons who work in hazardous occupations, and persons who have other life conditions or life styles that may affect the risks of or susceptibility to particular toxins or diseases. If it turns out that a particular individual plaintiff with a disease has few or none of these risk factors, than a RR = 1.9 is a serious underestimate of the effects of his or her exposure, as age, smoking, weight, etc., did not

---

[23] If one thousand cases appear "naturally" and another 1000 are due to exposure (the result of a RR = 2.0), then of every 2000 cases only 1000 or 50% would seem to be a result of the exposure.

contribute to the development of the disease and should not be used to discount the risk from exposure, as is done in an epidemiological study when these factors are "adjusted" for.

iii. **Understanding that the calculation of the "attributable risk" that is the supposed foundation for the RR = 2.0 criterion depends in a basic way on a (usually unstated) model of causation:** This point has been made repeatedly in the literature, accompanied with graphic examples of how a study that produces a RR less than 2.0 could result from an exposure in which all of the cases, some of the cases, or none of the cases were the result of exposure. Without a specification of the underlying causation model (which is in almost all cases unknown in sufficient detail to allow an accurate calculation, or even any calculation, of the fraction of cases due to exposure), the doubling of the RR or OR is useless as a criterion for evidentiary admissibility. The fact that it is sometimes used for this purpose has been described in the scientific literature as "a methodologic error that has become a social problem."[24]

SUMMARY OF THE MEANS BY WHICH THE INTERNAL VALIDITY OF STUDIES IS EVALUATED: Evaluating internal validity requires the assessment of the roles played by bias, chance, and real effect. Each can operate, sometimes reinforcing other factors, sometimes offsetting them. There is often disagreement among experts, stemming from differing weights each places on the influence of bias, chance and real effect. Such differences in science are common, both in and out of court. The fact that two scientists have different judgments about how much weight to give a study does not demonstrate that either has failed to use scientifically acceptable reasoning, but only that the ultimate opinion about the weight to accord a study is inherently part of the subjective judgment process of scientists.

An evaluation of internal validity helps a scientist in deciding how much to rely on the specific results of a particular experiment or study. It does *not* tell a scientist how much to extend that result to contexts or situations different than the one studied in the particular study, *i.e.,* how much to *generalize* the result. A separate evaluation for *external* validity is needed.

---

[24] For the latest such comment, see Greenland S, "Relation of probability of causation to relative risk and doubling dose: A methodologic error that has become a social problem," American Journal of Public Health 89:1166-1169, 1999.

b. External validity: Can valid, reliable, and useful generalizations be drawn from the results of a particular study?

Scientists observe and experiment in order to generalize, that is, to explain as much of the world as possible. Generalization is the source of science's fascination, power of explanation, and practical importance in the world outside the community of scientists. The limits and extent of the generalization that can be drawn from a given study constitute the dimensions by the study's external validity. For present purposes, the question is whether research results and conclusions developed in one context (e.g., a high-dose animal study) can be generalized to cover other contexts (e.g., human exposures and disease).

Because there are no fixed, definite, and generally agreed upon rules about how -- and how far -- to generalize, each study must be evaluated in a specific context. Still, certain generic questions arise frequently, which I illustrate here with a brief example.

In this matter, I express an opinion about the capacity of chlorinated pesticides and other exposures from a Texas pesticide company to affect the health of workers and nearby residents. How does a scientist legitimately assert that such a generalization is valid and reliable? In essence, scientists put forth reasons why such a generalization makes sense, for example, that the animals involved are similar in pertinent respects to humans, followed by an examination of reasons that might limit the generalization, for example, that the high doses used may alter the process sufficiently that it no longer applies to human exposures.[25] Defining and constraining generalizations is an active process for forming opinions about studies. Again, there is ample scope for shades of opinion among experts who devote their professional time, resources, and best efforts to these areas of inquiry.

