R. Brent Wisner, Esq. (SBN: 276023)
rbwisner@baumhedlundlaw.com
Michael L. Baum, Esq. (SBN: 119511)
mbaum@baumhedlundlaw.com
BAUM, HEDLUND, ARISTEI, & GOLDMAN, P.C.
12100 Wilshire Blvd., Suite 950
Los Angeles, CA 90025
Telephone: (310) 207-3233
Facsimile: (310) 820-7444

Attorneys for Plaintiffs
(Additional attorneys on signature page)

UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE: ROUNDUP PRODUCTS LIABILITY LITIGATION THIS DOCUMENT RELATES TO: ALL ACTIONS	MDL No. 2741 **PLAINTIFFS' POST-ARGUMENT SUBMISSION** HON. VINCE CHHABRIA

Pursuant to the Court's instruction today during oral argument, the Plaintiffs submit the following literature relevant to the general causation inquiry:

1. Kenneth Rothman, *Six Persistent Research Misconceptions*, 7 J GEN INTERN MED. 29, 1060-65 (Jul. 2014), Exh. 79;
2. Aaron Blair, et al, *Methodological Issues Regarding Confounding and Exposure Misclassification in Epidemiological Studies of Occupational Exposures*, 50 AM. J. IND. MED. 199-206 (2007), Exh. 31;
3. Krista Christensen, et al, *The Use Epidemiology in Risk Assessment: Challenges and Opportunities*, 21 Human and Ecological Risk Assess., 1644-1663 (2015);
4. Michael Green, et al, *Reference Guide on Epidemiology*, Reference Manual on Scientific Evidence, Federal Judicial Center, 3rd Ed. (2011);
5. Lorelei A. Mucci, *Maternal Smoking and Childhood Leukemia and Lymphoma Risk among*

1,440,542 Swedish Children, 13 CANCER EPID., BIOMARKERS & PREVENTION 9, 1528-1533 (2004);

6. [DRAFT] Manisha Pahwa, et al, *An evaluation of glyphosate use and the risk of non-Hodgkin lymphoma major histological sub-types in the North American Pooled Project (NAPP)* – unpublished draft, Exh. 106;
7. George M. Gray, et al, *The Federal Government's Agricultural Health Study: A Critical Review with Suggested Improvements*, 6 HUMAN AND ECOLOGICAL RISK ASSESSMENT 1, 47-71 (2000), Exh. 13;
8. Aaron Blair, et al, *Reliability of Reporting on Life-Style and Agricultural Factors by a Sample of Participants in the Agricultural Health Study from Iowa*, 13 EPIDEMIOLOGY 1, 94-99 (2002), Exh. 30;
9. Scott Weichenthal, et al, *A Review of Pesticide Exposure and Cancer Incidence in the Agricultural Health Study Cohort*, 118 ENVIRONMENTAL HEALTH PERSPECTIVES 8, 1117-25 (2010), Exh. 35;
10. Christopher Portier, et al (94 authors), *Differences in the carcinogenic evaluation of glyphosate between the International Agency for Research on Cancer (IARC) and the European Food Safety Authority (EFSA)*, 0 EPIDEMIOLOGY COMMUNITY HEALTH MONTH 0, 1-4 (2016), Exh. 78; and
11. Neil Pearce, et al (110 authors), *IARC Monographs: 40 Years of Evaluating Carcinogenic Hazards to Humans*, 123 ENVIRONMENTAL HEALTH PERSPECTIVES 6, 507-14 (2015), Exh. 116.

DATED: March 14, 2018

Respectfully submitted,

By: /s/ R. Brent Wisner

R. Brent Wisner, Esq. (SBN: 276023)
rbwisner@baumhedlundlaw.com
Michael L. Baum, Esq. (SBN: 119511)
mbaum@baumhedlundlaw.com
BAUM, HEDLUND, ARISTEI, & GOLDMAN, P.C.
12100 Wilshire Blvd., Suite 950
Los Angeles, CA 90025
Telephone: (310) 207-3233
Facsimile: (310) 820-7444

Aimee Wagstaff
aimee.wagstaff@andruswagstaff.com
ANDRUS WAGSTAFF, P.C.
7171 West Alaska Drive
Lakewood CO 80226
Ph 303-376-6360
F 303-376-6361

Robin Greenwald
rgreenwald@weitzlux.com
WEITZ & LUXENBERG, P.C.
700 Broadway
New York NY 10003
Telephone: (212) 558-5500
Facsimile: (212) 344-5461

Michael Miller
mmiller@millerfirmllc.com
THE MILLER FIRM, LLC
108 Railroad Ave
Orange VA 22960
Telephone: (540) 672 4224
Facsimile: (540) 672-3055

Attorneys for Plaintiffs

CERTIFICATE OF SERVICE

I, R. Brent Wisner, hereby certify that, on March 14, 2018, I electronically filed the foregoing with the Clerk for the United States District Court for the Northern District of California using the CM/ECF system, which shall send electronic notification to counsel of record.

/s/ R. Brent Wisner
R. Brent Wisner

REVIEWS

Six Persistent Research Misconceptions

Kenneth J. Rothman, DrPH[1,2]

[1]Research Triangle Institute, Research Triangle Park, NC, USA; [2]Boston University School of Public Health, Boston, MA, USA.

Scientific knowledge changes rapidly, but the concepts and methods of the conduct of research change more slowly. To stimulate discussion of outmoded thinking regarding the conduct of research, I list six misconceptions about research that persist long after their flaws have become apparent. The misconceptions are: 1) There is a hierarchy of study designs; randomized trials provide the greatest validity, followed by cohort studies, with case–control studies being least reliable. 2) An essential element for valid generalization is that the study subjects constitute a representative sample of a target population. 3) If a term that denotes the product of two factors in a regression model is not statistically significant, then there is no biologic interaction between those factors. 4) When categorizing a continuous variable, a reasonable scheme for choosing category cut-points is to use percentile-defined boundaries, such as quartiles or quintiles of the distribution. 5) One should always report P values or confidence intervals that have been adjusted for multiple comparisons. 6) Significance testing is useful and important for the interpretation of data. These misconceptions have been perpetuated in journals, classrooms and textbooks. They persist because they represent intellectual shortcuts that avoid more thoughtful approaches to research problems. I hope that calling attention to these misconceptions will spark the debates needed to shelve these outmoded ideas for good.

KEY WORDS: study design; data interpretation; epidemiologic methods; representativeness; evaluation of interaction; multiple comparisons; percentile boundaries; statistical significance testing.
J Gen Intern Med 29(7):1060–4
DOI: 10.1007/s11606-013-2755-z

A surprising number of misconceptions persist in the conduct of research involving human subjects. Some persist despite teachings to the contrary, and some because of teachings that should be to the contrary. To spark discussion of these issues, I list here six persistent research misconceptions, and offer a capsule summary of the problems with each of them.

Received November 01, 2013
Revised November 27, 2013
Accepted December 18, 2013
Published online January 23, 2014

Misconception 1. *There is a hierarchy of study designs; randomized trials provide the greatest validity, followed by cohort studies, with case–control studies being least reliable.*

Randomized trials, though often considered the "gold standard" of study types, are not perfect, even in concept. Furthermore, the premise that the comparative validity of study results can be inferred from the type of study is wrong.

Although some believe that evidence from a randomized trial is as compelling as a logical proof, no empirical finding can provide absolute certainty. If randomized trials were perfect, how could they give divergent results? In fact, they are subject to various errors.[1] Obviously there is random error, as one would expect from a study based on random assignment. But there is also systematic error, or bias. For example, randomized trials are usually analyzed using the "intent to treat" principle, which compares the groups that are initially assigned by randomization, regardless of any subsequent non-adherence. Non-adherence results in underestimation of any treatment effect. This bias is usually considered acceptable because it is outweighed by the advantages achieved by random assignment. Underestimation of effects, however, is not acceptable in a safety trial aimed at uncovering adverse effects of the treatment. Another important source of bias in a randomized trial comes from errors in assessing the outcome, such as undercounting of outcome events. Also, even if randomization provides a balance of risk factors between groups at the start of the trial, with extended follow-up, the study groups may become progressively imbalanced through differential attrition or changes in risk factor distributions. With long-term trials, the benefits of random assignment may therefore fade with time.

In short, trials are far from perfect. Furthermore, both cohort and case–control studies will yield valid results when properly designed and carried out. Therefore, mindlessly ascribing greater validity to a study based on a hierarchy of designs[2,3] is fallacious. For example, the relation between cigarette smoking and lung cancer is well established, based on findings from cohort and case–control studies. The connection was never shown clearly in a randomized trial. It is not easy to assign people randomly to smoke or not smoke; however, when smoking cessation was studied as part of a multi-pronged intervention in the randomized Multiple Risk Factor Intervention Trial,[4] those who were

urged to cease smoking actually developed more lung cancer than those who did not receive the cessation encouragement. The results of the trial did not overthrow the findings of the many cohort and case–control studies conducted without randomization. Rather, the discrepancy was ascribed to problems with the trial.

In another high-profile example, results from large cohort studies[5,6] indicated that risk of coronary heart disease was reduced among postmenopausal hormone users, but later results from two randomized trials indicated either no association or an increased risk.[7,8] The reaction in the scientific community and the popular press[9] was to discredit the results from the cohort studies, presuming that they had been refuted by the randomized trials. Many continue to believe that interpretation, but in an elegant reanalysis, Hernan et al.[10] showed that the study populations in the cohort studies and the randomized trials were different, and that the effects of postmenopausal hormone use varied greatly according to age and time since menopause. When studies were restricted to new users of hormones, Hernan et al. showed that differences in the distribution of age and time since menopause could explain all of the apparent discrepancies. Although it is common to ascribe such discrepancies to inherent weaknesses of the nonexperimental studies, it is simplistic to assign validity based on a presumed hierarchy of study types.[11]

Similarly, discrepancies between cohort studies and case–control studies should not be explained away superficially by a presumed validity advantage for cohort studies over case–control studies. Properly designed case–control studies will produce the same results as properly designed cohort studies. When conflicts arise, they could stem from problems in either or both types of study. Although case–control studies have long been disparaged as being backwards versions of cohort studies, starting from disease and tracing back to possible causes, epidemiologists today understand case–control studies to be conceptually identical to cohort studies, apart from an efficiency gain that comes from sampling the denominators rather than conducting a complete census. Indeed, the efficiency gain may allow more resources for exposure assessment or case validation in case–control studies, resulting in less bias than in corresponding cohort studies of the same relation.

Those who view case–control studies as backwards versions of cohort studies sometimes make the false analogy that the controls should closely resemble the cases, except that they lack the case-defining disease. In fact, the control group in a case–control study is intended to be a sample of the population denominator that gives rise to the cases, a substitute for the full denominators obtained in a cohort study. Thus, the control group should resemble the entire study population, rather than the cases.[12,13] When properly designed, case–control studies can achieve the same excellent validity as properly designed cohort studies, whereas a poorly designed trial can be unreliable. The type of study should not be taken as a guide to a study's validity.

Misconception 2. *An essential element of making valid generalizations from a study is that the study subjects constitute a representative sample of a target population.*

This misconception is tied to the view that scientific generalization involves the mechanical extrapolation of results from a sample to its source population. But that describes statistical generalization; scientific generalization is different: it is the process of constructing a correct statement about the way nature works.

Scientific generalization is the ultimate goal of scientific inquiry, but a prerequisite is designing a study that has internal validity, which is enhanced by keeping all disturbing variables constant. When have we heard of animal researchers who seek a statistically representative sample of animals? Instead, their operating principle is nearly the opposite of seeking representativeness. Thus, biologists studying mice prefer to study mice that are homogeneous with respect to genes and environment, and that differ only in respect to the experimentally manipulated variable. Unlike the statistical generalization of opinion polls or survey sampling, which merely calls for extrapolation from sample to source population, scientific generalization proceeds by informed guesses, but only from the secure platform of a valid study. Consequently, studies are stronger if they limit variability of confounding factors, as opposed to seeking representativeness. Doll and Hill[14] studied the mortality of male British physicians in relation to their smoking habits. Their findings were considered broadly generalizable despite the fact that their study population was unrepresentative of the general population of tobacco users with regard to sex, race, ethnicity, social class, nationality and many other variables.

When there is a legitimate question about whether an overall association varies by subgroup of some third variable, such as age or ethnic group, it may be necessary to include people drawn from a broad range of values of that third variable, but even then it is counterproductive for the study population to be representative of the source population for that variable. The goal in that case would be to include study subjects distributed evenly across the range, or in a distribution that enhances overall study efficiency. A sample that is representative of the source population will be suboptimal.[15,16]

Misconception 3. *If a term that denotes the product of two factors in a regression model is not statistically significant, then there is no biologic interaction between those factors.*

"Biologic" is meant here broadly, to encompass biochemical, psychological, behavioral and physical interactions. The

problem is that interaction is usually evaluated through regression models, in which the product term addresses statistical interaction rather than biologic interaction.

Biologic interaction refers to two or more causes acting in the same mechanism, with effects that are mutually dependent. It describes a state of nature. If basic effects are measured as changes in disease risk, synergistic (i.e. positive) biologic interaction is present when the joint effect of two causal factors is more than the sum of their effects acting separately.[17] In contrast, statistical interaction does not describe nature; it describes a mathematical model. It is typically assessed with a product term for two variables in a regression model. Its magnitude depends on the choice of measures and scale of measurement. Statistical interaction implies only that the basic functional form of a specific mathematical model is not an apt description of the relation among variables. Two factors that show biologic interaction may or may not exhibit statistical interaction, depending on the model used.

Product terms in regression models have units that can defy interpretation. If one variable is fat consumption, measured in grams per day, and another variable is pack-years of cigarettes smoked, what is the interpretation of a variable that has units of grams/day multiplied by pack-years? The challenge of interpreting such product term coefficients has fostered a focus on the p value accompanying the coefficient, rather than the magnitude of the coefficient itself. Focusing on the pvalue, or on whether the coefficient of a product term is statistically significant, only worsens the problem of mistaking statistical interaction for biologic interaction (see misconception 6). A more meaningful assessment of interaction would be to focus on the proportion of cases of a disease that one could attribute to biologic interaction.[17,18]

Consider a simple example from the TREAT trial (Trial to Reduce Cardiovascular Events with Aranesp Therapy),[19] which evaluated the risk of stroke among 4,038 patients with diabetes mellitus, chronic kidney disease, and anemia randomized to receive darbepoetin alfa or placebo. Among patients without a history of stroke, the risk of stroke during the study period was 2 % among patients receiving placebo and 4 % among patients receiving darbepoeitin alfa. Among patients with a history of stroke, the corresponding risks were 4 % and 12 %. The authors noted that the risk increase was greater for darbepoeitin alfa among those with a history of stroke, but they dismissed this interaction because the product term in a logistic regression model was not statistically significant. The increased risk attributable to darbepoeitin alfa was 2 % in the patients without a history of stroke and 8 % among patients with a history of stroke, indicating strong biologic interaction between darbepoeitin alfa and history of stroke. If the risks were merely additive, the risk would be 6 % among those with both risk factors, instead of the actual 12 %. Thus, half of the risk among those with both risk factors appears attributable to biologic interaction, despite the authors' claim that there was no interaction.

Misconception 4. *When categorizing a continuous variable, a reasonable scheme for choosing category cut-points is to use percentile-defined boundaries, such as quartiles or quintiles of the distribution.*

There are two reasons why using percentiles is a poor method for choosing category boundaries. First, these boundaries may not correspond to the parts of the distribution where biologically meaningful changes occur. Suppose you were conducting a study of vitamin C intake and scurvy risk in the U.S. If you decided to categorize vitamin C intake by quintiles, you would find that the entire relation between vitamin C consumption and scurvy was confined to the lowest quintile, and within that category, to only a small proportion of people who were outliers in their low vitamin C intake. 10 mg/day of vitamin C can prevent scurvy, but those consuming less than that represent a fraction of 1 % of the population in the U. S.[20] Using percentile-based categories would make it impossible to find the effect of inadequate vitamin C intake on scurvy risk, because all intake above 10 mg/d is essentially equivalent. If we routinely use percentile cut-points, we may not know if we are facing the same problem as we would face in the study of vitamin C and scurvy. A more effective alternative would be to begin with many narrow categories, merging neighboring categories until meaningful breaks in risk become evident.

The second problem with percentile-based categories is the difficulty in comparing results across studies, because categories across studies using percentile category boundaries are unlikely to correspond. This problem can be averted by expressing boundary points in terms of the natural units of the variable (such as mg/d for vitamin C intake). It is also useful to report within-category means or medians.

Misconception 5. *One should always report P values or confidence intervals that have been adjusted for multiple comparisons.*

Traditional adjustments for multiple comparisons involve inflating the P value or the width of a confidence interval according to the number of comparisons conducted. If one is analyzing biological data that are replete with actual associations, the premise for traditional adjustments is shaky and the adjustments are difficult to defend. The concern for multiple comparisons stems from fear of finding falsely significant findings (type I errors in the lingo of statistics). In misconception 6, we discuss the problems with using statistical significance testing for data analysis in the first place. But before considering those problems, let us consider the rationale for adjusting reported results for multiple comparisons.

Despite the fact that a single significance test is intended to have a 5 % probability (at the conventionally used level) of being significant when the null hypothesis is true, and

therefore multiple tests when properly carried out should each have this property, there is a concern that when making multiple tests, the probability of a spurious result is increased. Of course, as the number of tests increases, the probability that one or more of them would be falsely positive increases, but that is only because many tests are being conducted. Adjustments for multiple comparisons will reduce these type I errors, but they do so at the expense of increasing type II errors, which are nonsignificant test results in the presence of a real association. When observed associations are all the result of chance, type I errors can occur, but type II errors cannot occur. Conversely, when the observed associations all reflect actual relationships, type II errors can occur, but type I errors cannot. Thus, the context of any analysis has fundamental implications regarding the interpretation of the data. In particular, it is absurd to make adjustments that reduce type I errors at the expense of increasing type II errors without some evaluation of the estimated relative cost and frequency of each type of error.

If scientists were put to work studying random numbers instead of biologic data, all the significant results they reported would represent type I errors, and adjustments for multiple comparisons would make sense; some skeptics believe that studies of genome-wide association scans may approximate this situation.[21] But when scientists are studying biological relations rather than random numbers, the premise that type I errors are the major concern may be wrong.[22] A more rigorous evaluation of the need for multiplicity adjustments would begin with an assessment of the tenability of the thesis that the data are essentially random numbers. If one is studying experiments on psychic phenomena, skepticism about the results might lend support to multiplicity adjustments. If one is studying physiologic effects of pharmaceutical agents, real associations are to be expected and the adjustments are more difficult to defend. Studying single nucleotide polymorphisms in relation to a given disease might be a middle ground. One approach to this issue that is theoretically more defensible is a Bayesian approach, which assigns prior credibility to various levels of association and adjusts by using Bayes' theorem to calculate posterior credibility.[23,24]

Misconception 6. *Significance testing is useful and important for the interpretation of data.*

Significance testing has led to far more misunderstanding and misinterpretation than clarity in interpreting study results.[25–28] A significance test is a degraded version of the P value, a statistic that blends precision with effect size, thus confusing two essential aspects of data interpretation. Measuring effect size and its precision as separate tasks is a more direct and clearer approach to data interpretation.

For research studies that aim to measure associations, and infer whether they reflect causal connections, focusing on the magnitude of these associations ought to be the primary goal: estimation of effects is decidedly preferable to statistical testing. Ideally, a study estimates the magnitude of the effect size, and analyzes the possible errors that might have distorted it. Systematic errors such as confounding from measured factors can be dealt with through analytic methods; other systematic errors, such as the effects of measurement error or selection bias, can be addressed through sensitivity analyses (also known as bias analysis). Random error is typically expressed through confidence intervals, giving a range of parameter values that are consistent with the data to a specified level.