---

[25] It should be noted here that high dose animal studies *are* generally accepted by scientists and regulators. Cf. , for example, Huff, et al.., "Carcinogenesis studies: Results of 398 experiments in 104 chemicals from the US National Toxicology Program," *Ann NY Acad Sci* 534:1-30, 1988. Cf. also, Reference Guide on Toxicology, pp. 190- 191.

DEVELOPING AN OPINION ABOUT CAUSATION

1. Arriving at an Explanation: Assembling the Picture

Clinical observations and case reports, epidemiological and animal studies, and toxicological experiments, including those summarized in the Toxicological Profile for Benzene (ATSDR, 1992), are like the pieces of picture puzzle, albeit with the difference that the pieces are being fit into a picture that is being formed in the mind of the scientist on the basis of, and at the same time, that the individual pieces are being discovered and taking shape, and with additional caveats that some existing pieces may not fit (and thus may not be used) and that not all of the pieces that might be needed to fill in the picture are available for placement in the picture when the scientist completes the process, let alone when he or she starts the process. All in all, fitting the pieces into a scientific picture is a fluid, dynamic, and difficult process.

Depending upon a scientist's judgment of the internal validity (or inherent quality) of a particular study, an individual "piece" may be clear and well defined, or fuzzy and indefinite. Depending upon a scientist's judgment of external validity of a particular study, he or she may decide that an individual piece forms a large and central part of the picture, or is just a small piece on the periphery of the picture, or not even part of the picture at all[26]. In addition, a scientist's experience, expertise and basic judgment are involved. The objective for the scientist is to take the available picture pieces, judge their internal and external validity, and assemble a picture (a theory or working diagnosis), that uses the majority of the clear and definite (*i.e.*, internally valid) and the most relevant (*i.e.*, externally valid) pieces into a coherent, sensible, comprehensive, and "elegant" picture of "reality," *i.e.*, a picture that represents his or her decision about "what is happening."[27]

. Thus, a toxicologist studies cancer in the Zymbal gland in the rat and surmises that this is a mechanism whereby benzene produces damage in that species and which may or may not be relevant to other species, while an epidemiologist looks at cancer risks in human populations and concludes that benzene causes cancer in the human species. Each sees a part of the picture.

---

[26] External and internal validity are thus analogous to the "reliability" and "fit" criteria of the *Daubert* Court.

[27] *See* Kuhn, op cit.

As already noted, interpreting a scientific study for use in assembling a coherent picture requires the use of critical thinking to weigh the various factors that might be responsible for the observed association. This includes evaluating the part played by bias, chance, and real effect, together and separately, and judgments on what generalizations are valid. In such a complex process and with practical matters of consequence at stake, it is not surprising that differences of opinion develop. It is also not surprising that such differences are highlighted and, indeed, magnified by the adversary process. But even when so magnified, such disagreements are not merely *artifacts* of the adversary process, but actually essential features of science as it is routinely practiced rather than evidence of flawed scientific reasoning or methodology.

Scientific Evidence that Chlorinated Pesticides Cause non-Hodgkin's Lymphoma

Numerous epidemiologic studies conducted by different investigators over the past two decades have demonstrated increased risk of non-Hodgkin's lymphoma (NHL) in those exposed to a variety of chlorinated pesticides, including DDT, toxaphene, chlordane/heptachlor, dieldrin and lindane. For example, in a study of NHL deaths in Northern Sweden, those exposed to chlorophenols had a significantly increased risk (RR=4.3; 95% C.I.=2.7-6.9) compared to controls (Hardell, et al., 1980). In an early study in Utah, those employed in farming occupations were at increased risk of dying of NHL (Schumacher MC, 1985). In a Kansas study of patients dying of NHL showed a significant increase (RR=6.0; 95% C.I.=1.9-19.5) if they sprayed chlorinated pesticides, primarily 2,4,-D, more than 20 days per year (Hoar, et al., 1986). Later, a study in Washington state (Woods, et al., 1987) showed an increased risk of NHL in forestry herbicide applicators (RR=4.8; 95% C.I.=1.2-19.4), but this study was able to estimate risk for those exposed to DDT, specifically (RR=1.82; 95% C.I.=1.04-3.2).