It is unfortunate that a confidence interval, from which both an estimate of effect size and its measurement precision can be drawn, is typically used merely to judge whether it contains the null value or not, thus converting it to a significance test. Significance tests are a poor classification scheme for study results; strong effects may be incorrectly interpreted as null findings because authors fallaciously interpret lack of statistical significance to imply lack of effect, or weak effects may be incorrectly interpreted as important because they are statistically significant. Rather than be used as surrogate significance tests, confidence intervals ought to be interpreted as quantitative measures indicating magnitude of effect size and degree of precision, with little attention paid to the precise location of the boundaries of the confidence interval. This advice is backed by the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, but nevertheless often overlooked even by reviewers and editors whose journals support the requirements.[29]

Many misconceptions derive from reliance on statistical significance testing. The focus on the statistical significance of interaction terms instead of measuring interaction, as discussed above, is one example. The evaluation of dose–response trends simply by declaring that there is or is not a significant trend, rather than expressing the magnitude and ideally the shape of that trend, is another. Yet another is the advice sometimes offered to calculate the power of a study when reporting results, especially if those results are not statistically significant. Reporting the power of a study as part of the results is called "post-hoc" power calculation.[30] Power calculations are based on a hypothesis about the level of association that is to be distinguished from a null association, but when the study results are on hand, there is no longer any need to hypothesize about the magnitude of the association, because you now have an estimate of it. A confidence interval for the estimated association conveys all the relevant information; nothing further is to be gained from a power calculation.

The unfortunate consequence of the focus on statistical significance testing has been to foster a dichotomous view of relationships that are better assessed in quantitative terms. This distinction is more than a nicety. Every day there are important, regrettable and avoidable misinterpretations of data that results from the confusing fog of

statistical significance testing. Most of these errors could be avoided if the focus were shifted from statistical testing to estimation.

CONCLUSION

Why do such important misconceptions about research persist? To a large extent these misconceptions represent substitutes for more thoughtful and difficult tasks. It is simpler to resolve a discrepancy between a trial and a nonexperimental study in favor of the trial, without undertaking the laborious analysis that Hernan et al. did.[10] It is easy to declare that a result is not statistically significant, falsely implying that there is no indication of an association, rather than to consider quantitatively the range of associations that the data actually support. These misconceptions involve taking the low road, but when that road is crowded with others taking the same path, there may be little reason to question the route. Indeed, these misconceptions are often perpetuated in journals, classrooms and textbooks. I believe that the best prospect for improvement is to raise consciousness about the issues, with reasoned debate. Max Planck once said, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."[31] To the extent that this cynical view is correct, we can expect to see outmoded concepts fade away slowly at best. I hope that calling attention to these misconceptions will spark the needed debates and be a catalyst for change.

Acknowledgements: *I received helpful criticism from Susana Perez, Andrea Margulis, Manel Pladevall, and Jordi Castellsague.*

Conflict of Interest: *The author declares no conflict of interest.*

Corresponding Author: *Kenneth J. Rothman, DrPH; Research Triangle Institute, Research Triangle Park, NC, USA (e-mail: KRothman@rti.org).*

REFERENCES

1. **Hernán MA, Hernández-Díaz S, Robins JM.** Randomized trials analyzed as observational studies. Ann Intern Med. 2013;159:560–2. doi:10.7326/0003-4819-159-8-201310150-00709
2. **Ioannidis JPA.** Why most published research findings are false. PLoS Med. 2005;2(8):e124.
3. **Hiatt WR.** Observational studies of drug safety–aprotinin and the absence of transparency. N Engl J Med. 2006;355:2171–3.
4. **Shaten BJ, Kuller LH, Kjelsberg MO, Stamler J, Ockene JK, Cutler JA, Cohen JD.** Lung cancer mortality after 16 years in MRFIT participants in intervention and usual-care groups. Multiple Risk Factor Intervention Trial. Ann Epidemiol 1997;7:125–36.
5. **Grodstein F, Manson JE, Colditz GA, et al.** A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. Ann Intern Med. 2000;133:933–41.
6. **Varas-Lorenzo C, García-Rodríguez LA, Pérez-Gutthann S, et al.** Hormone replacement therapy and incidence of acute myocardial infarction. Circulation. 2000;101:2572–8.
7. **Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, Vittinghoff E.** Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA. 1998;280:605–13. doi:10.1001/jama.280.7.605.
8. **Manson JE, Hsia J, Johnson KC, et al.** Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med. 2003;349:523–34.
9. **Taubes G.** Do we really know what makes us healthy? New York Times, September 16, 2007.
10. **Hernán MA, Alonso A, Logan R, Grodstein F, Michels K, Willett WC, Manson JE, Robins JM.** Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19:766–79. doi:10.1097/EDE.0b013e3181875e61
11. **Concato J.** Observational versus experimental studies: what's the evidence for a hierarchy? NeuroRx. 2004;1:341–7.
12. **Vandenbroucke JP, Pearce N.** Case–control studies: basic concepts. Int J Epidemiol. 2012;41:1480–9. doi:10.1093/ije/dys147.
13. **Rothman KJ.** Chapter 5, Types of Epidemiologic Studies, in *Epidemiology, An Introduction*, 2nd Edition. Oxford University Press, New York, 2012.
14. **Doll R, Hill AB.** The mortality of doctors in relation to their smoking habits: a preliminary report. Br Med J. 1954;ii:1451–5.
15. **Rothman KJ, Gallacher J, Hatch EE.** Why representativeness should be avoided. Int J Epidemiol. 2013;42:1012–4. doi:10.1093/ije/dys223
16. **Rothman KJ, Gallacher J, Hatch EE.** When it comes to scientific inference, sometimes a cigar is just a cigar. Int J Epidemiol. 2013;42:1026–8. doi:10.1093/ije/dyt124
17. **Rothman KJ.** Chapter 11, Measuring Interaction, in *Epidemiology, An Introduction*, 2nd Edition. Oxford University Press, New York, 2012.
18. **Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI.** Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol. 2007;36:1111–8.
19. **Skali H, Parving HH, Parfrey PS, Burdmann EA, Lewis EF, Ivanovich P, Keithi-Reddy SR, McGill JB, McMurray JJ, Singh AK, Solomon SD, Uno H, Pfeffer MA.** TREAT Investigators: Stroke in patients with type 2 diabetes mellitus, chronic kidney disease, and anemia treated with Darbepoetin Alfa: the trial to reduce cardiovascular events with Aranesp therapy (TREAT) experience. Circulation. 2011;124:2903–8.
20. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. Institute of Medicine, The National Academies Press, Washington, D. C., 2000.
21. **Dudbridge F, Gusnanto A.** Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008;32:227–34.
22. **Rothman KJ.** No adjustments are needed for multiple comparisons. Epidemiology. 1990;1:43–6.
23. **Greenland S, Robins J.** Empirical-Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology. 1991;2:244–51.
24. **Greenland S, Poole C.** Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health. 1994;48:9–16.
25. **Rothman KJ.** A show of confidence (editorial). N Engl J Med. 1978;299:1362–3.
26. **Poole C.** Beyond the confidence interval. Am J Public Hlth. 1987;77:195–9.
27. **Rothman KJ.** Significance questing (Editorial). Ann Int Med. 1986;105:445–7.
28. **Gelman A, Stern H.** The difference between "Significant" and "Not Significant" is not itself statistically significant. Amer Statistician. 2006;60:328–31.
29. Uniform Requirements for Manuscripts Submitted to Biomedical Journals, http://www.icmje.org/manuscript_1prepare.html (accessed May 2, 2013)
30. **Smith AH, Bates MN.** Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. Epidemiology. 1992;3:449–52.
31. **Planck M.** *Scientific Autobiography and Other Papers*, Philosophical Library, New York, 1968, trans. F. Gaynor (New York, 1949), pp. 33–34

AMERICAN JOURNAL OF INDUSTRIAL MEDICINE 50:199–207 (2007)

Methodological Issues Regarding Confounding and Exposure Misclassification in Epidemiological Studies of Occupational Exposures

Aaron Blair, PhD,[1]* Patricia Stewart, PhD,[1] Jay H. Lubin, PhD,[1] and Francesco Forastiere, MD[2]

Background *Confounding and exposure misclassification are issues that concern epidemiologists because of their potential to bias results of studies and complicate interpretations. In occupational epidemiology both are routinely raised to argue that an observed result is either a false positive or a false negative finding. Although it is important to consider the potential for limitations of epidemiologic investigations, judgment regarding their importance should be based on their actual likelihood of occurrence.*

Methods *This paper is based on our experience in epidemiologic analyses and a brief review of the literature regarding confounding and exposure misclassification.*

Results *Examples of substantial confounding are rare in occupational epidemiology. In fact, even for studies of occupational exposures and lung cancer, tobacco-adjusted relative risks rarely differ appreciably from the unadjusted estimates. This is surprising because it seems the perfect situation for confounding to occur. Yet, despite the lack of evidence that confounding is a common problem, nearly every epidemiologic paper includes a lengthy discussion on uncontrolled or residual confounding. On the other hand, exposure misclassification probably occurs in all studies. The only question is, how much? The direction and magnitude of nondifferential exposure misclassification (the type most likely to occur in cohort studies) on estimates of relative risks can be largely predicted given knowledge on the degree of misclassification, that is, relatively small amounts of misclassification can bias relative risks substantially towards the null. The literature, however, is full of discussions implying that misclassification of exposure is an explanation for a positive finding.*

Conclusions *These comments are not to suggest that all potential limitations for epidemiologic studies should not be considered and evaluated. We do believe, however, that the likelihood of occurrence and the direction and magnitude of the effect should be more carefully and realistically considered when making judgments about study design or data interpretation.* Am. J. Ind. Med. 50:199–207, 2007. © *2006 Wiley-Liss, Inc.*

KEY WORDS: confounding; exposure misclassification; methods; occupational epidemiology

INTRODUCTION

The potential limitations of observational epidemiologic studies are well described in textbooks on epidemiology. These limitations include confounding, selection bias, information bias, and lack of validity and precision of exposure, and disease determinations. Concerns over these

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland
[2]Department of Epidemiology, ASL Roma E, Via Santa Costanza 53, Roma, Italy
Contract grant sponsor: NIH.
*Correspondence to: Aaron Blair, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Executive Plaza South, Room 8118, Bethesda, MD. 20892

Accepted 28 December 2005
DOI 10.1002/ajim.20281. Published online in Wiley InterScience (www.interscience.wiley.com)

limitations are also raised and discussed in most epidemiologic papers. A critical assessment of the strengths and weaknesses of all scientific studies is a crucial component of the scientific method and this process serves to identify false leads, to consider alternative explanations, and to improve study designs. In occupational epidemiology, two limitations that receive considerable attention are confounding and exposure misclassification. Theoretical issues regarding these concepts have been well thought out, can be found in most epidemiologic texts [e.g., Breslow and Day, 1980; Checkoway et al., 2004], and are taught in all epidemiology training programs. We worry, however, that many ''potential'' limitations in epidemiology, particularly confounding and exposure misclassification, have assumed an aura of ''actual'' limitations, where it is not necessary to provide any evidence that the proposed limitation is present. Simply the mention of the possibility of a theoretical limitation is often sufficient to discount the study findings. In the field of occupational epidemiology, it seems that we are especially prone to react in this way on issues that are complex, contentious and hotly debated. These are, of course, the situations where we should demand data, not just opinions. Perhaps we should follow the proposition of Levitt and Dubner [2005] that ''conventional wisdom is often wrong'' and that a hypothesis of bias requires direct evidence to corroborate or refute, just like a hypothesis for a causal relationship.

We emphasize that we are not proposing that potential limitations be ignored. It is important to consider the possible impact of confounding and exposure misclassification on study results. We are concerned, however, that as a discipline, our assessment of the likelihood and impact of these two factors on study findings is unbalanced and this may lead to invalid conclusions, poor decision-making, and faulty public policy. This is clearly a scientific issue, but could also be construed as an ethical issue. Although interpretation of data is not usually recognized as an ethical issue, the American College of Epidemiology Ethics Guidelines identify ''making appropriate interpretations from the data analysis'' as one the criteria in the section on ''Adhering to the highest scientific standards'' [American College of Epidemiology, 2000].

CONFOUNDING

Confounding occurs when a factor is associated with the outcome in the absence of the exposure of interest and also with the exposure of interest. For confounding to occur, the factor must be a risk factor for the outcome and also correlated with the exposure of interest [Checkoway et al., 2004]. What may not be as well appreciated is that for confounders to have much of an impact, both associations (i.e., risk factor for the disease and correlation with the exposure of interest) must be strong [Breslow and Day, 1980]. If this is not the case, the impact of confounding cannot be large. Situations fulfilling these requirements are not common. Despite these rather stringent requirements, we find that many scientific discussions about potential confounding seem to assume that it is common and its impact is sizable. Typically the potential for confounding is hypothesized because some putative risk factor for the outcome of interest, or because some factor thought to be correlated with the exposure of interest has not been addressed in the study design or in the analyses. For example, in evaluating a study of a specific pesticide and lung cancer risk, suspected or established lung carcinogens (with no evidence of a linkage with the pesticide of interest), or other exposures that may coincide with the pesticide of interest (with no indication that they cause lung cancer) may be suggested as possible confounders. In such discussions, it is unusual for both associations to be considered and even rarer for the magnitude of these associations to be evaluated and for supportive data to be provided.

In occupational epidemiology, tobacco or other occupational exposures are commonly raised as potential confounders, particularly with retrospective cohort studies, since these studies often lack information on these factors. However, even without direct information on their occurrence or magnitude in the population under study, the possible impact of such confounding can be estimated.

For example, consider tobacco use as a confounder. Axelson [1978] made an extremely important contribution to this issue when he demonstrated that confounding from tobacco use in occupational studies of lung cancer was unlikely to entirely explain relative risks greater than 1.6. So, even without information on tobacco use, the Axelson approach [1978] could be used to set boundaries regarding the likely impact of smoking confounding. This approach was further evaluated and extended to additive models by Gail et al. [1988]. Using these approaches, the occurrence and likely magnitude of confounding by tobacco can be reasonably estimated because we have a considerable amount of information on relative risks from tobacco use for many diseases, as well as information on tobacco use by various occupations or exposures [Brackbill et al., 1988; Stellman et al., 1988]. With this information, it is relatively easy to estimate the potential impact of confounding by smoking, as suggested by Axelson [1978], thus, negating the need for pure speculation. Kriebel et al. [2004] extended this technique of indirect adjustment in a quantitative evaluation of the possible effects of confounding by tobacco and alcohol use in occupational studies. They concluded that changes of greater than 20% were unlikely.

The potential for confounding by tobacco can also be evaluated by assessing the correlation between smoking and specific occupational exposures. This may not always be possible because the necessary information is often not available. However, Siemiatycki et al. [1988b] evaluated the

relationship between level of exposure to 10 common occupational exposures (sulfur dioxide, welding fumes, engine emissions, gasoline, lubricating oil, solvents, paints/varnishes, adhesives, excavation dust, and wood dust) and tobacco use using data from a case-control study in Montreal. They found no correlation between occupational exposure indices for any of these substances and smoking history. Of course, tobacco use could be associated with other occupational exposures, but these data suggest that a strong association between smoking and specific exposures is unlikely.

Another approach to assess the magnitude and importance of confounding is to examine the impact of adjustment for possible confounders on estimates of relative risks. It has been our experience from numerous analyses for many potential confounders in our own studies that, just as theory indicates [Checkoway et al., 2004; Breslow and Day, 1980], confounding sufficient to affect interpretations of the data is extremely rare. We have not made a thorough review of the literature on this point for this paper, but we present a few examples. Table I presents odds ratios (ORs) for lung cancer by industry and occupation from a case-control study [Levin

TABLE I. Unadjusted and Adjusted (Age and Smoking) Odds Ratios for Lung Cancer by Occupation/Industry [From Levin et al., 1988]

Occupation/industry	Number cases/controls	Unadjusted OR	OR adjusted for age and smoking
Industry			
Agriculture, forestry, fishing	63/47	1.4	1.4
Food manufacturing	28/31	0.9	0.9
Textile	89/128	0.7	0.7
Sewing	34/30	1.2	1.3
Furniture	16/10	1.7	1.3
Chemical	34/25	1.4	1.7
Pharmaceuticals	12/10	1.3	1.2
Rubber and plastic	15/18	0.9	1.0
Metallurgical	84/73	1.2	1.1
General machinery	135/151	0.9	0.9
Electric equipment	27/33	0.8	0.9
Transportation	45/40	1.2	1.1
Precision machinery	23/19	1.3	1.5
Building construction	73/57	1.4	1.2
Food and beverage	198/225	0.9	1.0
Education, culture, arts	61/57	1.1	1.2
Scientific research	14/13	1.1	1.0
State organizations	93/92	1.1	1.0
Occupation			
Professionals/technicians	150/163	0.9	1.1
Service workers	189/172	1.2	1.2
Agricultural workers	54/37	1.6	1.6
Metal smelting	675/57	1.2	1.1
Chemical workers	17/11	1.6	1.4
Textile workers	38/53	0.7	0.7
Tanning and furs workers	12/11	1.1	0.9
Tailoring and sewing workers	21/25	0.9	1.0
Food and beverage workers	21/14	1.6	1.6
Metal forgers, tool makers	114/86	1.4	1.4
Machinery assemblers	53/65	0.8	0.9
Electrical equipment installers	19/25	0.8	0.8
Pipefitters, welders	26/30	0.9	0.9
Glass, ceramic workers	12/17	0.7	0.6
Painters	15/10	1.6	1.4
Construction workers	44/30	1.6	1.4
Power equipment operators	27/20	1.4	1.2
Transportation equipment operators	109/104	1.1	1.1

et al., 1988] where both crude and smoking-adjusted ORs were presented. Of the 36 comparisons of unadjusted and adjusted ORs, 26 were identical or differed by only 0.1, seven changed by 0.2, and two changed by 0.3. The results from an analysis of pooled data from several case-control studies of lung cancer in Germany were similar (Table II)[Bruske-Hohlfeld et al., 2000]. After adjusting for smoking and asbestos exposure, the biggest change in the ORs for lung cancer was about 0.3 or more and most ORs hardly changed at all. Similar results were found in a case-control study of lung cancer in Italy [Richiardi et al., 2005]. Likewise, Siemiatycki et al. [1988a] compared 75 smoking adjusted and smoking unadjusted relative risks for lung, bladder and stomach cancer. Only eight comparisons had differences of 20% or greater (seven for lung cancer and one for bladder cancer). Adjustment for smoking in a cohort study where the prevalence of smoking was positively correlated with the estimated level of exposure to acrylonitrile is shown in Table III [Blair et al., 1998]. The prevalence of ever smoking increased from 62% among workers in the lowest exposure

TABLE II. Odds Ratios for Lung Cancer With Different Adjustments (Age; Age, Smoking; and Age, Smoking, Asbestos) by Occupation/Industry [From Bruske-Hohlfeld et al., 2000]

	Number case/controls	Age adjusted OR	Age, smoking adjusted OR	Age, smoking, asbestos adjusted OR
Industry				
Agriculture, forestry, fishing	812/951	1.29	1.30	1.32
Energy and mining	274/440	1.72	1.47	1.44
Chemicals and oil	98/117	1.23	1.19	1.16
Rubber and plastics	43/85	2.04	1.94	1.89
Stone, glass, pottery	165/276	1.80	1.55	1.50
Metal production	574/764	1.45	1.37	1.27
Engine/vehicle building	791/1000	1.40	1.32	1.21
Electrical and sheet metal	499/446	0.89	0.90	0.87
Paper, wood, and printing	362/426	1.24	1.28	1.31
Food and tobacco	232/276	1.23	1.04	1.07
Construction	706/1004	1.63	1.35	1.32
Wholesale trade	475/404	0.83	0.71	0.73
Shipping and storage	318/410	1.37	1.13	1.14
Financing and insurance	119/97	0.79	0.76	0.79
Restaurants and hotels	128/166	1.36	1.04	1.06
Education, health, research	99/156	1.60	1.24	1.27
Occupation				
Farmer, agricultural workers	662/770	1.26	1.29	1.31
Forestry worker, fisherman	125/179	1.52	1.57	1.61
Miner	211/380	1.92	1.64	1.65
Stone cutter and carver	75/96	1.34	1.07	1.04
Chemical processor	104/170	1.69	1.56	1.55
Paper maker, printer	76/71	0.95	0.87	0.89
Cabinet maker	274/314	1.20	1.32	1.36
Metal producer and processor	460/731	1.77	1.49	1.42
Machinery mechanic, plumber	904/983	1.14	1.13	0.99
Electrician	286/246	0.87	0.87	0.82
Textile and leather worker	157/180	1.20	1.13	1.17
Food and beverage processor	218/281	1.35	1.14	1.17
Bricklayer, carpenter	330/498	1.65	1.39	1.33
Plasterer, insulator, upholsterer	108/152	1.43	1.37	1.34
Painter and lacquerer	96/147	1.60	1.39	1.42
Architect, technician, engineer	754/409	0.49	0.61	0.60
Sales worker	565/447	0.76	0.70	0.73
Medical, dental, veterinary worker	83/43	0.50	0.58	0.60
Social worker, teacher, scientist	361/122	0.32	0.39	0.41

TABLE III. Relative Risks for Lung Cancer by Estimated Level of Acrylonitrile Exposure Adjusted for Age, Calendar Time, Gender, and Race and also for Cigarette Use [From Blair et al., 1998]

	Quintile of estimated exposure					
Analysis group	**Lowest**	**2nd**	**3rd**	**4th**	**Highest**	***P* for trend**
% Ever smoked cigarettes	62%	64%	68%	72%	75%	
Full cohort	1.1	1.3	1.2	1.0	1.5	0.65
Selected smoking subcohort, not adjusted for smoking	0.8	1.1	1.0	0.9	1.5	0.70
Smoking subcohort with information on cigarette use, not adjusted for smoking	0.3	0.9	1.0	1.0	1.7	0.80
Smoking subcohort, adjusted for number of cigarettes per day	0.3	0.7	1.1	1.0	1.7	0.96
Full cohort with estimated changes from the smoking subcohort	1.1	1.0	1.1	0.9	1.4	

quintile to 75% in the highest. In this case, smoking was associated with the exposure of interest and we thought that the nonsignificant excess for lung cancer in the highest quintile (relative risk (RR) = 1.5) could be due to confounding. Because lung cancer was the a priori disease of interest, information on tobacco use was obtained from the next-of-kin of all the lung cancer cases and a 10% sample of the cohort on the noncases. Adjustment for smoking did not eliminate the elevated RR in the upper quintile. The RR for lung cancer in the highest quintile of exposure of the smoking subcohort increased from 1.5 to 1.7 without adjustment for tobacco use, but was unchanged when adjusted for smoking. Thus, the smoking–exposure relationship observed was apparently not tight enough to have much of an effect on the acrylonitrile/lung cancer relationship. Similar conclusions have been made in other analyses and surveys evaluating possible confounding by smoking in occupational studies [Blair et al., 1985; Simonato et al., 1988]. Data from studies of well-established occupational carcinogens also indicate that tobacco use does not confound these associations. For example, respiratory cancer is a well-demonstrated consequence of arsenic exposure among smelter workers [Lubin et al., 2000]. Tobacco does not appear to be associated with level of exposure and, consequently, does not confound the arsenic-respiratory cancer relationship [Welch et al., 1982]. Radon exposure among uranium miners has a sizable impact on lung cancer and this relationship is not confounded by smoking [Labbe et al., 1991]. Thus, these findings on well-established carcinogens indicate that confounding by tobacco use in occupational studies of lung cancer is rare and is not likely to be an explanation for positive study findings. We think the fact that tobacco use, which is the major risk factor for lung cancer and which differs by occupation and sometimes by estimated exposure to specific chemicals, rarely confounds disease risks from occupational associations is instructive. If tobacco does not confound lung cancer risks in occupational studies, it is even less likely that more modest risk factors for various diseases and with no known association with the exposure of interest would have a substantial effect.

Potential confounding from other exposures in the workplace is more difficult to evaluate [Blair et al., 1995]. This is because information is seldom available on the correlation between different occupational exposures, although we know that most work places have multiple exposures. What is often available, however, is information regarding the potential for these "other" exposures to cause the disease of interest. If experimental and epidemiologic studies do not suggest an association between a potential confounder and the disease, then perhaps we need not be as concerned that these factors function as confounders. Experience from our own studies and the article by Bruske-Hohlfeld et al. [2000] indicates that confounding by other work place exposures is also rare. Similarly, adjustment for asbestos exposure had little effect on the relationship between crystalline silica and lung cancer in diatomaceous earth workers [Checkoway et al., 1997]. Thus, a cursory examination of the literature suggests that confounding from other occupational exposures is not likely to be a common occurrence.

EXPOSURE MISCLASSIFICATION

It is important to note that the definition of exposure and the presence of exposure misclassification is tied to the objectives of the research. For example, if the study goal is to evaluate the association between airborne measurements of radon gas and lung cancer in underground miners, then an exposure assessment based on airborne measurements (if performed appropriately) may not suffer from much misclassification (measurement error would still occur) and the estimates of relative risk are unbiased. In contrast, misclassification is more likely to occur if the goal is to evaluate the risk of lung cancer by delivered dose of radiation to the lung tissue and exposure estimates were based on entirely on airborne measurements.

For etiologic research, a reasonable theoretical construct for exposure is "delivered dose to the target cell." Although desirable, this definition of exposure is largely unachievable. In practice, measured levels or estimates in air, water, dust, or

biologic tissues serve as surrogates for delivered dose. Axelson [1985] noted that assessment for relatively short (hours or days) exposures, for example, accidents or similar events, or constant/life-long exposures can sometimes be relatively easy, but these situations are rare. More typically, exposures over a longer period of time period are of interest. Since occupational exposures vary in intensity over time, it is difficult to create an accurate time-dependent exposure model.

The theoretical underpinnings for exposure misclassification are well developed. Checkoway et al. [2004] describe this as information bias. Exposure misclassification can either be non-differential (the probability or degree of misclassification is the same among diseased and non-diseased subjects), or differential (the probability or degree of misclassification is not the same among the diseased and non-diseased). Non-differential misclassification tends to bias relative risks toward the null for dichotomous exposure classifications. Although it can move estimates of relative risks away from null for some categories in multi-level exposure indices, in the highest exposure category it can only diminish the relative risk [Dosemeci et al., 1990]. Thus, in multi-level exposure analyses, non-differential misclassification tends to disrupt exposure-response trends and diminish our confidence that a causal association exists. In cohort studies, exposure misclassification is typically thought to be nondifferential because exposure assessment is independent from diagnosis of disease. In contrast, differential misclassification of exposure can bias the relative risks toward or away from the null. This type of misclassification is typically thought of as more of a concern in case-control studies because information on exposure is often obtained after diagnosis of disease. It is our impression that clear evidence for differential misclassification in case-control studies is relatively uncommon, but we did not perform a thorough review of the epidemiologic literature on this point. The likely occurrence of nondifferential misclassification of exposure, however, does not insure that the relative risks are underestimated. This is because misclassification of exposure is not the only source of bias and other sources could move the risk estimates away from the null [Jurek et al., 2005]. On the other hand, nondifferential misclassification itself is unlikely to create false positive findings.

It is more difficult to evaluate the impact of exposure misclassification on relative risks in occupational studies than for confounding because of the absence of information on the level of misclassification present. The theoretical impact of exposure misclassification on relative risks, however, can be estimated with information on the validity of exposure measurements/assessments and predicted relative risks [Rothman and Greenland, 1998; Checkoway et al., 2004]. A number of publications have described the theoretical impact of misclassification. They demonstrate that the magnitude of the effect of exposure misclassification on estimates of relative risk varies by the degree of misclassification and prevalence of the exposure. It is clear from these publications that relatively small errors (i.e., 10%–20%) can have sizable effects on relative risks [Copeland et al., 1977; Flegal et al., 1986].

If the desired characterization of exposure in etiologic studies is delivered dose to the target tissue, then no epidemiologic study is free from exposure misclassification. Unfortunately, the difficulty of obtaining true ''gold standard'' measurements means we never precisely know where we stand on the misclassification scale. It is likely, however, that even when basing exposure estimates on environmental or biologic measurements, our estimate of exposure is not likely to be very accurate if ''delivered dose to the target organ'' is the desired construct. Thus, even in the best of circumstances, exposure misclassification is likely to be considerable, and most epidemiologic studies do not possess ''ideal'' exposure measures. Some indication of the accuracy of occupational exposure assessment, however, can be gleaned from reports that compare different methods to assess a particular exposure. The sensitivity and specificity or correlation between two methods of exposure assessment provides some indication of the possible magnitude of misclassification, although it is important to remember that neither is likely to represent a ''gold standard.''

Table IV displays a few selected comparisons of occupational exposure assessments from the literature. The level of agreement in these studies shows Kappa values from ranging from 0.40 to 0.70 and correlations from 0.10 to 0.70. These values are roughly equivalent to the degree of misclassification and indicate that the level of disagreement between different measures of exposure is likely to exceed 30% in most circumstances and maybe as high as 70%. Use of these values as the actual range of misclassification assumes that one of the measures represents the ''gold standard.'' Since they do not, we are unsure of the how well the relationship between these two factors reflects that actual amount of misclassification.

The effect on relative risks from nondifferential misclassification in the 30%, percent range is sobering. Just in terms of relative ranking of subjects, Walker and Blettner [1985] showed that the classification of subjects by an exposure estimate that has a correlation of 0.70 with the true measure results in only 40% of the subjects being placed in the correct quintile of exposure. Even accepting correct placement by quintile as a success, it means that about 60% of the subjects would not be in the correct exposure quintile. Moreover, misclassification of exposure of this magnitude would have a considerable impact on estimates of relative risk. For example, Table V shows the impact of exposure misclassification on relative risks in hypothetical situations with a sensitivity of 0.7, specificity of 0.7 or 1.0, exposure prevalences of 10%, 30%, or 50% and true relative risks of

TABLE IV. Studies Reporting Different Exposure Assessment Techniques

Reference	Type of estimate	Exposure	Agreement
Friesen et al. [2003]	Expert estimate and measurements	Coal tar pitch volatiles	r = 0.42
Benke et al. [1997]	Expert estimate and measurements	Cutting, fluids-welding, fumes, lubricating oils	Kappa = 0.64, Kappa = 0.57, Kappa = 0.42
Ahrens et al. [1993]	JEMs, JEMs + questionnaires	Asbestos, asbestos	Kappa = 0.67, Kappa = 0.40
Baugher [1994]	PK model and measurements	2,4-D	r = 0.65
Steenland et al. [1999]	Expert estimates and serum measurements	2,3,7,8 tetrachloro-dibenzo-p-dioxin	r = 0.70
Stewart et al. [2003]	Deterministic, ratio, and homogeneous group methods with exposure measurements	Acrylonitrile	r = 0.63, r = 0.64, r = 0.66
Nieuwenhaijsen et al. [1995]	Estimates for average, cumulative, peak levels	Allergens	r range from 0.39 to 0.68
Stewart et al. [2000]	Expert estimates	Formaldehyde	r = 0.4 to 0.5

2.0 or 3.0. We chose these sensitivity and specificity values because they are roughly similar to level of exposure misclassification from the studies in Table IV. The amount of downward bias observed in situations displayed here is of such a magnitude that a reasonable interpretation of some of these observed relative risks would be that no association exists, even for a true relative risk of three. The observed relative risks are similar to what we often see in occupational studies, raising the question that we may be missing many occupational hazards because of exposure misclassification. In many, probably most, occupational studies, the sensitivity and specificity of exposure assessment may not reach 70% as assumed here and the level of bias would be even greater than displayed in Table V.

MISCLASSIFICATION OF A CONFOUNDER

Confounding factors can also suffer from misclassification. This is probably a common occurrence. The effects of confounder misclassification have been well discussed by Savitz and Baron [1989]. They make the point that in the presence of confounding, statistical adjustment is likely to be incomplete because of misclassification of the confounder and that the amount of confounding remaining is likely to be proportional to the amount removed in the adjustment process. Thus, misclassification of actual confounders would result in a general under assessment of the amount of confounding. Concern about residual confounding would be confined to situations where a meaningful difference is found between the adjusted and unadjusted point estimates, unless exposure assessment for the confounder is completely random.

TABLE V. Observed Relative Risks Based on Sensitivity, Specificity, Exposure Prevalence and True Relative Risks

True relative risk and prevalence of exposure	Sensitivity = 0.7; specificity = 1.0	Sensitivity = 0.7; specificity = 0.7
True relative risk = 2.0		
Exposure prevalence = 10%	1.94 (0.808)	1.15 (0.194)
Exposure prevalence = 30%	1.80 (0.760)	1.30 (0.359)
Exposure prevalence = 50%	1.63 (0.700)	1.31 (0.400)
True relative risk = 3.0		
Exposure prevalence = 10%	2.82 (0.808)	1.29 (0.194)
Exposure prevalence = 30%	2.44 (0.766)	1.53 (0.359)
Exposure prevalence = 50%	2.05 (0.700)	1.50 (0.400)

Kappas for the corresponding sensitivity, specificity, and exposure prevalence are in parentheses.

CONCLUSIONS

We believe of the two of the major methodologic issues raised in epidemiologic studies of occupational exposures, that is, confounding and exposure misclassification, the latter is of far greater concern. It is rare to find substantial confounding in occupational studies (or in other epidemiologic studies for that matter), even by risk factors that are strongly related to the outcome of interest. On the other hand, exposure misclassification probably occurs in nearly every epidemiologic study. For nondifferential misclassification, the type of misclassification most likely in cohort studies, the direction of the bias is largely predictable, that is, a bias of relative risks toward the null. In addition, the magnitude from relatively small amounts of misclassification can be sufficient to lead to an interpretation of no effect. Thus, interpretation of epidemiologic data and evaluations of epidemiologic studies should be more concerned about exposure assessment than confounding.

We find this is not usually the case. Extensive discussion of potential for confounding from specific, and sometimes unspecified, factors occurs routinely. Confounding is often raised as an explanation for positive findings without providing any information that the very specific conditions

required for it to occur actually do. On the other hand, discussions of exposure misclassification, if they occur at all, often imply that it may have created a false positive finding, even for cohort studies where nondifferential misclassification is likely to have the opposite effect. We think the relative attention paid to potential biases from confounding and exposure misclassification is unbalanced. To provide sound evaluations of epidemiology data, comments on confounding and exposure misclassification need to indicate the probability of occurrence, and magnitude and direction of possible effects to make sound scientific judgments and public policy decisions.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the NIH (National Cancer Institute, Division of Cancer Epidemiology and Genetics).

REFERENCES

Ahrens W, Jockel KH, Brochard P, Bolmaudorff U, Grossgarten K, Iwatsubo Y, Orlowski E, Pohlabeln H, Berrino F. 1993. Retrospective assessment of asbestos exposure. 1. Case-control analysis in a study of lung cancer efficientcy of job-specific questionnaires and job exposure matrices. Int J Epidemiol 22:s83–s95.

American College of Epidemiology. 2000. American College of Epidemiology Ethics Guidelines. Ann Epidemiol 10:487–497.

Axelson O. 1978. Aspects on confounding in occupational health epidemiology. Scand J Work Environ Health 10:325–326.

Axelson O. 1985. Dealing with the exposure variable in occupational and environmental epidemiology. Scand J Soc Med 18: 147–152.

Baugher DG. 1994. Validation of a pharmacokinetic model to predict exposure of ground boom mixer-loader-applicators to chlorophenoxyacetic acid herbicides—Comparison to conventional approaches. ACS Symposium Ser 542:214–230.

Benke G, Sim M, Forbes A, Salzberg M. 1997. Retrospective assessment of occupational exposure to chemicals in community-cased studies; validity and repeatability of industrial hygiene panel ratings. Int J Epidemiol 26:635–642.

Blair A, Hoar S, Walrath J. 1985. Comparison of crude and smoking-adjusted standardized mortality rates. J Occup Med 27:881–884.

Blair A, Stewart WF, Stewart PA, Sandler DP, Axelson O, Vineis P, Checkoway H, Savitz D, Pearce N, Rice C. 1995. A philosophy for dealing with hypothesized uncontrolled confounding in epidemiological investigations. Med Lav 86:106–110.

Blair A, Stewart PA, Pottern L, Zey JN, Bloom TF, Miller B, Ward E, Lubin J. 1998. Mortality of industrial workers exposed to acrylonitrile. Scand J Work Environ Health 24(Suppl 2):25–41.

Brackbill R, Frazier T, Shilling S. 1988. Smoking characteristics of US workers, 1978–1980. Am J Ind Med 13(1):5–42.

Breslow NE, Day NE. 1980. Statistical methods in cancer research, Vol. 1. In: The analysis of case-control studies. Lyon: IARC Sci Publ No. 338p.

Bruske-Hohlfeld I, Mohner M, Pohlabein H, Ahrens W, Bolm-Audorff U, Kreienbrock L, Kreuzer M, Jahn I, Wichmann H-E, Jockel K-H. 2000. Occupational lung cancer risk in Germany: Results from a pooled case-control study. Am J Epid 151:384–395.

Checkoway H, Heyer NJ, Seixas NS, Welp EA, Demers PA, Hughes JM, Weill H. 1997. Dose-response associations of silica and nonmalignant respiratory disease and lung cancer mortality in the diatomeous earth industry. Am J Epidemiol 145(8):680–688.

Checkoway H, Pearce N, Kriebel D. 2004. Research methods in occupational epidemiology. New York: Oxford University Press. 393p.

Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. 1977. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol 105:488–493.

Dosemeci M, Wacholder S, Lubin JH. 1990. Does nondifferential misclassification of exposure always bias the true effect toward the null? Am J Epidemiol 132:746–748.

Friesen MC, Demers PA, Spinelli JJ, Le ND. 2003. Validation of a semi-quantitative job exposure matrix at a Soderberg aluminum smelter. Ann Occup Hyg 47:477–484.

Gail MH, Wacholder S, Lubin JH. 1988. Indirect corrections for confounding under multiplicative and additive risk models. Am J Ind Med 13:119–130.

Jurek AM, Greenland S, Maldonado G, Church TR. 2005. Proper interpretation of non-differential misclassification effects: Expectations vs. observations. Int J Epid 34:680–687.

Kriebel D, Zeka A, Eisen EA, Wegman DH. 2004. Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. Int J Epidemiol 13:1–6.

Labbe KA, Howe GR, Burch JD, Miller AB, Abbatt J, Band P, Choi W, Du J, Feather J, Gallager R, Hill G, Mathews V. 1991. Radon exposure, cigarette-smoking, and other mining experience in the Beaverlodge uranium miners cohort. Health Phys 60:489–495.

Levin L, Zheng W, Blot WJ, Gao Y-T, Fraumeni JF, Jr. 1988. Occupation and lung cancer in Shanghai: A case-control study. Brit J Ind Med 45:450–458.

Levitt SD, Dubner SJ. 2005. Freakonomics: A rogue economist explores the hidden side of everything. New York: HarperCollins Books. 242p.

Lubin JH, Pottern LM, Stone BJ, Fraumeni JF, Jr. 2000. Respiratory cancer in a cohort of copper smelter workers: Results from more than 50 years of follow-up. Am J Epid 151:554–565.

Nieuwenhaijsen MJ, Lowson D, Venables KM, Newman Taylor AJ. 1995. Correlation between different measures of exposure in a cohort of bakery workers and flour millers. Ann Occup Hy 39:291–298.

Richiardi L, Forastiere F, Boffetta P, Simonato L, Merletti F. 2005. Effect of different approaches to treatment of smoking as a potential confounder in a case-control study on occupational exposures. Occup Environ Med 62(2):101–104.

Rothman JK, Greenland S. 1998. Modern epidemiology. 2nd edn. Philadelphia: Lippincott-Raven.

Savitz D, Baron AE. 1989. Estimating and correcting for confounder misclassification. Am J Epid 129:1062–1071.

Siemiatycki J, Wacholder S, Dewar R, Cardis E, Greenwood C, Richardson L. 1988a. Degree of confounding bias related to smoking, ethnic-group, and socioeconomic-status in estimates of the associations between occupation and cancer. J Occup Environ Med 30: 617–625.

Siemiatycki J, Wacholder S, Dewar R, Wald L, Begin D, Richardson L, Rosenman K, Gerin M. 1988b. Smoking and degree of occupational exposures: Are internal analyses in cohort studies

likely to be confounded by smoking status? Am J Ind Med 13:59–70.

Simonato L, Vineis P, Fletcher AC. 1988. Extimates of the proportion of lung cancer attributable to occupational exposure. Carcinogenesis 9:1159–1165.

Steenland K, Piacitelli L, Deddens J, Fingerhut M, Chang LI. 1999. Cancer, heart disease, and diabetes in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. JNCI 91:779–786.