A mortality study of forest conservationists and soil conservationists in the United States Department of Agriculture found significant excess deaths from NHL (PMR=2.4; 95% C.I.=1.5-3.6). A case-control analysis found an excess risk of NHL among the soil conservationists last

employed after 1960 (employed more recently than others) (OR=2.6; 95% C.I. not reported) (Alavanja et al., 1989).

In a study in Iowa and Minnesota, among 622 men diagnosed with NHL, men who ever farmed were at an increased risk of NHL, compared to nonfarmers (OR=1.2; 95% C.I.=1.0-1.5). Further, elevated risks were found among those who had handled pesticides, including chlordane (OR=1.7; 95% C.I.=1.0-2.9), DDT (OR=1.2; 95% C.I.=0.9-1.7), lindane (OR=1.4; 95% C.I.=1.0-2.1), and toxaphene (OR=0.8; 95% C.I.=0.3-2.0) (Cantor et al., 1992).

A study of Danish gardeners exposed to pesticides found a two-fold increased incidence of NHL (SMbR 200; 95% C.I.= 86-393). Among male gardeners, who have a predominance of direct or indirect exposure to herbicides, the NHL SMbR=173 (95% C.I.=63-376); among female gardeners, who have a predominance of indirect exposure to insecticides (including chlorinated compounds such as DDT, chlordane, and lindane) the NHL SMbR=364 (95% C.I.=44-1,314) (Hansen et al., 1992).

In a re-analysis of an earlier study in Northern Sweden, researchers found an excess of NHL in those exposed to DDT (Hardell, et al., 1994). The univariate odds ratio was 2.4, which was reduced in multivariate analysis taking into account other potential confounding exposures (OR=1.5; 95% C.I.=0.6-3.6).

NHL and chronic lymphocytic leukaemia (CLL) cases in an Italian province were interviewed for information on several factors, including occupational history and use of pesticides. The case-control analysis found no significant association between exposure to insecticides by farmers-animal breeders and NHL and CLL (OR=1.12; 95% C.I.=0.64-1.96). An elevated but not statistically significant risk was found for several pesticides, including DDT (OR=1.70; 95% C.I.=0.91-3.17) (Nanni et al., 1996).

A 1997 state of the art review of pesticides and cancer states that NHL is among the cancers generally found to be excessive among farmers and other occupational groups potentially exposed to pesticides (Zahm et al., 1997). Specific pesticides with reported associations with NHL in epidemiologic studies include the organochlorine insecticides including DDT, chlordane (study cited in the review: Cantor et al., 1992), lindane (study cited in the review: Blair et al., 1998) and toxaphene (study cited in the review: Cantor et al., 1992) have been reported as associated with lymphoma.

A pooled analysis combined three population-based case-control studies examining the risk of NHL among male farmers in four U.S. states. In the pooled analysis, use of DDT resulted in an elevated but non-significant OR of 1.2 (95% C.I.=1.0-1.6) which diminished as adjustments were made (e.g., for age, use of other pesticides [including lindane]). Farmers with more frequent use of DDT (more than 5 days per year) had an increased risk of NHL (OR=2.6; 95% C.I.=1.1-5.9) which also diminished after adjusting for other pesticides (reduced to 0.9 and 1.9). The authors concluded that the excess risk may be explained by the use of other pesticides (Baris et al., 1998).