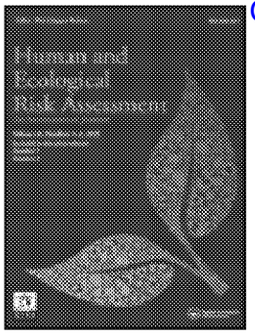
Stellman SD, Boffetta P, Garfinkel L. 1988. Smoking habits of 800,000 American men and women in relation to their occupation. Am J Ind Med 13(1):43–58.

Stewart PA, Carel R, Schairer C, Blair A. 2000. Comparison of industrial hygienists' exposure evaluations for an epidemiologic study. Scand J Work Environ Health 26:44–51.

Stewart PA, Lees PSJ, Correa A, Breysse P, Gail M, Gruabard BI. 2003. Evaluation of three retrospective exposure assessment methods. Ann Occup Hyg 47:399–411.

Walker AM, Blettner M. 1985. Comparing imperfect measures of exposure. Am J Epid 121:783–790.

Welch K, Higgins I, Oh M, Burchfiel C. 1982. Arsenic exposure, smoking, and respiratory cancer in copper smelter workers. Arch Environ Health 37:325–335.



Taylor & Francis
Taylor & Francis Group

Human and Ecological Risk Assessment: An International Journal

ISSN: 1080-7039 (Print) 1549-7860 (Online) Journal homepage: http://www.tandfonline.com/loi/bher20

The Use of Epidemiology in Risk Assessment: Challenges and Opportunities

Krista Christensen, Carol H. Christensen, J. Michael Wright, Audrey Galizia, Barbara S. Glenn, Cheryl Siegel Scott, Jennifer K. Mall, Thomas F. Bateson, Patricia A. Murphy & Glinda S. Cooper

To cite this article: Krista Christensen, Carol H. Christensen, J. Michael Wright, Audrey Galizia, Barbara S. Glenn, Cheryl Siegel Scott, Jennifer K. Mall, Thomas F. Bateson, Patricia A. Murphy & Glinda S. Cooper (2015) The Use of Epidemiology in Risk Assessment: Challenges and Opportunities, Human and Ecological Risk Assessment: An International Journal, 21:6, 1644-1663, DOI: 10.1080/10807039.2014.967039

To link to this article: http://dx.doi.org/10.1080/10807039.2014.967039

Accepted online: 25 Sep 2014.Published online: 09 Feb 2015.

Human and Ecological Risk Assessment, 21: 1644–1663, 2015
ISSN: 1080-7039 print / 1549-7860 online
DOI: 10.1080/10807039.2014.967039

The Use of Epidemiology in Risk Assessment: Challenges and Opportunities

Krista Christensen,[1] Carol H. Christensen,[2] J. Michael Wright,[3] Audrey Galizia,[1] Barbara S. Glenn,[1] Cheryl Siegel Scott,[1] Jennifer K. Mall,[1] Thomas F. Bateson,[1] Patricia A. Murphy,[4] and Glinda S. Cooper[1]

[1]National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, USA; [2]Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, DC, USA; [3]National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, USA; [4]National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Edison, NJ, USA

ABSTRACT

The assessment of risk from environmental and occupational exposures incorporates and synthesizes data from a variety of scientific disciplines including toxicology and epidemiology. Epidemiological data have offered valuable contributions to the identification of human health hazards, estimation of human exposures, quantification of the exposure–response relation, and characterization of risks to specific target populations including sensitive populations. As with any scientific discipline, there are some uncertainties inherent in these data; however, the best human health risk assessments utilize all available information, characterizing strengths and limitations as appropriate. Human health risk assessors evaluating environmental and occupational exposures have raised concerns about the validity of using epidemiological data for risk assessment due to actual or perceived study limitations. This article highlights three concerns commonly raised during the development of human health risk assessments of environmental and occupational exposures: (a) error in the measurement of exposure, (b) potential confounding, and (c) the interpretation of non-linear or non-monotonic exposure–response data. These issues are often the content of scientific disagreement and debate among the human health risk

Received 27 February 2014; revised manuscript accepted 14 September 2014.
Address correspondence to Carol H. Christensen, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, 1200 Pennsylvania Avenue NW, Mailcode 7509P, Washington, DC 20460, USA. E-mail: Christensen.Carol@epa.gov
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/bher.

assessment community, and we explore how these concerns may be contextualized, addressed, and often ameliorated.

Key Words: epidemiology, risk assessment, bias, measurement error, confounding, exposure–response, misclassification.

INTRODUCTION

Human health risk assessment (HHRA) is a process used to estimate the nature and probability of adverse health effects in humans who may be exposed to chemical and non-chemical stressors in environmental media (*e.g.*, air, water, soil, or food) or in the workplace (USEPA 2013). The risk assessment paradigm is comprised of four steps: hazard identification, exposure assessment, dose–response modeling, and risk characterization (NRC 1983). A risk assessment may be designed to address questions such as "What types of health problems may be caused by different environmental and occupational stressors such as chemicals, microbes, or radiation?" or "What is the probability that an adverse health effect will occur within a specific range of concentration or dose of these stressors?" The answers to these questions and others determine the scope of the human health risk assessment and influence what actions may be necessary for public health protection from environmental and occupational hazards. HHRA includes a synthesis of data from a variety of scientific disciplines including toxicology, epidemiology, industrial hygiene, and exposure science. Each of these types of scientific data has strengths as well as limitations for use in risk assessment.

Epidemiological data can provide valuable contributions to all stages of a HHRA, including hazard identification, exposure–response evaluation, and risk characterization. For several decades, different authors have extensively discussed the challenges of using epidemiological data in regulatory risk assessment—but have also emphasized the need to overcome these challenges, as human data provide unique information beyond what can be gleaned from traditional toxicology-based risk assessments (Gibb *et al.* 2002; Goldbohm *et al.* 2006; Gordis 1988; Graham *et al.* 1995; Hertz-Picciotto 1995; Johnson 2010; Lavelle *et al.* 2012; Samet *et al.* 1998; Schwartz 2002; Stayner *et al.* 2002; Whittemore 1986).

Over the past few decades, environmental epidemiology has advanced significantly, particularly with regard to exposure assessment methods, facilitating greater use of these data in risk assessment. For example, the Agricultural Health Study (AHS) developed a pesticide exposure metric for use in the prospective cohort study using data collected through self-report questionnaire (Alavanja *et al.* 1996). Exposure assessment methods developed further over the course of the follow-up of this cohort, and includes collection of additional biomonitoring data and other information to validate and improve the original algorithm (Coble et al. 2005, 2011; Thomas et al. 2010). In air pollution epidemiology, researchers and policy-makers have been working together to make best use of available time-series data to assess human health risk to particulate matter ($PM_{2.5}$) (Fann et al. 2011, 2012). In addition, researchers and policy-makers are looking beyond standard single chemical

exposures in HHRA, and are considering the role of multiple, cumulative chemical exposures as well as non-chemical exposures such as psycho-social stressors on health (Morello-Frosch and Shenassa 2006; Sexton and Hattis 2007). Given these advancements, this is an auspicious time to re-commit to the use of epidemiology in risk assessment to improve public health. For example, human data from modern epidemiology studies can inform the identification of hazards for which an animal model does not exist. These data can also inform estimates of risk in the low range of exposure and in the species of interest, and aid in the characterization of risks in sensitive populations (Burke 1995; Hertz-Picciotto 1995; Nachman *et al.* 2011; Samet *et al.* 1998). Consequently, many federal and international agencies that perform human health risk assessment state that epidemiological data should be preferentially incorporated into risk assessments when available (USEPA 2005; IARC 2000; NRC 2009).

Despite these recommendations, epidemiological data have been used in regulatory risk assessment relatively infrequently. For example, human data have been used to support less than 10% of risk assessments in the U.S. Environmental Protection Agency's (USEPA's) Integrated Risk Information System (IRIS) program (Persad and Cooper 2008), even in instances in which human data were available and could have been used more extensively in the risk assessment (Nachman *et al.* 2011; Persad and Cooper 2008). Concerns relating to the limitations and perceived insensitivity of epidemiological methods to meet the demands of risk assessment have been raised as a rationale against greater incorporation of these data in HHRA. One major limitation of observational studies is the potential for errors in assigning exposure values to study participants, possibly leading to misclassification of exposure and biased study results. Characterization of the anticipated direction, and even the magnitude, of this potential bias may be able to address this limitation. Another challenge involves the inadequate measure or control of potentially confounding variables. We discuss how the phenomenon of (strong) confounding such that study inference is incorrect is less common than presumed in published environmental and occupational epidemiology studies, and that there are strict criteria that must be met for a variable to bias study results in this way (Blair *et al.* 2007). Lastly, another misconception is that a non-linear or non-monotonic exposure–response trend in an epidemiology study is evidence of a non-causal relationship between exposure and disease. However, research from multiple scientific disciplines has shown that many true exposure–response relations are inherently non-linear or non-monotonic in nature, the identification of which adds scientific value to the risk assessment (Conolly and Lutz 2004; Vandenberg *et al.* 2012).

Understanding and constructively addressing the challenges noted above is critical for moving the field of environmental public health forward. Observational studies of environmental and occupational exposures reflect "real world" exposure–disease associations as opposed to experimentally controlled scenarios. As such, risk assessment models will benefit from incorporating these data, when appropriate. Situations in which the epidemiological data cannot be integrated into risk models in an easy or straightforward manner will inevitably lead to informative discussion within the multi-disciplinary team. In this article, we explore how data

from epidemiology studies can make a key contribution to understanding hazard and risk in human populations.

SPECIFIC ISSUES TO CONSIDER

Exposure Issues

Characterizing the degree to which humans come in contact with chemical, biological, radiological, or other agents in the environment or in the workplace is challenging. An accurate and precise measure of human exposure must reflect the timing, frequency, duration, and intensity of these exposures during a biologically relevant time period (*e.g.*, a lifetime cancer risk, or the period of gestational susceptibility). This may require extensive and, therefore, potentially expensive, exposure measurement efforts. The importance of these efforts is underscored by methodological research that indicates that misclassification as a result of incorrect exposure measurement likely influences bias in epidemiology studies to a far greater extent than confounding in epidemiology (Blair *et al.* 2007). The challenge of accurate and precise human exposure assessment notwithstanding, the use of human exposure information in human health risk assessment remains far superior to alternatives (*e.g.*, extrapolation from high-dose animal studies) (USEPA 2005, 2013; IARC 2000; Schwartz 2002).

Exposure assessment approaches can vary widely across occupational and environmental epidemiology studies. The type of disease or exposure under study (*e.g.*, acute or chronic), study population (occupationally or environmentally exposed), and the availability and feasibility of measurable exposure information will affect the type and quality of epidemiological exposure assessment. The extent to which epidemiology studies may contribute to a risk assessment will depend in large part on the exposure assessment. Many perceived flaws or inadequacies of epidemiology studies relate to the quality of the exposure assessment. These include the use of ecologic (group-level) versus individual-level exposure information; the grouping of exposure utilizing qualitative or semi-quantitative versus quantitative exposure categorization methods; and, the potential for error or mistakes in the measurement or classification of exposure. We posit, however, that studies using these imperfect methods may still inform risk assessment.

Assessment of Environmental and Occupational Exposures

Epidemiologists have a suite of exposure assessment approaches available to characterize human exposure to occupational and environmental agents. These include use of questionnaires; environmental or workplace measurement either alone or in combination with exposure modeling (*e.g.*, air dispersion modeling); personal or biological exposure monitoring; and use of exposure assessment tools such as job-exposure matrices. Using any of these exposure assessment methods, the actual exposure of interest (*i.e.*, the level of the agent or its active metabolite in the target tissue at the critical window of time) is rarely known with certainty, but available methods do allow the epidemiologist to rank or order participants in a study with high accuracy, thus allowing valid (unbiased) estimation of risk.

There are a variety of ways in which exposure assessment results may be utilized to estimate risk. Qualitative exposure measures are the least information-intensive approach, followed by semi-quantitative measures; quantitative measures are generally the most information intensive. In the experience of the authors, there may be a perception that epidemiology studies utilizing qualitative or semi-quantitative methods to categorize exposure are uninformative to the risk assessment process. However, there are valid uses for these data. An example of a qualitative exposure measure would be to characterize all workers in a particular job category as exposed to a substance, and compare them to workers in a different job category, considered to be "not exposed." Often a wide variety of industrial hygiene data are used to define exposure status for specific jobs or tasks within an industry. Such exposure assignments were made by studies of exposure to perchloroethyene in the dry cleaning industry; dry cleaners were classified as exposed, and launderers were classified as unexposed (Eskenazi *et al.* 1991; Gold *et al.* 2008; Raisanen *et al.* 2001). Epidemiology studies in which exposure is based on a dichotomized categorization (*i.e.*, exposed and unexposed) can inform the potential for hazard (or harm), but cannot support evaluation of exposure–response relationships without additional sources of information. Importantly, in some instances where the database of information is limited, studies with qualitative exposure measures represent the "best available" exposure measurement approach and may provide the only human data on an important public health issue.

Semi-quantitative exposure measures may also be used in epidemiology studies. These measures reflect more detailed information on each subject's individual exposure than qualitative methods and allow for an ordinal categorization (*e.g.*, low, medium, high) based on knowledge of a variety of factors including duration, frequency, and intensity of exposure, or based on knowledge of relative exposures in different types of jobs. The use of semi-quantitative exposure categories of increasing magnitude provides stronger evidence of a human health hazard than strictly qualitative (*e.g.*, none, low, or high) approaches and in some cases would allow evaluation of the relative exposure–response relationship.

Quantitative exposure classification can increase the accuracy of exposure estimates and should most closely represent the "true" (human) exposure experience. However, it should be emphasized that quantitative estimates (*e.g.*, individual air concentrations of a chemical during an 8-hour work shift, or individual biomarkers of a chemical) are not necessarily the "true" exposure of interest, but still a surrogate for this generally unknowable value. With that said, in addition to adding to the body of evidence in a hazard identification evaluation, studies with quantitative exposure data may inform the exposure-assessment phase of a risk assessment for a specific target population and may be used to estimate exposure–response relationships in greater detail for an exposed population.

A common misperception of environmental epidemiology studies is that they must include individual-level, quantitative exposure information in order to accurately characterize exposure for use in risk assessment. However, even with a complete lack of individual-level quantitative exposure measurements (*i.e.*, only group-level data are available), it may still be possible to apply externally derived exposure data for the characterization of risk. For example, external information

sources such as predictive models of exposure, literature-based exposure databases, or geographical information systems may be used to verify exposure trends observed on the group level, and even to develop individual-level exposure estimates in qualitative or semiquantitative categories (Henn *et al.* 2010; Ritz and Costello 2006; Teschke *et al.* 2002).

In summary, there are many ways to assess environmental exposures in epidemiology studies, each with inherent strengths and weaknesses. Even relatively crude qualitative measures of exposure such as "ever" or "never" exposed can be useful in identifying hazards associated with an exposure, and semi-quantitative and quantitative measures can further be used to support exposure–response analyses. In the next section, we discuss the implications of errors in the measurement of exposure, and the ability to correctly discern the magnitude and direction of the risk estimate despite these potential errors.

Exposure Measurement Error and Effect on Exposure–Outcome Associations

As described above, epidemiologists aspire to have exact dose or quantitative exposure information on each individual in the study population, but often this information is not feasible to obtain. Thus, nearly all exposure estimates are approximations or surrogates of delivered dose and are assumed to reflect some degree of error and misclassification (Smith 2002). Conceptually, it is useful to consider exposure measurement error in epidemiology studies as the difference between the "ideal" and the "actual" exposure estimate. The "error" is the difference between what epidemiologists would like to ideally measure and what is practically feasible to measure (Savitz 2003). Different types of measurement error can arise from a variety of sources. Some of these sources include analytical limitations (such as limited sensitivity of exposure measurement instruments resulting in more uncertainty in concentration measurements), sampling from a non-representative time period, and missing data. Appreciation of the different types of measurement error, and their effects on epidemiological measures of association, is critical in judging the influence of measurement error on the validity of the study as well as upon the utility of a study to assess the relationship between exposure and health outcomes. In the authors' experience, many perceive that an error in measurement renders the results of an epidemiology study unusable or unreliable. While that may be true in some instances, much of the time the magnitude and direction of the error can be predicted or characterized to allow accurate epidemiological inference (Smith 1988).

Measurement error is classified as either differential or non-differential with reference to the other comparison group. Non-differential error refers to an exposure assessment error that is independent of the health outcome status of the participants. Differential error occurs when the error is dependent on a person's outcome status. Recall bias is an example of this, where individuals with disease may remember more details about previous exposures than healthy individuals in a case-control study, or conversely the illness being examined may interfere with the ability of an individual to recall and report information on past exposures. The manner in which the misclassification is related to the disease outcome of interest influences the confidence in the resulting effect estimate.

Exposure misclassification bias can influence risk estimates derived from epidemiology studies, but the potential for error of this nature does not preclude the use of an epidemiology study for human health risk assessment. It is generally understood that in most instances, although there are some exceptions, non-differential error in exposure measurements (where the exposure error is independent of the health outcome status) for a dichotomized exposure results in an attenuation of the observed effect (*i.e.*, bias toward the null value of the measure of association) as well as inaccurate estimates of the precision of epidemiological effect estimates (*i.e.*, the standard error estimates are artificially small) (Blair *et al.* 2007; Deddens and Hornung 1994; Deklerk *et al.* 1989). However, this is not always true for continuous exposure measures, depending on the nature of the measurement error (*e.g.*, Berksonian bias; Armstrong 1990). Non-differential exposure measurement error in otherwise well-conducted epidemiology studies, while undesirable, would generally not be expected to create a false positive association (Correa-Villasenor *et al.* 1995; Jurek *et al.* 2008). That is, if the true odds ratio was actually 1.0 (no association between exposure and outcome), non-differential exposure measurement error is an unlikely explanation of a higher observed odds ratio such as 2.0. Sensitivity analyses such as assessing the effect estimates in relation to varying proportions of study participants presumed to be misclassified would aid characterization of this uncertainty in risk assessment. If the investigator has some knowledge about the exposure measurement error, statistical inferences may also be directly adjusted to account for this error (Stayner *et al.* 2007).

Exposure estimates may also be evaluated as quantitative measure (continuous data) or semi-quantitative (use of categorical variables). Classification of exposure using an ordinal scale (such as 1, 2, and 3 for low, medium, and high exposure) can be particularly useful for hazard identification or assessing relative trends but may be of limited use for quantifying exposure–response relationships, particularly if assumptions regarding homogeneous exposure and risk within these categories are not met. Misclassification bias is a particular concern when continuous exposure data are split into categories. For example, an unexposed participant may be mistakenly classified as exposed based on an arbitrary dichotomous exposure cut-point. Errors in exposure categorization can occur as a result of errors in data collection or data entry, failure to recall an exposure in a self-reported exposure questionnaire, or reliance on current exposure information as a proxy for exposures in the past, among others factors. Misclassification into adjacent categories is more likely than across several levels (*i.e.*, between medium and high exposure versus low and high exposure), and this misclassification can result in biased and imprecise study results.

In summary, while errors in classifying exposures of individual study participants occurs, methodological research into the effects of different types of classification errors allows informed epidemiological (causal) inference. Therefore, even when exposure measurement error is present, epidemiological data can still provide valuable information for risk assessment. Information is often available in epidemiology studies that can help characterize the direction or magnitude of errors to estimate their impact on the association between exposure and health outcome. Such information may come from the broader literature on exposure assessment, from methods papers on the study in question, or from supplemental information from the researchers (such as that found in appendices or online supplements). These

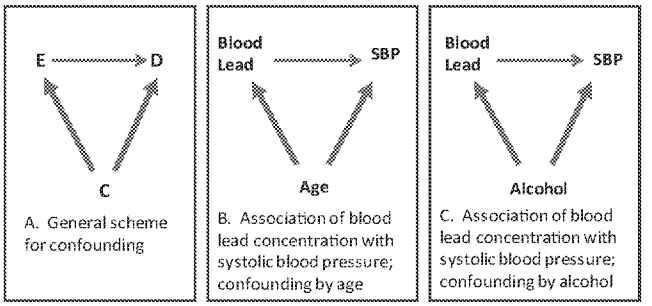


Figure 1. Panel A: Graphical depiction of general scheme for required relationships between exposure (E), disease (D), and potential confounder (C). Panels B and C depict the association of blood lead concentration with systolic blood pressure and potential relationships with age and alcohol consumption. Age and alcohol consumption both are associated with increases in blood lead concentration (Falq *et al.* 2011; Hense *et al.* 1992; Lee *et al.* 2005) and with systolic blood pressure (Marchi *et al.* 2014; Scinicariello *et al.* 2011). However, alcohol, primarily wine, contains lead (Ajtony *et al.* 2008), and adjustment for this source of lead exposure may remove its contribution to the variation in blood lead concentration. Thus, decisions regarding what confounders to adjust for can be complicated.

other sources of information may help clarify exposure-related issues, and thereby aid in the use of these studies for risk assessment (Fann *et al.* 2011). Although limitations in exposure assessment remains a challenge (Bailer 1999; Gordis 1988; Graham *et al.* 1995), the uncertainty in exposure measurement in epidemiology studies is likely to be small in comparison to the uncertainty in extrapolating from high doses in experimental animals to the complex human experience (Hertz-Picciotto 1995; Schwartz 2002; Smith 1988).