The same data used in the Baris et al. study was further examined in a detailed review of the risk of NHL among those farmers exposed to lindane (Blair et al., 1998). Farmers who reported use of lindane had an increased risk of NHL as compared to those farmers who did not report using lindane (OR=1.5; 95% C.I.=1.1-2.0). The excess risk was largely confined to those who first used the pesticide 20 years or more prior to the interview (as compared with those who used the pesticide more recently) (OR=1.7; 95% C.I.=1.1-2.5 compared to OR=1.3; 95% C.I.=0.7-2.3). Risk was higher among those with more frequent use (OR=2.0; 95% C.I.=0.6-6.4 for use 5 or more days per year compared to OR=1.6; 95% C.I.=0.6-4.0 for fewer than five days

per year). Some ORs were lower, however (e.g., for data from direct respondents versus proxy respondents), and the authors conclude that lindane may have a small etiologic role in the development of NHL.

In a population-based case-control study of Swedish males diagnosed with NHL, exposure to insecticides, including DDT and pyrethrins, increased the risk of NHL (OR=1.2; 95% C.I.=0.8-1.7), although not statistically significantly (Hardell and Eriksson 1999).

A recent review article summarizing evidence on the epidemiology of lymphomas notes that some, but not all studies have shown an increased risk for NHL following exposure to pesticides, including organochlorine insecticides (positive association seen in Blair et al., 1998 and Cantor et al., 1992; modest relationship between insecticides and NHL found by Hardell and Eriksson 1999). The authors note that inconsistencies in the results may be due to methodological challenges of studies of pesticide exposure and NHL, especially misclassification of exposure through evaluating general classes of pesticides rather than specific pesticides (Baris and Zahm 2000).

More recently, a Canadian study of NHL patients was able to show increased risk of NHL in those exposed to specific insecticides such as aldrin (OR=4.19; 95% C.I.=1.48-11.96), lindane (OR=2.06; 95% C.I.=1.01-4.22) and DDT (OR=1.73; 95% C.I.=1.08-2.76) after controlling for various potential confounding factors (McDuffie, et al., 2001). Similarly, a study in Iowa and Minnesota (Schroeder, et al., 2001) focusing on NHL patients with a specific type of genetic damage showed significantly increased risk in those exposed to chlorinated pesticides including dieldrin (OR=3.7; 95% C.I.=1.9-7.0), toxaphene (OR=3.0; 95% C.I.=1.5-6.1), lindane (OR=2.3; 95% C.I.=1.3-3.9) and elevated but not statistically significant increased risk in those exposed to aldrin and chlordane.

Pooled data from three Midwestern U.S. NHL case-control studies of farmers were analyzed to estimate the effect of specific pesticides and certain pesticide combinations (adjusted for the use of other pesticides) on NHL incidence. Several individual pesticides were associated with increased NHL incidence, including the insecticides chlordane (OR=1.3; 95% C.I.=0.8-2.1), dieldrin (OR=1.4; 95% C.I.=0.8-2.6), lindane (OR=1.2; 95% C.I.=0.8-1.9), and toxaphene (OR=1.1; 95% C.I.=0.6-2.0) (De Roos et al., 2003).

In a study of organochlorine pesticides in those who died of NHL and whose tissue was collected in a national adipose tissue survey (Quintana, et al., 2004), there was an increased risk and a significant dose-response relationship in those exposed to DDT (p for trend=0.04), dieldrin (p for trend=0.0002), oxychlordane (p for trend=0.0002), and heptachlor epoxide (p for trend=0.0001) among other findings.

These and additional studies are summarized in Table 1.

Author, Year	**Exposure/Chemical**	**Study Location**	**RR or OR, CI**
Woods et al., 1987	Forestry herbicide application	Washington (state)	RR= 4.8; 95% C.I.=1.2-19.4
	DDT		RR= 1.82; 95% C.I.=1.04-3.2
Hardell et al., 1980	Chlorophenols	Northern Sweden	RR=4.3; 95% C.I.=2.7-6.9
Schumacher, 1985	Farming occupation	Utah	OR=1.3; 90% C.I.= 0.9-2.3
Cantor et al., 1992	Ever farmed	Iowa and Minnesota	OR=1.2; 95% C.I.= 1.0-1.5
	Chlordane		OR=1.7; 95% C.I.=1.0-2.9
	DDT		OR=1.2; 95% C.I.=0.9-