CONFOUNDING

Valid epidemiology studies must ensure that risk estimates from the factors (exposures) of primary interest are not unduly influenced by the presence of other risk factors, also known as confounders. Most of the major health outcomes influenced by exposure to environmental chemicals have several contributing causes (*i.e.*, multifactorial etiology) and may cluster within specific groups defined by common characteristics such as age, sex, race, socioeconomic status, or lifestyle. It is, therefore, important to account for potential differences in these factors between groups being compared (e.g., cases and controls, exposed and unexposed).

As illustrated in Figure 1, Panel A, confounders are factors that are: (1) associated *both* with the outcome, (2) *and also* with the exposure, (3) *but* do not mediate the effect of the exposure on the disease process (*i.e.*, be an intermediary factor in the causal pathway) (Szklo and Nieto 2004). All three criteria must be met for a variable to potentially confound the exposure–outcome association. For example, previous

studies indicate that age and alcohol consumption are potential confounders of the association between blood lead concentration and systolic blood pressure because both factors are associated with exposure and outcome (Falq *et al.* 2011; Hense *et al.* 1992; Lee *et al.* 2005; Marchi *et al.* 2014; Scinicariello *et al.* 2011). Decisions about confounding can be complicated, however; in this example, alcohol consumption can be a source of lead exposure (Ajtony *et al.* 2008). Therefore, adjustment for alcohol consumption could remove some of the contribution to the increased risk of high blood pressure due to blood lead concentration.

The potential for confounding including the inadequate control of confounding (known as residual confounding) is often noted as an impediment to the use of epidemiology studies in the evaluation of hazard and risk of an environmental agent (Hertz-Picciotto 1995). Studies will often not evaluate confounding by every possible known or hypothesized risk factor, in some cases simply because new or newly suspected risk factors may be identified after a study was completed. Although many factors may be suspected confounders, it is important to examine the available data (including previous studies on the same exposure and/or outcome) to determine if confounding is truly a concern. In many instances, suspected confounding variables are not truly confounding the exposure–disease relation under study because they do not meet the aforementioned requisite three criteria for confounding.

The Evaluation of Confounding

To address potential confounding in epidemiology studies, efforts are needed to ensure that comparison groups (*e.g.*, exposed and unexposed, cases and controls) are as similar as possible with the exception of the factor being evaluated (Savitz 2003). Some epidemiology and toxicology studies attempt to control for potential confounders through the randomization step of the experimental design (*i.e.*, similar distribution of potential confounders across animal exposure groups) (Festing and Altman 2002). Since randomization is not generally feasible in occupational or environmental epidemiology studies, potential confounding can be addressed through study design and statistical analysis. Potential confounders, such as age, sex, and race, are often controlled by techniques such as defining exclusion/inclusion criteria for subject recruitment, matching during study design and recruitment, or restriction in the data collection or analysis phases. For example, if age is suspected to be a confounder of a chemical being studied, a study might include only those in a certain age range, or exposed and unexposed participants might be matched by age or age group (Aschengrau and Seage 2003; Last 2001). These design features allow investigators to select study subjects so that potential confounders are distributed more equally among exposed and unexposed groups.

When appropriate data have been collected, potential confounders also can be controlled for during the data analysis phase by such methods as standardization, stratification, or statistical modeling. Standardization and stratification are two methods that can be used to develop a summary risk estimate while accounting for differences between comparison groups with respect to potential confounding characteristics (Aschengrau and Seage 2003; Rothman *et al.* 2008). The particular method

or methods selected to control for confounding are determined by the type of data available. For example, if vital statistics data (such as national or state mortality rates) are examined, then standardization can be used to control for confounding by comparing rates in the population under study with the rates in the general population using the same distributions for age, sex, and race. Controlling for confounding also can be achieved through statistical adjustment in multivariate models, a technique that easily allows simultaneous adjustment or stratification for several variables (Aschengrau and Seage 2003). In effect, statistical adjustment for confounding creates strata of individuals with similar values of the confounder for analysis. If the effect estimates are meaningfully different when potential confounders are included or not included in the model, then confounding may be present (often a difference of roughly +/− 10% in the effect estimate is considered evidence of confounding). The ability to meaningfully adjust for confounders in an analysis is dependent on the quality of the data, including the amount and type of measurement error in the confounding variables that are being examined.

Although statistical modeling is a powerful tool for addressing potential confounding, it is necessary to carefully select the factors to include in the exposure–response models, rather than including every possible variable or to rely solely on statistical criteria to determine which variables may be potential confounders. This is important because including extraneous risk factors in a regression can reduce precision and even produce unintended confounding due to the interrelationship of the included covariates, resulting in a biased effect estimate. Causal diagrams may be useful in judging whether including certain potential confounders in the model is necessary (Greenland *et al.* 1999; Hernandez-Diaz *et al.* 2008). For a more complete explanation of the use of causal diagrams in modeling decisions, see these citations: Howards *et al.* (2012), Schisterman *et al.* (2009), and VanderWeele (2009).

Influence of Confounding on Effect Estimates

If a confounder is identified as a concern during the planning of a study, the control of potential confounding may be relatively straightforward through aspects of the design and analysis described above. Evaluating the role of confounding for factors not considered in the design is more difficult, but still possible. The first consideration is whether there is any evidence to suggest potential confounding and, if so, its influence (direction and magnitude) on the risk estimate. Recall that all three criteria must be met (confounder must be associated with both the outcome of interest and with the exposure, but the confounder must not mediate the effect of the exposure on the disease process) in order for a variable to have a potential confounding effect on the exposure–outcome relation of interest (Szklo and Nieto 2004). When all of the aforementioned three conditions for confounding are met, the magnitude and direction of the bias depends on the strength and direction of the associations between the confounder and both the exposure and also the outcome of interest in a particular study, as well as the prevalence of the confounder in the population of interest.

For a confounder to fully explain the association between exposure and outcome, the confounder must have as great an influence on the relative risk of the

outcome as the exposure of interest. For example, analyses of confounding in occupational studies have found that the associations of smoking with both exposure and outcome must be moderately to strongly correlated before there is a change in the estimated risk for the outcome (Blair *et al.* 2007; Kriebel *et al.* 2004). Even for studies of occupational exposures and lung cancer risks, analyses that adjusted for smoking rarely found that the adjusted relative risk was substantially different from the unadjusted relative risk (*i.e.*, odds ratios differed by no more than 0.3 in the studies evaluated; Blair et al. 2007). Researchers concluded that in the occupational studies they evaluated, relative risks for lung cancer of 1.5 or higher are unlikely to be entirely explained by uncontrolled confounding by smoking behavior (Axelson and Steenland 1988). This is because the distribution of nonsmokers, moderate and heavy smokers must be very different between the exposed group and comparison population for smoking to substantially change the effect estimate.

Concerns about the influence of confounding on observed effect estimates may arise for studies involving populations exposed to more than one chemical or pollutant at a time. Co-exposures with moderate correlation should be considered as potential confounders in statistical models, if they also are risk factors for the health outcome under study and are not part of the exposure-to-response trajectory (*i.e.*, mediators in the causal pathway). Use of multi-variable regression techniques or other statistical tools such as factor analysis can isolate the exposure–disease association of interest, while controlling for the effect of co-exposures. In addition, if more than one study is available to evaluate the exposure–response relationship, then consistency in the collection of studies, including those that did or did not adjust for a particular co-exposure, can help determine if confounding by a specific co-exposure is likely. Although every individual is exposed to many agents, both chemical and non-chemical stressors, via various routes (oral, inhalation, dermal), it is likely that only a small subset of possible exposures would both be correlated with the exposure of interest, and also be risk factors for the health outcome of interest. Recall that both associations must be present at moderately strong correlations for confounding to occur. For example, Patel *et al.* (2012) found that in the National Health and Nutrition Examination Survey, biomarkers of exposure were generally not strongly correlated with each other; exceptions included compounds in the same chemical family (*e.g.*, polychlorinated biphenyls) that generally occur as mixtures in environmental media (Patel *et al.* 2012).

When a study population is exposed to multiple agents, and these exposures are highly correlated (*e.g.*, $\rho > .80$), it may be difficult to analytically disentangle individual exposure effects. This issue has been encountered in studies of many environmental contaminants, including air pollutants (Bell *et al.* 2007, 2009), drinking water contaminants (Rivera-Núñez and Wright 2013), and certain pesticides (Alavanja *et al.* 2003; Bell *et al.* 2007, 2009). In this situation, confounding may be difficult to address with statistical analysis. However, one may be able to draw insights from studies in other locations or exposure scenarios where the correlation between the same or similar agents is lower (Bell *et al.* 2011). When two or more agents are always encountered together, evaluating the risk of the combined exposure is a relevant consideration for public health since they better reflect real-world exposure

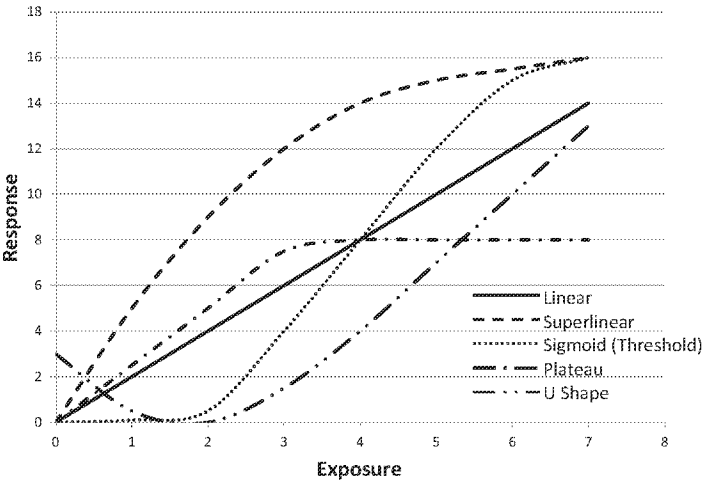


Figure 2. Examples of different exposure–response curves.

mixture scenarios and can offer some insight into potential combined effect of multiple exposures on human health.

THE EXPOSURE–RESPONSE RELATIONSHIP

The relation between environmental or occupational exposures and human diseases may take many different forms (linear or non-linear) (Figure 2). Critics may question the internal validity of an epidemiology study and its utility for quantitative risk assessment when the observed exposure–response relationship is not linear, or even non-monotonic. However, the observation of a non-monotonic curve in an individual study may be biologically plausible and can be used to inform a risk assessment (Wigle and Lanphear 2005). Further, the shape of the exposure–response relationship observed in a given study may depend on numerous factors including: population characteristics, the statistical model used, range of exposure, statistical power, and, as discussed previously, other factors including exposure measurement error (Brauer *et al.* 2002; Park and Stayner 2006). Consideration as to whether an observed exposure–response curve is a true representation of the underlying relation or an artifact of study design or conduct (*e.g.*, unbalanced observations per exposure category) requires expert consideration of many different factors.

The simplest exposure–response curve shape is linear, in which level of exposure is directly proportional to level of response. This type of relationship has been seen, for example, in epidemiology studies of methylmercury exposure and effects on neurodevelopment (NRC 2000). However, non-linear exposure–response curves are often observed in environmental and occupational epidemiology studies. A supra-linear relation in which exposure–response is linear at lower doses but attenuated at high doses, leading to an observed response plateau, is a frequently observed phenomenon in epidemiology (Blair *et al.* 1998; Cocco *et al.* 2001; Gibb *et al.* 2000; Hayes *et al.* 1996; Hertz-Picciotto and Smith 1993; Hornung and Meinhardt 1987;

Schubauer-Berigan *et al.* 2011; Stayner *et al.* 1993; Steenland *et al.* 1998, 1999, 2001). For example, birth weight and neurodevelopmental measures both have been observed to have a supralinear relationship with maternal and children's blood lead levels less than 10 μg/ml, respectively (Tellez-Rojo *et al.* 2006; Zhu *et al.* 2010). This plateau in the response curve may be due to factors such as exposure misclassification or a depletion of susceptible individuals in the population, or may represent a true biological phenomenon, such as receptor saturation or enzyme depletion (Stayner *et al.* 2003). Non-linearity may also arise in groups due to different exposure profiles, such as higher intensity and shorter duration of exposure, compared with lower intensity and longer duration (Lubin *et al.* 2008).

Another type of exposure–response relationship is a "U-shaped" curve in which the exposure–response association is lower in the mid-exposure range than at either the low or high ends of the exposure range. For example, both low and high levels of exposure to manganese in early life is related to risk of adverse neurodevelopmental effects, while exposures in the mid-range are not associated with these effects (Henn *et al.* 2010). Similarly, a U-shaped association between cadmium exposure and peripheral artery disease has been shown among non-smoking women (Tellez-Plaza *et al.* 2010).

Exposure–response relations may also exhibit an apparent threshold effect. This has been observed in the relation between PCBs exposure and neuropsychological function (Haase *et al.* 2009), where no response is observed below a certain dose (possibly due to compensatory mechanisms or lack of statistical power), but the exposure–response association is significant above a certain level of exposure. In epidemiology, as in experimental toxicology studies, however, it is difficult to detect effects at low exposures, and thus it is often difficult to establish the presence or absence of thresholds.

A statistical trend test is often used to examine the change in response over an entire range of exposures. For categorical analyses, differences in effect levels are compared between exposure groups. Statistically significant effect estimates may be observed in the highest exposure categories, with smaller and non-statistically significant effect estimates observed in the intermediate and/or lower exposure categories. This may be incorrectly interpreted to mean that the "trend" only starts at the point that statistical significance is reached, or, if statistical significance is not achieved for any exposure category, that there is an absence of an association between exposure and outcome. It may be the case that a monotonic trend is present, but statistical testing of individual grouped categories does not have sufficient power to demonstrate statistical significance compared to the more powerful trend test. As noted previously, misclassification of exposure and confounding variables may also result in bias, the result of which may be an inability to detect a true exposure–response relationship.

As noted above, several factors related to exposure and response can influence the observed relationship between the two factors. First, the range of exposures evaluated affects the shape of the curve. For example, no association may be observed if exposures in the study population were very low or were very similar among all study participants; however, an increasing trend in risk may exist over a wider exposure range. For categorical exposure comparisons, the choice of the

referent group (*i.e.*, the unexposed or a combination of those with no and low exposure) also affects comparisons made with higher exposure levels and can alter the exposure–response relationships (Stewart and Correa-Villasenor 1991). If there are positive associations at low exposure levels, inclusion of individuals with low to moderate exposure in the referent group (either by study design or due to exposure misclassification) can greatly influence effect estimates for the upper exposure categories, and decrease the slope of the observed exposure–response curve. Similarly, decisions as to how to categorize the exposure groups (*e.g.*, quartiles, or any exposure versus none) may affect the observed exposure–response relationship (Greenland 1995; Schulz *et al.* 2001; Van Wijngaarden 2005). For example, if the range within each exposure category is too broad the overall relationship may be obscured.

Exposure measurement error introduces variation and can lead to bias in the observed exposure–response relationship. For example, when exposure is classified into more than two categories, non-differential misclassification of those with the highest exposure into the lowest exposure group and vice versa, could result in a systematic bias in the observed risk estimates, and incorrectly influence the direction of a trend across exposure categories (Dosemeci *et al.* 1990). In addition, the response may also vary depending on such factors as the timing and dose of exposure, genetic susceptibility, and other factors that can influence absorption, metabolism and excretion rates across individuals; such variation will affect the shape of the exposure–response curve in a given population (Rothman 1976).

In summary, certain environmental and occupational exposure–response trends may truly be non-linear or non-monotonic in nature. Therefore, the observation of a non-linear exposure–response relationship is not necessarily an indicator of a flaw in the study. Studies that report such non-linear curves can be informative and should not be dismissed; they may provide information on both hazard identification and exposure–response. Users of such epidemiological data can gain further insight into the reported relationships by graphing or plotting the curve when such data are available to do so. Such visual representation yields information on the range of the data overall and within each group, as well as the magnitude of differences between the groups. Additionally, interpretation of other evidence including mechanistic understanding of the key biological events can provide further insight on the shape of exposure–response curves and inform causal inference.

CONCLUSION

Epidemiological data provide valuable contributions to all stages of health risk assessment, and should be used whenever possible to help reduce uncertainty in risk estimates. This article outlined some considerations when using epidemiological data for risk assessment, relating to exposure measurements, confounding, and the shape of the observed exposure–response relationship. The improvements in epidemiological methods seen in studies published in recent years make this an auspicious time to re-commit to the use of epidemiology in risk assessment to improve public health.

DISCLAIMER

The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency or the Federal government.

ACKNOWLEDGMENT

The authors thank Dr. Aaron Blair, National Cancer Institute and Dr. Thomas Luben, U.S. Environmental Protection Agency, for their helpful review of this article.