				1.7
		Lindane		OR=1.4; 95% C.I.=1.0 2.1
		Toxaphene		OR=0.8; 95% C.I.=0.3-2.0
	Alavanja et al., 1989	Soil conservationist occupation, employed after 1960	United States	OR=2.6; (p<.05) 95% C.I. not reported
	Hansen et al., 1992	Gardening occupation, exposure to pesticides	Denmark	SMbR 200; 95% C.I.= 86-393
		Female gardeners: indirect exposure to insecticides (including chlorinated compounds such as DDT, chlordane, and lindane)		SMbR=364; 95% C.I.=44-1,314
	Sathiakumar et al., 1992	Agricultural chemical manufacturing plant	Alabama	SMR= 203; 95% C.I.=55-520
	Blair et al., 1993	Farming occupation	23 US states	White male farmers PCMR 1.21; 95% C.I.=1.13-1.3
				White male farmers, central region (PCMR 1.32; 95% C.I. not reported
	Hardell, et al., 1994	Herbicides, DDT, solvents	Northern Sweden	OR=1.5; 95% C.I.=0.6-3.6
	Nanni et al., 1996	Insecticide exposure	Italy	exposure matrices [more power] OR=1.12; 95% C.I.=0.64-1.96
		DDT		exposure matrices OR=1.70; 95% C.I.=0.91-3.17
	Quintana et al., 2004	DDT, dieldrin, oxychlordane, heptachlor epoxide	United States	Significant dose-response relationships, p<.05
	Baris et al., 1998	Agricultural pesticides	Nebraska, Iowa,	OR=1.2; 95% C.I.=1.0-1.6; diminished with

		Minnesota, Kansas	adjustments for other variables including other pesticides
Blair et al., 1998	Lindane	Nebraska, Iowa, Minnesota, Kansas	OR=1.5; 95% C.I.=1.1-2.0
	First use >20 years prior to interview		OR=1.7; 95% C.I.=1.1-2.5
	Use 5 or more days per year		OR=2.0; 95% C.I.=0.6-6.4
Hardell and Eriksson 1999	Insecticide exposure	Sweden	OR=1.2; 95% C.I.=0.8-1.7
Schroeder et al., 2001	Dieldrin	Iowa and Minnesota	OR=3.7; 95% C.I.=1.9-7.0
De Roos et al., 2003	Chlordane	Nebraska, Iowa, Minnesota, Kansas	OR=1.3; 95% C.I.=0.8-2.1
	Dieldrin		OR=1.4; 95% C.I.=0.8-2.6

Based on this review of the literature, it is my opinion, to a reasonable degree of scientific certainty, that exposure to chlorinated pesticides and specific chemicals chlordane, dieldrin, lindane, toxaphene and DDT cause or contribute to the development of non-Hodgkin's lymphoma in exposed humans.

Signed on February ___, 2005, at Boston, MA.

[redacted]Clapp, D.Sc., MPH

References Cited

Alavanja MC, Blair A, Merkle S, Teske J, Eaton B, Reed B. Mortality among forest and soil conservationists. Arch Environ Health. 1989 Mar-Apr;44(2):94-101.

Baris D, Zahm SH, Cantor KP, Blair A. Agricultural use of DDT and risk of non-Hodgkin's lymphoma: pooled analysis of three case-control studies in the United States. Occup Environ Med. 1998 Aug;55(8):522-7.

Baris D, Zahm SH. Epidemiology of lymphomas. Curr Opin Oncol. 2000 Sep;12(5):383-94.

Blair A, Dosemeci M, Heineman EF. Cancer and other causes of death among male and female farmers from twenty-three states. Am J Ind Med. 1993 May;23(5):729-42.