REFERENCES

Ajtony Z, Szoboszlai N, Susko EK, *et al.* 2008. Direct sample introduction of wines in graphite furnace atomic absorption spectrometry for the simultaneous determination of arsenic, cadmium, copper and lead content. Talanta 76(3):627–34

Alavanja MC, Sandler DP, McMaster SB, *et al.* 1996. The agricultural health study. Environ Health Perspect 104(4):362–9

Alavanja MCR, Samanic C, Dosemeci M, *et al.* 2003. Use of agricultural pesticides and prostate cancer risk in the agricultural health study cohort. Am J Epidemiol 157(9):800–14

Armstrong BG. 1990. The effects of measurement errors on relative risk regressions. Am J Epidemiol 132(6):1176–84

Aschengrau A and Seage GR. 2003. Essentials of Epidemiology in Public Health. Jones & Bartlett Publishers Sudbury, MA, USA

Axelson O and Steenland K. 1988. Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Indust Med 13(1):105–18

Bailer AJ. 1999. Uncertainty in risk assessment. Current efforts and future hopes. Ann NY Acad Sci 895:367–72

Bell ML, Kim JY, and Dominici F. 2007. Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies. Environ Health Persp 115(11):1591–5

Bell ML, Ebisu K, Peng RD, *et al.* 2009. Hospital admissions and chemical composition of fine particle air pollution. Am J Resp Crit Care 179(12):1115–20

Bell ML, Ebisu K, and Peng RD. 2011. Community-level spatial heterogeneity of chemical constituent levels of fine particulates and implications for epidemiological research. J Expo Sci Env Epid 21(4):372–84

Blair A, Stewart PA, Zaebst DD, *et al.* 1998. Mortality of industrial workers exposed to acrylonitrile. Scand J Work Env Hea 24:25–41

Blair A, Stewart P, Lubin JH, *et al.* 2007. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Indust Med 50(3):199–207

Brauer M, Brumm J, Vedal S, *et al.* 2002. Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects. Risk Anal 22(6):1183–93

Burke TA. 1995. The proper role of epidemiology in regulatory risk assessment: A regulators perspective. In: Graham JD (ed), The Role of Epidemiology in Regulatory Risk Assessment., Elsevier Publishers Inc., New York, NY, USA

Coble J, Arbuckle T, Lee W, *et al.* 2005. The validation of a pesticide exposure algorithm using biological monitoring results. J Occupat Environ Hyg 2(3):194–201

Coble J, Thomas KW, Hines CJ, *et al.* 2011. An updated algorithm for estimation of pesticide exposure intensity in the agricultural health study. International J Environ Res Public Hlth 8(12):4608–22

Cocco P, Rice CH, Chen JQ, *et al.* 2001. Lung cancer risk, silica exposure, and silicosis in Chinese mines and pottery factories: The modifying role of other workplace lung carcinogens. Am J Indust Med 40(6):674–82

Conolly RB and Lutz WK. 2004. Nonmonotonic dose-response relationships: Mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol Sci 77(1): 151–7

Correa-Villasenor A, Stewart WF, Franco-Marina F, *et al.* 1995. Bias from nondifferential misclassification in case-control studies with three exposure levels. Epidemiology 6(3):276–81

Deddens JA and Hornung R. 1994. Quantitative examples of continuous exposure measurement errors that bias risk estimates away from the null. In: Smith CM, Christiani DC, and Kelsey KT (eds), Chemical Risk Assessment of Occupational Health: Current Applications, Limitations, and Future Prospects, pp 77–85. Auburn House, Westport, CT, USA

Deklerk NH, English DR, and Armstrong BK. 1989. A review of the effects of random measurement error on relative risk estimates in epidemiological-studies. Internat J Epidemio 18(3):705–12

Dosemeci M, Wacholder S, and Lubin JH. 1990. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 132(4): 746–8

Eskenazi B, Wyrobek AJ, Fenster L, *et al.* 1991. A study of the effect of perchloroethylene exposure on semen quality in dry cleaning workers. Am J Indust Med 20(5): 575–91

Falq G, Zeghnoun A, Pascal M, *et al.* 2011. Blood lead levels in the adult population living in France the French Nutrition and Health Survey (ENNS 2006-2007). Environ Internat 37(3):565–71

Fann N, Bell ML, Walker K, *et al.* 2011. Improving the linkages between air pollution epidemiology and quantitative risk assessment. Environ Health Perspect 119(12):1671–5

Fann N, Baker KR, and Fulcher CM. 2012. Characterizing the PM(2).(5)-related health benefits of emission reductions for 17 industrial, area and mobile emission sectors across the U.S. Environ Internat 49:141–51

Festing MF and Altman DG. 2002. Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR Journal/National Research Council, Institute of Laboratory Animal Resources 43(4):244–58

Gibb HJ, Lees PSJ, Pinsky PF, *et al.* 2000. Lung cancer among workers in chromium chemical production. Am J Indust Med 38(2):115–26

Gibb HJ, Checkoway H, and Stayner L. 2002. Improving risk assessment: Priorities for epidemiologic research. Hum Ecol Risk Assess 8(6):1397–404

Gold LS, De Roos AJ, Waters M, *et al.* 2008. Systematic literature review of uses and levels of occupational exposure to tetrachloroethylene. J Occupat Environ Hyg 5(12): 807–39

Goldbohm RA, Tielemans ELJP, Heederik D, *et al.* 2006. Risk estimation for carcinogens based on epidemiological data: A structured approach, illustrated by an example on chromium. Regul Toxicol Pharm 44(3):294–310

Gordis L (ed). 1988. Epidemiology and Health Risk Assessment, Oxford University Press, New York, NY, USA

Graham J, Paustenbach D, and Butler W. 1995. Epidemiology and risk assessment: Divorce or marriage? In: Graham JD (ed), The Role of Epidemiology in Regulatory Risk Assessment, pp. 1–14. Elsevier Publishers Inc., New York, NY, USA

Greenland S. 1995. Dose-response and trend analysis in epidemiology—Alternatives to categorical analysis. Epidemiology 6(4):356–65

Greenland S, Pearl J, and Robins JM. 1999. Causal diagrams for epidemiologic research. Epidemiology 10(1):37–48

Haase RF, McCaffrey RJ, Santiago-Rivera AL, *et al.* 2009. Evidence of an age-related threshold effect of polychlorinated biphenyls (PCBs) on neuropsychological functioning in a Native American population. Environ Res 109(1):73–85

Hayes RB, Yin SN, Dosemeci M, *et al.* 1996. Mortality among benzene-exposed workers in China. Environ Health Persp 104:1349–52

Henn BC, Ettinger AS, Schwartz J, *et al.* 2010. Early postnatal blood manganese levels and children's neurodevelopment. Epidemiology 21(4):433–9

Hense HW, Filipiak B, Novak L, *et al.* 1992. Nonoccupational determinants of blood lead concentrations in a general population. Internat J Epidemiol 21(4): 753–62

Hernandez-Diaz S, Wilcox AJ, Schisterman EF, *et al.* 2008. From causal diagrams to birth weight-specific curves of infant mortality. Eur J Epidemiol 23(3):163–6

Hertz-Picciotto I. 1995. Epidemiology and quantitative risk assessment: A bridge from science to policy. Am J Public Health 85(4):484–91

Hertz-Picciotto I and Smith AH. 1993. Observations on the dose-response curve for arsenic exposure and lung-cancer. Scand J Work Env Hea 19(4):217–26

Hornung RW and Meinhardt TJ. 1987. Quantitative risk assessment of lung-cancer in United-States uranium miners. Hlth Physics 52(4):417–30

Howards PP, Schisterman EF, Poole C, *et al.* 2012. "Toward a clearer definition of confounding" revisited with directed acyclic graphs. Am J Epidemiol 176(6):506–11

IARC (International Agency for Research on Cancer). 2000. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Ionizing Radiation, Part 1: X- and Gamma (g)-Radiation, and Neutrons. Lyon, France. Available at http://monographs.iarc.fr/ENG/Monographs/vol75/mono75.pdf (accessed July 15, 2013)

Johnson BL. 2010. The vital importance of epidemiology in risk assessment. Hum Ecol Risk Assess 16(4):669–72

Jurek AM, Greenland S, and Maldonado G. 2008. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? Internat J Epidemiol 37(2):382–5

Kriebel D, Zeka A, Eisen EA, *et al.* 2004. Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. Internat J Epidemio 33(5):1040–5

Last JM. 2001. A Dictionary of Epidemiology, 4th edit. Oxford University Press, New York, NY, USA

Lavelle KS, Robert Schnatter A, Travis KZ, *et al.* 2012. Framework for integrating human and animal data in chemical risk assessment. Regul Toxicol Pharm 62(2):302–12

Lee MG, Chun OK, and Song WO. 2005. Determinants of the blood lead level of US women of reproductive age. J Am Coll Nutrit 24(1):1–9

Lubin JH, Moore LE, Fraumeni JF, Jr., et al. 2008. Respiratory cancer and inhaled inorganic arsenic in copper smelters workers: A linear relationship with cumulative exposure that increases with concentration. Environ Health Perspect 116(12):1661–5

Marchi KC, Muniz JJ, and Tirapelli CR. 2014. Hypertension and chronic ethanol consumption: What do we know after a century of study? World J Cardiol 6(5):283–94

Morello-Frosch R and Shenassa ED. 2006. The environmental "Riskscape" and social inequality: Implications for explaining maternal and child health disparities. Environ Health Persp 114(8):1150–3

Nachman KE, Fox MA, Sheehan MC, *et al.* 2011. Leveraging epidemiology to improve risk assessment. The Open Epidemio J 4:3–29

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. The National Academies Press, Washington, DC, USA

NRC. 2000. Toxicological Effects of Methylmercury. Washington, DC, USA. Available at http://www.nap.edu/openbook.php?isbn=0309071402 (accessed July 15, 2013)

NRC. 2009. Science and Decisions: Advancing Risk Assessment. National Research Council (U.S.). Committee on Improving Risk Analysis Approaches Used by the U.S. EPA, Washington, DC, USA. Available at http://www.nap.edu/catalog/12209/science-and-decisions-advancing-risk-assessment (accessed July 15, 2013)

Park RM and Stayner LT. 2006. A search for thresholds and other nonlinearities in the relationship between hexavalent chromium and lung cancer. Risk Anal 26(1): 79–88

Patel CJ, Cullen MR, Ioannidis JP, *et al.* 2012. Systematic evaluation of environmental factors: Persistent pollutants and nutrients correlated with serum lipid levels. Internat J Epidemiol 41(3):828–43

Persad AS and Cooper GS. 2008. Use of epidemiologic data in Integrated Risk Information System (IRIS) assessments. Toxicol Appl Pharmacol 233(1):137–45

Raisanen J, Niemela R, and Rosenberg C. 2001. Tetrachloroethylene emissions and exposure in dry cleaning. J Air Waste Manage Assoc 51(12):1671–5

Ritz B and Costello S. 2006. Geographic model and biomarker-derived measures of pesticide exposure and Parkinson's disease. Living in a Chemical World: Framing the Future in Light of the Past 1076:378–87

Rivera-Núñez Z and Wright JM. 2013. Associations between trihalomethane and haloacetic acid exposure and fetal growth and preterm delivery. J Occupat Environ Med 55(10):1125–34

Rothman KJ. 1976. Causes. Am J Epidemiol 104(6):587–92

Rothman KJ, Greenland S, and Lash TL. 2008. Modern Epidemiology, 3rd edit. Lipincott Williams and Wilkins, Philadelphia, PA, USA

Samet JM, Schnatter R, and Gibb H. 1998. Epidemiology and risk assessment. Am J Epidemiol 148(10):929–36

Savitz DA. 2003. Interpreting Epidemiologic Evidence: Strategies for Study Design and Analysis. Oxford University Press, Oxford, UK

Schisterman EF, Cole SR, Platt RW. 2009. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 20(4):488–95

Schubauer-Berigan MK, Deddens JA, Couch JR, *et al.* 2011. Risk of lung cancer associated with quantitative beryllium exposure metrics within an occupational cohort. Occupat Environ Med 68(5):354–60

Schulz MR, Hertz-Picciotto I, Van Wijngaarden E, *et al.* 2001. Dose-response relation between acrylamide and pancreatic cancer. Occupat Environ Med 58(9):609

Schwartz J. 2002. The use of epidemiology in environmental risk assessment. Hum Ecol Risk Assess 8(6):1253–65

Scinicariello F, Abadin HG, and Murray HE. 2011. Association of low-level blood lead and blood pressure in NHANES 1999–2006. Environ Res 111(8):1249–57

Sexton K, and Hattis D. 2007. Assessing cumulative health risks from exposure to environmental mixtures - Three fundamental questions. Environ Health Persp 115(5): 825–32

Smith AH. 1988. Epidemiologic input to environmental risk assessment. Arch Environ Health 43(2):124–9

Smith TJ. 2002. Issues in exposure and dose assessment for epidemiology and risk assessment. Hum Ecol Risk Assess 8(6):1267–93

Stayner L, Steenland K, Greife A, *et al.* 1993. Exposure-response analysis of cancer mortality in a cohort of workers exposed to ethylene-oxide. Am J Epidemio 138(10): 787–98

Stayner L, Toraason M, and Hattis D. 2002. Risk assessment at the crossroads of the 21st century: Opportunities and challenges for research. Hum Ecol Risk Assess 8(6): 1195–202

Stayner L, Steenland K, Dosemeci M, *et al.* 2003. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health 29(4):317–24

Stayner L, Vrijheid M, Cardis E, *et al.* 2007. A Monte Carlo maximum likelihood method for estimating uncertainty arising from shared errors in exposures in epidemiological studies of nuclear workers. Radiat Res 168(6):757–63

Steenland K, Deddens J, and Stayner L. 1998. Diesel exhaust and lung cancer in the trucking industry: Exposure-response analyses and risk assessment. Am J Indust Med 34(3): 220–8

Steenland K, Piacitelli L, Deddens J, *et al.* 1999. Cancer, heart disease, and diabetes in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. J Natl Cancer I 91(9): 779–86

Steenland K, Sanderson W, and Calvert GM. 2001. Kidney disease and arthritis in a cohort study of workers exposed to silica. Epidemiology 12(4):405–12

Stewart WF and Correa-Villasenor A. 1991. False positive exposure errors and low exposure prevalence in community-based case-control studies. Appl Occupat Environ Hyg Appl Occupat Environ Hyg 6(6):534–40

Szklo MA and Nieto FJ. 2004. Identifying Noncausal Associations: Confounding. In: Epidemiology: Beyond the Basics, pp 177–210. Jones and Bartlett Publishers, Boston, MA, USA

Tellez-Plaza M, Navas-Acien A, Crainiceanu CM, *et al.* 2010. Cadmium and peripheral arterial disease: Gender differences in the 1999-2004 US National Health and Nutrition Examination Survey. Am J Epidemio 172(6):671–81

Tellez-Rojo MM, Bellinger DC, Arroyo-Quiroz C, *et al.* 2006. Longitudinal associations between blood lead concentrations lower than 10 microg/dL and neurobehavioral development in environmentally exposed children in Mexico City. Pediatrics 118(2):e323–30

Teschke K, Chow Y, Brauer M, *et al.* 2002. Exposures and their determinants in radiographic film processing. Aihaj 63(1):11–21

Thomas KW, Dosemeci M, Coble JB, *et al.* 2010. Assessment of a pesticide exposure intensity algorithm in the agricultural health study. J Expo Sci Environ Epidemiol 20(6): 559–69

USEPA (US Environmental Protection Agency). 2005. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F Risk Assessment Forum. Available at http://www.epa.gov/osa/mmoaframework/pdfs/CANCER-GUIDELINES-FINAL-3-25-05[1].pdf (accessed July 15, 2013)

USEPA. 2013. Human Health Risk Assessment. Available at http://epa.gov/riskassessment/health-risk.htm (accessed 16 September 2013)

Van Wijngaarden E. 2005. A graphical method to evaluate exposure-response relationships in epidemiologic studies using standardized mortality or morbidity ratios. Dose Response 3(4):465–73

Vandenberg LN, Colborn T, Hayes TB, *et al.* 2012. Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses. Endocr Rev 33(3): 378–455

VanderWeele TJ. 2009. On the relative nature of overadjustment and unnecessary adjustment. Epidemiology 20(4):496–9

Whittemore A. 1986. Epidemiology in risk assessment for regulatory policy. J Chron Dis 39(12):1157–68

Wigle DT and Lanphear BP. 2005. Human health risks from low-level environmental exposures: No apparent safety thresholds. PLoS Med 2(12):1232–4

Zhu M, Fitzgerald EF, Gelberg KH, *et al.* 2010. Maternal low-level lead exposure and fetal growth. Environ Health Perspect 118(10):1471–5

Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

Reference Guide on Epidemiology

MICHAEL D. GREEN, D. MICHAL FREEDMAN, AND LEON GORDIS

Michael D. Green, J.D., is Bess & Walter Williams Chair in Law, Wake Forest University School of Law, Winston-Salem, North Carolina.

D. Michal Freedman, J.D., Ph.D., M.P.H., is Epidemiologist, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland.

Leon Gordis, M.D., M.P.H., Dr.P.H., is Professor Emeritus of Epidemiology, Johns Hopkins Bloomberg School of Public Health, and Professor Emeritus of Pediatrics, Johns Hopkins School of Medicine, Baltimore, Maryland.

CONTENTS

Examining a study for potential sources of bias is an important task that helps determine the accuracy of a study's conclusions. In addition, when a source of bias is identified, it may be possible to determine whether the error tended to exaggerate or understate the true association. Thus, bias may exist in a study that nevertheless has probative value.

Even if one concludes that the findings of a study are statistically stable and that biases have not created significant error, additional considerations remain. As repeatedly noted, an association does not necessarily mean a causal relationship exists. To make a judgment about causation, a knowledgeable expert[121] must consider the possibility of confounding factors. The expert must also evaluate several criteria to determine whether an inference of causation is appropriate.[122] These matters are discussed below.

C. Could a Confounding Factor Be Responsible for the Study Result?[123]

The third major reason for error in epidemiologic studies is confounding. Confounding occurs when another causal factor (the confounder) confuses the relationship between the agent of interest and outcome of interest.[124] (Confounding and selection bias (Section IV.B.1, *supra*) can, depending on terminology, overlap.) Thus, one instance of confounding is when a confounder is both a risk factor for the disease and a factor associated with the exposure of interest. For example, researchers may conduct a study that finds individuals with gray hair have a higher rate of death than those with hair of another color. Instead of hair color having an impact on death, the results might be explained by the confounding factor of age. If old age is associated differentially with the gray-haired group (those with gray hair tend to be older), old age may be responsible for the association found between hair color and death.[125] Researchers must separate the relationship between gray hair and risk of death from that of old age and risk of death. When researchers find an association between an agent and a disease, it is critical to determine whether the association is causal or the result of confounding.[126] Some

121. In a lawsuit, this would be done by an expert. In science, the effort is usually conducted by a panel of experts.

122. For an excellent example of the authors of a study analyzing whether an inference of causation is appropriate in a case-control study examining whether bromocriptine (Parlodel)—a lactation suppressant—causes seizures in postpartum women, see Kenneth J. Rothman et al., *Bromocriptine and Puerpal Seizures,* 1 Epidemiology 232, 236–38 (1990).

123. *See* Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991) (discussing the possibility that confounders may lead to an erroneous inference of a causal relationship).

124. *See* Rothman et al., *supra* note 61, at 129.

125. This example is drawn from Kahn & Sempos, *supra* note 31, at 63.

126. Confounding can bias a study result by either exaggerating or diluting any true association. One example of a confounding factor that may result in a study's outcome understating an

epidemiologists classify confounding as a form of bias. However, confounding is a reality—that is, the observed association of a factor and a disease is actually the result of an association with a third, confounding factor.[127]

Confounding can be illustrated by a hypothetical prospective cohort study of the role of alcohol consumption and emphysema. The study is designed to investigate whether drinking alcohol is associated with emphysema. Participants are followed for a period of 20 years and the incidence of emphysema in the "exposed" (participants who consume more than 15 drinks per week) and the unexposed is compared. At the conclusion of the study, the relative risk of emphysema in the drinking group is found to be 2.0, an association that suggests a possible effect). But does this association reflect a true causal relationship or might it be the product of confounding?

One possibility for a confounding factor is smoking, a known causal risk factor for emphysema. If those who drink alcohol are more likely to be smokers than those who do not drink, then smoking may be responsible for some or all of the higher level of emphysema among those who do not drink.

A serious problem in observational studies such as this hypothetical study is that the individuals are not assigned randomly to the groups being compared.[128] As discussed above, randomization maximizes the possibility that exposures other than the one under study are evenly distributed between the exposed and the control cohorts.[129] In observational studies, by contrast, other forces, including self-selection, determine who is exposed to other (possibly causal) factors. The lack of randomization leads to the potential problem of confounding. Thus, for example, the exposed cohort might consist of those who are exposed at work to an agent suspected of being an industrial toxin. The members of this cohort may, however, differ from unexposed controls by residence, socioeconomic or health status, age, or other extraneous factors.[130] These other factors may be causing (or

association is vaccination. Thus, if a group exposed to an agent has a higher rate of vaccination for the disease under study than the unexposed group, the vaccination may reduce the rate of disease in the exposed group, thereby producing an association that is less than the true association without the confounding of vaccination.

127. Schwab v. Philip Morris USA, Inc., 449 F. Supp. 2d 992, 1199–1200 (E.D.N.Y. 2006), *rev'd on other grounds,* 522 F.3d 215 (2d Cir. 2008), describes confounding that led to premature conclusions that low-tar cigarettes were safer than regular cigarettes. Smokers who chose to switch to low-tar cigarettes were different from other smokers in that they were more health conscious in other aspects of their lifestyles. Failure to account for that confounding—and measuring a healthy lifestyle is difficult even if it is identified as a potential confounder—biased the results of those studies.

128. Randomization attempts to ensure that the presence of a characteristic, such as coffee drinking, is governed by chance, as opposed to being determined by the presence of an underlying medical condition.

129. *See* Rothman et al., *supra* note 61, at 129; *see also supra* Section II.A.

130. *See, e.g., In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 783 (E.D.N.Y. 1984) (discussing the problem of confounding that might result in a study of the effect of exposure to Agent Orange on Vietnam servicemen), *aff'd,* 818 F.2d 145 (2d Cir. 1987).

protecting against) the disease, but because of potential confounding, an apparent (yet false) association of the disease with exposure to the agent may appear. Confounders, like smoking in the alcohol drinking study, do not reflect an error made by the investigators; rather, they reflect the inherently "uncontrolled" nature of exposure designations in observational studies. When they can be identified, confounders should be taken into account. Unanticipated confounding factors that are suspected after data collection can sometimes be controlled during data analysis, if data have been gathered about them.

To evaluate whether smoking is a confounding factor, the researcher would stratify each of the exposed and control groups into smoking and nonsmoking subgroups to examine whether subjects' smoking status affects the study results. If the relationship between alcohol drinking and emphysema in the smoking subgroups is the same as that in the all-subjects group, smoking is not a confounding factor. If the subjects' smoking status affects the relationship between drinking and emphysema, then smoking is a confounder, for which adjustment is required. If the association between drinking and emphysema completely disappears when the subjects' smoking status is considered, then smoking is a confounder that fully accounts for the association with drinking observed. Table 4 reveals our hypothetical study's results, with smoking being a confounding factor, which, when accounted for, eliminates the association. Thus, in the full cohort, drinkers have twice the risk of emphysema compared with nondrinkers. When the relationship between drinking and emphysema is examined separately in smokers and in nonsmokers, the risk of emphysema in drinkers compared with nondrinkers is not elevated in smokers or in nonsmokers. This is because smokers are disproportionately drinkers and have a higher rate of emphysema than nonsmokers. Thus, the relationship between drinking and emphysema in the full cohort is distorted by failing to take into account the relationship between being a drinker and a smoker.

Even after accounting for the effect of smoking, there is always a risk that an undiscovered or unrecognized confounding factor may contribute to a study's findings, by either magnifying or reducing the observed association.[131] It is, however, necessary to keep that risk in perspective. Often the mere possibility of uncontrolled confounding is used to call into question the results of a study. This was certainly the strategy of some seeking, or unwittingly helping, to undermine the implications of the studies persuasively linking cigarette smoking to lung cancer. The critical question is whether it is plausible that the findings of a given study could indeed be due to unrecognized confounders.

In designing a study, researchers sometimes make assumptions that cannot be validated or evaluated empirically. Thus, researchers may assume that a missing potential confounder is not needed for the analysis or that a variable used was adequately classified. Researchers employ a sensitivity analysis to assess the effect of those assumptions should they be incorrect. Conducting a sensitivity analysis

131. Rothman et al., *supra* note 61, at 129; *see also supra* Section II.A.

Table 4. Hypothetical Emphysema Study Data[a]

Drinking Status	Total Cohort				Smokers				Nonsmokers			
	Total	Cases	Incidence	RR	Total	Cases	Incidence	RR	Total	Cases	Incidence	RR
Nondrinkers	471	16	0.034	1.0[b]	111	9	0.081	1.0[b]	360	7	0.019	1.0[b]
Drinkers	739	41	0.069	2.0	592	48	0.081	1.0	147	3	0.020	1.0

[a] The incidence of disease is not normally presented in an epidemiologic study, but we include it here to aid in comprehension of the ideas discussed in the text.
[b] RR = relative risk. The relative risk for each of the cohorts is determined based on reference to the risk among nondrinkers; that is, the incidence of disease among drinkers is compared with nondrinkers for each of the three cohorts separately.

entails repeating the analysis using different assumptions (e.g., alternative corrections for missing data or for classifying data) to see if the results are sensitive to the varying assumptions. Such analyses can show that the assumptions are not likely to affect the findings or that alternative explanations cannot be ruled out.[132]

1. What techniques can be used to prevent or limit confounding?

Choices in the design of a research project (e.g., methods for selecting the subjects) can prevent or limit confounding. In designing a study, the researcher must determine other risk factors for the disease under study. When a factor or factors, such as age, sex, or even smoking status, are risk factors and potential confounders in a study, investigators can limit the differential distribution of these factors in the study groups by selecting controls to "match" cases (or the exposed group) in terms of these variables. If the two groups are matched, for example, by age, then any association observed in the study cannot be due to age, the matched variable.[133]

Restricting the persons who are permitted as subjects in a study is another method to control for confounders. If age or sex is suspected as a confounder, then the subjects enrolled in a study can be limited to those of one sex and those who are within a specified age range. When there is no variance among subjects in a study with regard to a potential confounder, confounding as a result of that variable is eliminated.

2. What techniques can be used to identify confounding factors?

Once the study data are ready to be analyzed, the researcher must assess a range of factors that could influence risk. In the hypothetical study, the researcher would evaluate whether smoking is a confounding factor by comparing the incidence of emphysema in smoking alcohol drinkers with the incidence in nonsmoking alcohol drinkers. If the incidence is substantially the same, smoking is not a confounding factor (e.g., smoking does not distort the relationship between alcohol drinking and the development of emphysema). If the incidence is substantially different, but still exists in the nonsmoking group, then smoking is a confounder, but does not wholly account for the association with alcohol drinking. If the association disappears, then smoking is a confounder that fully accounts for the association observed.

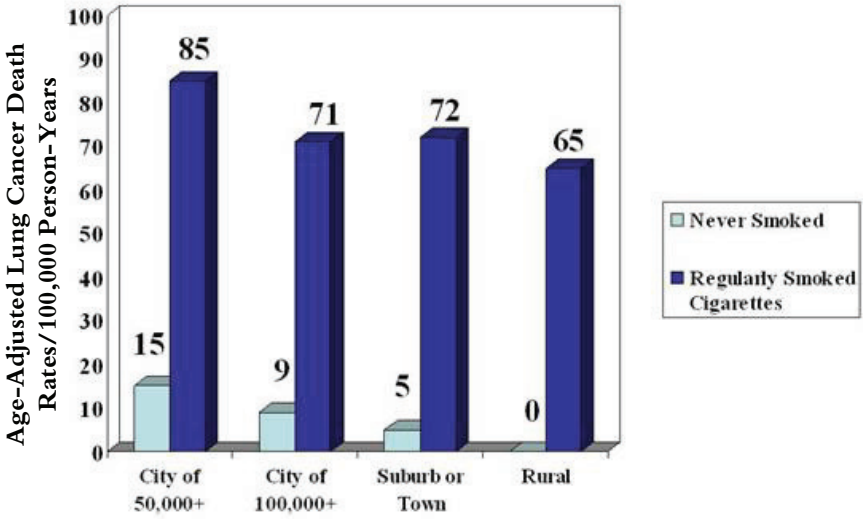
132. Kenneth Rothman & Sander Greenland, Modern Epidemiology (2d ed. 1998).

133. Selecting a control population based on matched variables necessarily affects the representativeness of the selected controls and may affect how generalizable the study results are to the population at large. However, for a study to have merit, it must first be internally valid; that is, it must not be subject to unreasonable sources of bias or confounding. Only after a study has been shown to meet this standard does its universal applicability or generalizability to the population at large become an issue. When a study population is not representative of the general or target population, existing scientific knowledge may permit reasonable inferences about the study's broader applicability, or additional confirmatory studies of other populations may be necessary.

3. What techniques can be used to control for confounding factors?

A good study design will consider potential confounders and obtain data about them if possible. If researchers have good data on potential confounders, they can control for those confounders in the data analysis. There are several analytic approaches to account for the distorting effects of a confounder, including stratification or multivariate analysis. Stratification permits an investigator to evaluate the effect of a suspected confounder by subdividing the study groups based on a confounding factor. Thus, in Table 4, drinkers have been stratified based on whether they smoke (the suspected confounder). To take another example that entails a continuous rather than dichotomous potential confounder, let us say we are interested in the relationship between smoking and lung cancer but suspect that air pollution or urbanization may confound the relationship. Thus, an observed relationship between smoking and lung cancer could theoretically be due in part to pollution, if smoking were more common in polluted areas. We could address this issue by stratifying our data by degree of urbanization and look at the relationship between smoking and lung cancer in each urbanization stratum. Figure 5 shows actual age-adjusted lung cancer mortality rates per 100,000 person-years by urban or rural classification and smoking category.[134]

Figure 5: Age-adjusted lung cancer mortality rates per 100,000 person-years by urban or rural classification and smoking category.



Source: Adapted from E. Cuyler Hammond & Daniel Horn, *Smoking and Death Rates—Report on Forty-Four Months of Follow-Up of 187,783 Men: II, Death Rates by Cause*, 166 JAMA 1294 (1958).

134. This example and Figure 4 are from Leon Gordis, Epidemiology 254 (4th ed. 2009).

For each degree of urbanization, lung cancer mortality rates in smokers are shown by the dark gray bars, and nonsmoker mortality rates are indicated by light gray bars. From these data we see that in every level (or stratum) of urbanization, lung cancer mortality is higher in smokers than in nonsmokers. Therefore, the observed association of smoking and lung cancer cannot be attributed to level of urbanization. By examining each stratum separately, we, in effect, hold urbanization constant, and still find much higher lung cancer mortality in smokers than in nonsmokers.

For each degree of urbanization, lung cancer mortality rates and smokers are shown by the dark-colored bars, and nonsmoker mortality rates are indicated by light-colored bars. For these data we see that in every level (or stratum) of urbanization, lung cancer mortality is higher in smokers than in nonsmokers. Therefore, the observed association of lung cancer cannot be attributed to level of urbanization. By examining each stratum separately, we are, in effect, holding urbanization constant, and we still find much higher lung cancer mortality in smokers than in nonsmokers.

Multivariate analysis controls for the confounding factor through mathematical modeling. Models are developed to describe the simultaneous effect of exposure and confounding factors on the increase in risk.[135]

Both of these methods allow for adjustment of the effect of confounders. They both modify an observed association to take into account the effect of risk factors that are not the subject of the study and that may distort the association between the exposure being studied and the disease outcomes. If the association between exposure and disease remains after the researcher completes the assessment and adjustment for confounding factors, the researcher must then assess whether an inference of causation is justified. This entails consideration of the Hill factors explained in Section V, *infra*.

V. General Causation: Is an Exposure a Cause of the Disease?

Once an association has been found between exposure to an agent and development of a disease, researchers consider whether the association reflects a true cause–effect relationship. When epidemiologists evaluate whether a cause–effect relationship exists between an agent and disease, they are using the term causation in a way similar to, but not identical to, the way that the familiar "but for," or sine qua non, test is used in law for cause in fact. "Conduct is a factual cause of

135. For a more complete discussion of multivariate analysis, see Daniel L. Rubinfeld, Reference Guide on Multiple Regression, in this manual.

References on Epidemiology

Causal Inferences (Kenneth J. Rothman ed., 1988).
William G. Cochran, Sampling Techniques (1977).
A Dictionary of Epidemiology (John M. Last et al. eds., 5th ed. 2008).
Anders Ahlbom & Steffan Norell, Introduction to Modern Epidemiology (2d ed. 1990).
Robert C. Elston & William D. Johnson, Basic Biostatistices for Geneticists and Epidemiologists (2008)
Encyclopedia of Epidemiology (Sarah E. Boslaugh ed., 2008).
Joseph L. Fleiss et al., Statistical Methods for Rates and Proportions (3d ed. 2003).
Leon Gordis, Epidemiology (4th ed. 2009).
Morton Hunt, How Science Takes Stock: The Story of Meta-Analysis (1997).
International Agency for Research on Cancer (IARC), Interpretation of Negative Epidemiologic Evidence for Carcinogenicity (N.J. Wald & R. Doll eds., 1985).
Harold A. Kahn & Christopher T. Sempos, Statistical Methods in Epidemiology (1989).
David E. Lilienfeld, *Overview of Epidemiology,* 3 Shepard's Expert & Sci. Evid. Q. 25 (1995).
David E. Lilienfeld & Paul D. Stolley, Foundations of Epidemiology (3d ed. 1994).
Marcello Pagano & Kimberlee Gauvreau, Principles of Biostatistics (2d ed. 2000).
Pharmacoepidemiology (Brian L. Strom ed., 4th ed. 2005).
Richard K. Riegelman & Robert A. Hirsch, Studying a Study and Testing a Test: How to Read the Health Science Literature (5th ed. 2005).
Bernard Rosner, Fundamentals of Biostatistics (6th ed. 2006).
Kenneth J. Rothman et al., Modern Epidemiology (3d ed. 2008).
David A. Savitz, Interpreting Epidemiologic Evidence: Strategies for Study Design and Analysis (2003).
James J. Schlesselman, Case-Control Studies: Design, Conduct, Analysis (1982).
Lisa M. Sullivan, Essentials of Biostatistics (2008).
Mervyn Susser, Epidemiology, Health and Society: Selected Papers (1987).

References on Law and Epidemiology

American Law Institute, Reporters' Study on Enterprise Responsibility for Personal Injury (1991).
Bert Black & David H. Hollander, Jr., *Unraveling Causation: Back to the Basics*, 3 U. Balt. J. Envtl. L. 1 (1993).
Bert Black & David Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 Fordham L. Rev. 732 (1984).

Gerald Boston, *A Mass-Exposure Model of Toxic Causation: The Content of Scientific Proof and the Regulatory Experience,* 18 Colum. J. Envtl. L. 181 (1993).

Vincent M. Brannigan et al., *Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases,* 12 Risk Analysis 343 (1992).

Troyen Brennan, *Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation,* 73 Cornell L. Rev. 469 (1988).

Troyen Brennan, *Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts,* 51 U. Pitt. L. Rev. 1 (1989).

Philip Cole, *Causality in Epidemiology, Health Policy, and Law,* 27 Envtl. L. Rep. 10,279 (June 1997).

Comment, *Epidemiologic Proof of Probability: Implementing the Proportional Recovery Approach in Toxic Exposure Torts,* 89 Dick. L. Rev. 233 (1984).

George W. Conk, *Against the Odds: Proving Causation of Disease with Epidemiological Evidence,* 3 Shepard's Expert & Sci. Evid. Q. 85 (1995).

Carl F. Cranor, Toxic Torts: Science, Law, and the Possibility of Justice (2006).

Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After* Daubert v. Merrell Dow Pharmaceuticals, Inc., 16 Va. Envtl. L.J. 1 (1996).

Richard Delgado, *Beyond Sindell: Relaxation of Cause-in-Fact Rules for Indeterminate Plaintiffs,* 70 Cal. L. Rev. 881 (1982).

Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact,* 7 Harv. Envtl. L. Rev. 429 (1983).

Jean Macchiaroli Eggen, *Toxic Torts, Causation, and Scientific Evidence After* Daubert, 55 U. Pitt. L. Rev. 889 (1994).

Daniel A. Farber, *Toxic Causation*, 71 Minn. L. Rev. 1219 (1987).

Heidi Li Feldman, Science and Uncertainty in Mass Exposure Litigation, 74 Tex. L. Rev. 1 (1995).

Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation,* 36 Jurimetrics J. 1 (1995).

Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy (1988).

Herman J. Gibb, *Epidemiology and Cancer Risk Assessment*, *in* Fundamentals of Risk Analysis and Risk Management 23 (Vlasta Molak ed., 1997).

Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence,* 96 Yale L.J. 376 (1986).

Leon Gordis, *Epidemiologic Approaches for Studying Human Disease in Relation to Hazardous Waste Disposal Sites,* 25 Hous. L. Rev. 837 (1988).

Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation,* 86 Nw. U. L. Rev. 643 (1992).

Michael D. Green, *The Future of Proportional Liability, in* Exploring Tort Law (Stuart Madden ed., 2005).

Sander Greenland, *The Need for Critical Appraisal of Expert Witnesses in Epidemiology and Statistics,* 39 Wake Forest L. Rev. 291 (2004).

Khristine L. Hall & Ellen Silbergeld, *Reappraising Epidemiology: A Response to Mr. Dore*, 7 Harv. Envtl. L. Rev. 441 (1983).

Jay P. Kesan, *Drug Development: Who Knows Where the Time Goes?: A Critical Examination of the Post-*Daubert *Scientific Evidence Landscape,* 52 Food Drug Cosm. L.J. 225 (1997).

Jay P. Kesan, *An Autopsy of Scientific Evidence in a Post-*Daubert *World,* 84 Geo. L. Rev. 1985 (1996).

Constantine Kokkoris, Comment, DeLuca v. Merrell Dow Pharmaceuticals, Inc.: *Statistical Significance and the Novel Scientific Technique,* 58 Brook. L. Rev. 219 (1992).

James P. Leape, *Quantitative Risk Assessment in Regulation of Environmental Carcinogens,* 4 Harv. Envtl. L. Rev. 86 (1980).

David E. Lilienfeld, *Overview of Epidemiology*, 3 Shepard's Expert & Sci. Evid. Q. 23 (1995).

Junius McElveen, Jr., & Pamela Eddy, *Cancer and Toxic Substances: The Problem of Causation and the Use of Epidemiology,* 33 Clev. St. L. Rev. 29 (1984).

Modern Scientific Evidence: The Law and Science of Expert Testimony (David L. Faigman et al. eds., 2009–2010).

Note, *Development in the Law—Confronting the New Challenges of Scientific Evidence,* 108 Harv. L. Rev. 1481 (1995).

Susan R. Poulter, *Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?* 7 High Tech. L.J. 189 (1992).

Jon Todd Powell, Comment, *How to Tell the Truth with Statistics: A New Statistical Approach to Analyzing the Data in the Aftermath of* Daubert v. Merrell Dow Pharmaceuticals, 31 Hous. L. Rev. 1241 (1994).

Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28, cmt. c & rptrs. note (2010).

David Rosenberg, *The Causal Connection in Mass Exposure Cases: A Public Law Vision of the Tort System,* 97 Harv. L. Rev. 849 (1984).

Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life-Cycle of Mass Torts,* 43 Hastings L.J. 301 (1992).

Joseph Sanders, *Scientific Validity, Admissibility, and Mass Torts After* Daubert, 78 Minn. L. Rev. 1387 (1994).

Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law,* 64 L. & Contemp. Probs. 107 (2001).

Palma J. Strand, *The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation,* 35 Stan. L. Rev. 575 (1983).

Richard W. Wright, *Causation in Tort Law*, 73 Cal. L. Rev. 1735 (1985).

Maternal Smoking and Childhood Leukemia and Lymphoma Risk among 1,440,542 Swedish Children

Lorelei A. Mucci,[1,2] Frederik Granath,[1,3] and Sven Cnattingius[1]

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [2]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; and [3]Clinical Epidemiology Unit, Department of Medicine at Karolinska Hospital, Karolinska Sjukhuset, Stockholm, Sweden

Abstract

Possible *in utero* effects of maternal smoking on hemopoietic cancer in the offspring have been addressed previously, although the results are inconclusive. In this investigation, we take advantage of population-based registers in Sweden to examine maternal smoking during pregnancy and childhood risk of leukemia and lymphoma. Prospective data were available from 1,440,542 Swedish children born between 1983 and 1997. Proportional hazard models were used to estimate hazard ratios (HR) and 95% confidence intervals (95% CI) controlling for potential confounders. In the study base, 750 hemopoietic cancers occurred across 11 million person-years. Incidence rates per 100,000 person-years were 4.7 for acute lymphocytic leukemia (ALL), 0.45 for acute myelogenous leukemia, and 0.76 for non-Hodgkin's lymphoma. Maternal smoking was associated with a lower risk of ALL (HR, 0.73; 95% CI, 0.58–0.91). On the other hand, there was a higher risk of acute myelogenous leukemia (HR, 1.41; 95% CI, 0.74–2.67) particularly among heavy (≥10 cigarettes per day) smokers (HR, 2.28; 95% CI, 1.05–4.94). The data also suggested a small excess risk of non-Hodgkin's lymphoma (HR, 1.25; 95% CI, 0.76–2.04). Evidence from this large cohort suggests that maternal smoking affects the risk of childhood leukemia and lymphoma in the offspring. The Swedish registries provide unique opportunities to examine this research question, with a design inherently free of selection and recall biases. The apparent protective effect with ALL needs to be explored further and in no way supports maternal smoking as beneficial, given its adverse association with common pregnancy outcomes. (Cancer Epidemiol Biomarkers Prev 2004;13(9):1528–33)

Introduction

The negative effects of cigarette smoking on cancer risk in adulthood are well documented and include convincing evidence of an increased risk of cancer of the lung and larynx (1), bladder (2), esophagus (3), and oral cavity (4). The possible *in utero* effects of maternal smoking during pregnancy on subsequent cancer risk in the offspring have been addressed more recently through epidemiologic studies, although the results are in large part inconclusive (5, 6). With respect to childhood leukemia and lymphoma, several case-control studies have observed a positive effect of maternal smoking during pregnancy on risk of acute lymphocytic leukemia (ALL; refs. 7, 8), acute myelogenous leukemia (AML; refs. 9, 10), and lymphomas (7, 11). Other studies have found no association between maternal smoking and risk of these cancers (7, 12), whereas others still showed some evidence of a protective effect at least for ALL (13-15) and AML (15, 16).

A well-conducted case-control study is an efficient design to examine *in utero* exposure to cigarette smoking and risk of childhood cancer. However, this study design is vulnerable to potential biases, including selection and recall biases, which could account for the diverging results of prior studies. Given the rarity of childhood leukemia and lymphoma, however, a cohort study, which would avoid these potential limitations, is often difficult to undertake with sufficient statistical power.