Blair A, Cantor KP, Zahm SH. Non-hodgkin's lymphoma and agricultural use of the insecticide lindane. Am J Ind Med. 1998 Jan;33(1):82-7.

Cantor KP, Blair A, Everett G, Gibson R, Burmeister LF, Brown LM, Schuman L, Dick FR. Pesticides and other agricultural risk factors for non-Hodgkin's lymphoma among men in Iowa and Minnesota. Cancer Res. 1992 May 1;52(9):2447-55.

De Roos AJ, Zahm SH, Cantor KP, Weisenburger DD, Holmes FF, Burmeister LF, Blair A. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. Occup Environ Med. 2003 Sep;60(9):E11.

Hansen ES, Hasle H, Lander F. A cohort study on cancer incidence among Danish gardeners. Am J Ind Med. 1992;21(5):651-60.

Hardell L, Eriksson M, Lenner P. [Malignant lymphoma and exposure to chemical substances, in particular organic solvents, chlorphenol and phenoxyacetates] [Article in Swedish] Lakartidningen. 1980 Jan 23;77(4):208-10.

Hardell L, Eriksson M, Degerman A. Exposure to phenoxyacetic acids, chlorophenols, or organic solvents in relation to histopathology, stage and anatomical location of non-Hodgkin's lymphoma. Cancer Research. 1994;14:2386-2389.

Hardell L, Eriksson M. A case-control study of non-Hodgkin lymphoma and exposure to pesticides. Cancer. 1999 Mar 15;85(6):1353-60. Comment in: Cancer. 1999 Aug 15;86(4):729-31.

McDuffie HH, Pahwa P, McLaughlin JR, Spinelli JJ, Fincham S, Dosman JA, Robson D, Skinnider LF, Choi NW. Non-Hodgkin's lymphoma and specific pesticide exposures in men: cross-Canada study of pesticides and health. Cancer Epidemiol Biomarkers Prev. 2001 Nov;10(11):1155-63.

Nanni O, Amadori D, Lugaresi C, Falcini F, Scarpi E, Saragoni A, Buiatti E. Chronic lymphocytic leukaemias and non-Hodgkin's lymphomas by histological type in farming-animal breeding workers: a population case-control study based on a priori exposure matrices. Occup Environ Med. 1996 Oct;53(10):652-7.

Quintana PJ, Delfino RJ, Korrick S, Ziogas A, Kutz FW, Jones EL, Laden F, Garshick E. Adipose tissue levels of organochlorine pesticides and polychlorinated biphenyls and risk of non-Hodgkin's lymphoma. Environ Health Perspect. 2004 Jun;112(8):854-61.

Sathiakumar N, Delzell E, Austin H, Cole P. A follow-up study of agricultural chemical production workers. Am J Ind Med. 1992;21(3):321-30.

Schroeder JC, Olshan AF, Baric R, Dent GA, Weinberg CR, Yount B, Cerhan JR, Lynch CF, Schuman LM, Tolbert PE, Rothman N, Cantor KP, Blair A. Agricultural risk factors for t(14;18) subtypes of non-Hodgkin's lymphoma. Epidemiology. 2001 Nov;12(6):701-9.

Schumacher MC. Farming occupations and mortality from non-Hodgkin's lymphoma in Utah. A case-control study. J Occup Med. 1985 Aug;27(8):580-4.

Zahm SH, Ward MH, Blair A. Pesticides and cancer. Occup Med. 1997 Apr-Jun;12(2):269-89.

Woods JS, Polissar L, Severson RK, Heuser LS, Kulander BG. Soft tissue sarcoma and non-Hodgkin's lymphoma in relation to phenoxyherbicide and chlorinated phenol exposure in western Washington. J Natl Cancer Inst. 1987 May;78(5):899-910.

Zahm SH, Ward MH, Blair A. Pesticides and cancer. Occup Med. 1997 Apr-Jun;12(2):269-89.