In the present investigation, we take advantage of existing population-based registers in Sweden to examine the effect of maternal smoking during pregnancy on childhood risk of leukemia and lymphoma among a cohort of 1,440,542 Swedish children born between 1983 and 1997.

Materials and Methods

Study Population. The study base for the present investigation consists of all live births in Sweden between January 1, 1983 and December 31, 1997 that were registered in the population-based Swedish Medical Birth Registry. The Birth Registry includes >99% of all births in Sweden (17). Follow-up data on this cohort were achieved through linkage of the Birth Registry with the Swedish Cancer Registry and the National Cause of Death Registry. Because each Swedish

Received 11/21/03; revised 4/13/04; accepted 4/15/04.

Grant support: Swedish Cancer Society grant 4242-B00-02XBB and Swedish Foundation for International Cooperation in Research and Higher Education grant, supporting collaboration between Karolinska Institutet and Harvard University.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Sven Cnattingius, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, P.O. Box 281, SE-171 77 Stockholm, Sweden. Phone: 46-8-524-86190; Fax: 46-8-31-49-57. E-mail: sven.cnattingius@meb.ki.se

Table 1. Characteristics of malignant childhood leukemia and lymphoma (*International Classification of Diseases, Seventh Edition* codes 200.0–207.0) in Sweden among cohort of 1,440,542 children born 1983–1997

	International Classification of Diseases, Seventh Edition code	*n*	Rate per 10^5 person-years	Mean (SD) age at diagnosis	% Male
ALL	204.0	505	4.75	3.7 (2.7)	53.7
AML	205.0	48	0.45	3.5 (3.9)	45.8
Chronic myelogenous leukemia	205.1	13	0.12	5.4 (3.8)	53.9
Other leukemias*	206–207	22	0.21	3.6 (3.7)	45.5
NHL†	200, 204.1	81	0.76	5.7 (3.0)	74.7
Hodgkin's disease	201	20	0.19	7.0 (3.7)	75.0
Reticulosis	202	61	0.57	2.4 (2.9)	57.4

*Includes 6 monocytic leukemias and 16 other and unspecified leukemias.
†Includes two chronic lymphocytic leukemia cases classified as NHL.

resident is assigned a national registration number, which is a unique identifier, it is possible to merge national databases.

Information on incident leukemia and lymphoma cases in the cohort came from the Swedish Cancer Registry, established by the National Board of Health and Welfare in 1958. Swedish law mandates and regulates physicians and pathologists, who confirm the diagnosis of cancer, to report on every newly diagnosed malignant tumor to the Swedish Cancer Registry. Since the early 1980s, all notifications of cancer diagnosis have been sent directly to one of six regional cancer registers, each of which has a strictly defined catchment area. All case reports are verified for completeness at the regional registries and subsequently computerized. Incidence statistics from the six regional registries are pooled in the Swedish Cancer Registry.

Information on all deaths in the cohort was available from the National Cause of Death Registry. The registry includes dates of death from specific causes, which is obtained from death certificates and coded according to the standards of the *International Classification of Diseases, Eighth, Ninth, and Tenth Editions*. Medical certification is carried out by the attending physician or coroner, with use of both clinical records and autopsy reports. This registry, which was established in 1961, maintains date and cause of death for >99% of residents who died after this year.

Among the 1,591,271 Swedish live births between 1983 and 1997, we excluded 3,627 (0.2%) infants who died within the first week of birth and 1,475 (0.1%) with Down syndrome. We excluded from the analysis an additional 97,905 (6.2%) births with missing information on maternal smoking, 47,573 (3.0%) with other missing covariate data, and 149 (0.01%) with erroneous follow-up information. Thus, the sample size of the final cohort for this analysis was 1,440,542 (90.5%) Swedish births during 1983 to 1997.

Data Collection. The Birth Registry includes standardized information from antenatal, obstetric, and neonatal medical records. During the first antenatal visit, normally at 8 to 12 gestational weeks, information from a standardized questionnaire is recorded by a nurse midwife. Information on maternal smoking during the first trimester has been collected routinely since 1983. Women were asked the number of cigarettes that they smoked, which was coded on the questionnaire as 0, 1 to 9, or ≥10 cigarettes per day. Additional covariate data include maternal demographic data, reproductive history, and birth characteristics and outcomes. Through linkage with the Education Registry, years of formal education attained as of December 31, 1998 were obtained from Statistics Sweden. Information on mother's country of birth was provided through linkage to the Immigration Registry and stratified into Nordic (Sweden, Denmark, Norway, Finland, and Iceland) or non-Nordic country of birth.

Lymphoma and Leukemia Cases. The incidence of lymphoma and leukemia (*International Classification of Diseases, Seventh Edition* codes 200–207) in the cohort was based on information provided by the Swedish Cancer Registry. Information available from the Swedish Cancer Registry includes date of diagnosis, malignancy, histologic subtype (WHO/HS/CANC/24.1 Histology Code), basis of diagnosis, and death from cancer. Observation time of the cohort was calculated from date of entry into the cohort (birth date) until the occurrence of a diagnosis of any primary lymphoma or leukemia cancer, or censoring since diagnosis of another cancer, death, or end of the observation period (December 31, 1997).

Statistical Analysis. The relation between maternal smoking and risk of childhood lymphoma or leukemia in the offspring was assessed using information on time to cancer event, which accounts for different amounts of follow-up time in the cohort. First, the incidence rates of cancer in the entire cohort were estimated by dividing the number of cases that occurred during follow-up by the total number of person-years at risk for a given level of exposure. Proportional hazard models using Proc PHREG in SAS version 8.2 were used to estimate the hazard ratio (HR) and 95% confidence interval (95% CI) of hemopoietic cancers, given smoking status, comparing nonsmokers as the reference. To assess whether the dose of cigarettes increased or decreased risk in a linear fashion, we calculated statistical tests for trend. The following covariates were evaluated as potential confounders: maternal age (categorically: ≤19, 20–24, 25–29, 30–34, ≥35 years), maternal education (categorically: ≤9, 10–11, 12, ≥13 years), parental status (cohabitating/not cohabitating), residence at birth (town or rural/large city), maternal birthplace (Nordic/non-Nordic), parity (categorically: 1, 2–3, ≥4), birth year (ordinal), and baby's gender (male/female). Because of concern that

birth weight (ordinally) and gestational age (categorically: <32, 32–36, ≥37 weeks) potentially could be considered on the causal pathway, we controlled for these variables in a secondary analysis.

Because of the early age at onset of ALL, we examined whether the effect of smoking was constant by age at diagnosis. To accomplish this, we stratified models into risk sets of 0 to 1 (completed), 2 to 4, and ≥5 years of follow-up and estimated the effect of maternal smoking in each risk group. Furthermore, we examined whether the effect of maternal smoking on ALL differed among male and female offspring, comparing the estimates formally with a test for interaction.

Results

This cohort of 1,440,542 children born in Sweden between 1983 and 1997 contributed almost 11 million person-years to the study base. ALL was by far the most common occurring of the leukemias and lymphomas, with an incidence rate of 4.75 per 100,000 person-years (Table 1). The characteristics of non-Hodgkin's lymphoma (NHL) and Hodgkin's disease cases were notably different than ALL and AML, with an older mean age and a predominance of male cases.

In Table 2, we present the prevalence of maternal smoking during pregnancy by demographic, reproductive, and birth characteristics. Overall, 24% of women smoked during pregnancy. The proportion of women smoking during pregnancy was higher among younger women, among those with lower levels of education, and among those born in Nordic countries. Maternal smoking was also associated with preterm birth and lower birth weight. Over the course of the study period, there was evidence of notable decreases in smoking prevalence.

Adjusting for potential confounders, maternal smoking was associated with a 30% lower risk of ALL (HR, 0.73; 95% CI, 0.58–0.91; Table 3). The risk reduction was similar for light (1–9 cigarettes per day) and heavy (≥10 cigarettes per day) smokers. On the other hand, there was evidence that maternal smoking was associated with a higher risk of AML. In particular, children whose mothers smoked ≥10 cigarettes per day during early pregnancy had a >2-fold higher risk of AML (HR, 2.28; 95% CI, 1.05–4.94) compared with women who did not smoke. The data also suggested a small excess risk of NHL, although because of the small number of cases, 95% CIs were wide. In the proportional hazard analyses, further adjustment by gestational age and birth weight did not substantial change the HRs, suggesting that these variables are neither confounders nor on the causal pathway.

In Table 4, we present estimates of the effect of maternal smoking on ALL stratified by age at diagnosis and sex. A decreased risk of ALL associated with maternal smoking was evident for each age at diagnosis (Table 4), although the effect was more consistent among those diagnosed at ages 0 to 1 years. Maternal smoking was associated with a significantly protective effect on risk of ALL among males only, but there was no evidence of a statistical interaction between maternal smoking and infant's sex on risk of ALL (P for interaction = 0.32).

Discussion

Evidence from this large cohort of Swedish children suggests that maternal smoking during pregnancy affects the risk of childhood leukemia in the offspring. The data are consistent with a small protective effect of smoking on risk of ALL and with an excess risk of AML. There is also some evidence that maternal smoking increases the risk of NHL, although small numbers of cases in the cohort prevent definitive conclusions. Although there is no statistical evidence of interaction, the effect of maternal smoking on ALL seems more consistent among male compared with female offspring and slightly stronger for infants during the first year of life.

Table 2. Frequency of smoking during pregnancy by maternal and reproductive characteristics among 1,440,542 Swedish births, Sweden, January 1983–December 1997

	N	Smoking during pregnancy (%)
Maternal age (y)		
≤19	37,243	43.6
20–24	311,861	29.8
25–29	538,653	22.2
30–34	379,602	21.1
≥35	173,183	21.7
Maternal education (y)		
≤9	243,553	43.1
10–11	593,128	28.3
12	177,235	15.7
13–14	252,578	12.3
≥15	174,048	8.4
Parental status		
Cohabitating	1,308,277	22.7
Not cohabitating	71,318	48.4
Town/city		
Large city	383,063	23.6
Town/rural	1,057,479	24.2
Maternal birthplace		
Nordic	1,323,945	24.8
Non-Nordic	116,597	15.4
Parity		
1	584,022	24.0
2–3	753,583	23.4
≥4	102,937	29.4
Multiple birth		
Singleton	1,406,909	24.1
Multiple	33,633	23.5
Offspring sex		
Female	700,348	24.0
Male	740,014	24.1
Gestational age (wk)		
≤31	8,143	31.7
32–36	57,624	28.9
≥37	1,373,847	23.8
Birth weight (g)		
<1,500	7,822	32.1
1,500–2,500	51,340	36.1
2,501–3,500	625,964	29.4
3,501–4,500	703,703	19.1
>4,500	46,871	12.7
Birth year		
1983–1986	343,557	30.3
1987–1990	402,512	26.5
1991–1994	432,496	21.6
1995–1997	261,977	16.1

Table 3. Crude and adjusted HRs for the effect of maternal smoking on leukemia and lymphoma, Sweden, January 1983–December 1997

	Cases (*n*)	Rate per 10^5 person-years	Crude HR	Adjusted HR* (95% CI)	Adjusted HR† (95% CI)
Maternal smoking					
ALL					
No	400	5.93	Reference	Reference	Reference
Yes	105	4.01	0.73	0.73 (0.58–0.91)	0.75 (0.60–0.93)
1–9 cigarettes	61	3.80	0.69	0.68 (0.52–0.89)	0.69 (0.52–0.91)
≥10 cigarettes	44	4.35	0.80	0.80 (0.58–1.10)	0.84 (0.61–1.15)
P for trend			0.016	0.012	0.043
AML					
No	33	0.49	Reference	Reference	Reference
Yes	15	0.57	1.28	1.41 (0.74–2.67)	1.28 (0.65–2.49)
1–9 cigarettes	6	0.37	0.83	0.91 (0.38–2.21)	0.75 (0.29–1.96)
≥10 cigarettes	9	0.89	2.00	2.28 (1.05–4.94)	2.20 (1.00–4.83)
P for trend			0.15	0.084	0.13
NHL					
No	56	0.83	Reference	Reference	Reference
Yes	25	0.96	1.17	1.25 (0.76–2.04)	1.22 (0.74–2.02)
1–9 cigarettes	15	0.93	1.14	1.21 (0.68–2.18)	1.15 (0.63–2.11)
≥10 cigarettes	10	0.99	1.21	1.30 (0.65–2.60)	1.33 (0.66–2.68)
P for trend			0.51	0.38	0.39
Reticulosis					
No	44	0.65	Reference	Reference	Reference
Yes	17	0.65	1.11	1.20 (0.67–2.16)	1.12 (0.61–2.05)
1–9 cigarettes	14	0.87	1.48	1.60 (0.86–3.00)	1.47 (0.77–2.79)
≥10 cigarettes	3	0.30	0.51	0.54 (0.17–1.77)	0.54 (0.16–1.77)
P for trend			0.74	0.855	0.739

*Data adjusted for maternal age, maternal education, maternal birthplace, parity, birth year, and baby's gender.
†Data also adjusted for gestational age and birth weight.

In evaluating the results of the study, there are several strengths to consider. The large size and duration of follow-up provide one of the few opportunities to evaluate the research question of maternal smoking on cancer risk using a cohort design. The Swedish Medical Birth and Cancer Registers include 99% of all births and 96% of cancer cases in Sweden (17, 18), respectively. Using these population-based resources almost eliminates the possibility of selection bias and loss to follow-up.

Maternal smoking in this study was assessed at the time women registered for prenatal care, during the first trimester. In this way, the possibility of recall bias is eliminated. However, we do lack exposure information over the course of pregnancy. Because it is unclear what the critical window of exposure is, we may have some misclassification of this time-varying exposure. For example, ~10% of smokers in Sweden cease cigarette smoking after the first antenatal care visit (19). Thus, if the relevant time window were later in pregnancy, we would have classified a small proportion of unexposed person-time as exposed. Moreover, the societal attitudes toward smoking may have led to underreporting of smoking during pregnancy. Because such misclassification of the exposure is nondifferential, the true associations between maternal smoking and leukemia and lymphoma may be greater than reported.

Because of the study design, there are few limitations to consider. The Medical Birth Register lacks information on some reported risk factors, such as exposure to ionizing radiation, parental occupation, and dietary data. These factors may have differed by maternal smoking status, thus leading to potential residual confounding. Of particular concern may be residual confounding by paternal smoking. Some studies suggest that, among

Table 4. Adjusted* HRs for the effect of maternal smoking on ALL stratified by age at diagnosis and sex, Sweden, January 1983–December 1997

	Age at diagnosis			Gender	
	0–1 y HR (95% CI)	2–4 y HR (95% CI)	≥5 y HR (95% CI)	Male HR (95% CI)	Female HR (95% CI)
Maternal smoking					
No	Reference	Reference	Reference	Reference	Reference
Yes	0.56 (0.31–1.01)	0.83 (0.62–1.11)	0.64 (0.42–0.97)	0.63 (0.46–0.86)	0.85 (0.62–1.16)
1–9 cigarettes	0.57 (0.28–1.15)	0.79 (0.55–1.13)	0.55 (0.32–0.95)	0.63 (0.43–0.92)	0.75 (0.50–1.10)
≥10 cigarettes	0.55 (0.22–1.37)	0.89 (0.59–1.35)	0.78 (0.44–1.40)	0.64 (0.40–1.02)	1.02 (0.66–1.57)
P for trend	0.071	0.33	0.10	0.008	0.59

*Data adjusted for maternal age, maternal education, maternal birthplace, parity, birth year, and baby's gender.

nonsmoking mothers, paternal smoking is associated with increased risk of ALL and lymphoma (16, 20). However, in Sweden, paternal smoking is closely associated with maternal smoking (19). Thus, if paternal smoking is associated with increased risk of ALL and NHL also in Sweden, we should have underestimated the protective effect of maternal smoking on ALL and overestimated the effect on NHL.

In this study, mean follow-up time of the cohort is ~8 years, and ~90% of the children were <15 years old at the end of the study. Thus, this study focused on cancers that occurred earlier in the cohort. This observation should be taken into consideration when assessing the generalizability of these findings to malignancies with later age at onset. If the *in utero* effects of smoking play a greater role on later rather than earlier onset cancers (15), then our effect estimates may not be directly applicable to the age groups under study. At least for ALL, our data do not suggest a different effect of smoking by age at diagnosis.

Our results agree with some, but not all, previous studies on the effect of maternal smoking on risk of childhood leukemia and lymphoma. The United Kingdom Cancer Study, which is a nationwide population-based case-control study, evaluated maternal smoking during the second trimester of pregnancy using structured interviews (15). The authors found that maternal smoking was associated with a 24% lower risk of leukemia (P for trend = 0.03). This protective effect was notable for both ALL and AML, however. A large, population-based case-control study undertaken in Germany assessed maternal smoking during the first trimester and found a protective effect for ALL and an increased risk for NHL (14). A meta-analysis based on eight studies, however, found no evidence of an effect of maternal smoking on leukemia (relative risk 1.05; CI, 0.82–1.34; ref. 21).

Few cohort studies examining maternal smoking and risk of childhood hemopoietic cancers have been undertaken. In a study including 54,795 live-born children, there was some evidence of a protective effect of maternal smoking on total leukemia, although the results were not statistically significant (22). In an initial follow-up for the Swedish birth cohort between 1982 and 1987, Pershagen et al. (23) reported no association between maternal smoking and cancers of the lymphatic and hemopoietic system (HR, 1.04; 95% CI, 0.71–1.52). However, in case-control studies nested within the cohort through 1989, there was evidence of a protective effect of ALL (13) and excess risks of AML (10) and NHL (11). Maternal smoking data from the Swedish nested case-control and cohort studies was derived in the same manner as the present study.

Given the inconclusiveness of earlier epidemiologic studies, we can turn to biological plausibility to assess the study findings. First, several components of cigarette smoke, such as benzo[*a*]pyrene and 4-aminobiphenyl, are known to cross the placental membrane and have been detected in the placenta and fetal blood of offspring (24-27). In addition, maternal smoking during pregnancy was positively associated with increased numbers of specific mutations such as deletions in lymphocytes of the offspring (28, 29). Thus, it is biologically plausible that maternal smoking during pregnancy increases the risk of NHL and AML, as observed in our study.

The protective effect of smoking and ALL is more difficult to understand, and little is known about the mechanism by which smoking could exert such an effect. In animal models in which progeny are exposed *in utero* to benzo[*a*]pyrene, a component of tobacco smoke, there is substantial evidence of generalized immune suppression after birth (30-32). In particular, *in utero* exposure to benzo[*a*]pyrene decreases prolymphocytic cells in animals (31) and suppresses B-cell lymphopoiesis and induces pre–B-cell apoptosis in bone marrow cultures (33). Such suppression of immune function could result in a decreased response and lower likelihood of clonal expansion.

Despite the apparent protective effect of smoking on ALL, this study in no way supports that maternal smoking is beneficial. Smoking during pregnancy is linked to several adverse effects, including fetal growth restriction, preterm birth, and perinatal mortality (33-35), outcomes that are significantly more common conditions. This evidence may simply outline a potential mechanism by which ALL could occur.

Clearly, the question of maternal smoking and risk of hemopoietic cancers remains. This study provides supportive evidence of positive associations with AML and NHL and an interesting protective effect with ALL, which needs to be explored further. With additional follow-up time, this unique cohort of Swedish children will help to further elucidate the role of maternal smoking on risk of childhood cancers.

References

1. Boffetta P, Trichopoulos D. Cancer of the lung, larynx and pleura. In: Adami HO, Hunter DH, Trichopoulos D, editors. Textbook of cancer epidemiology. New York: Oxford University Press; 2002. p. 248–80.
2. Kogevinas M, Trichopoulos D. Urinary bladder cancer. In: Adami HO, Hunter DH, Trichopoulos D, editors. Textbook of cancer epidemiology. New York: Oxford University Press; 2002. p. 446–66.
3. Nyrén O, Adami HO. Esophageal cancer. In: Adami HO, Hunter DH, Trichopoulos D, editors. Textbook of cancer epidemiology. New York: Oxford University Press; 2002. p. 137–61.
4. Mucci L, Adami HO. Oral and pharyngeal cancer. In: Adami HO, Hunter DH, Trichopoulos D, editors. Textbook of cancer epidemiology. New York: Oxford University Press; 2002. p. 115–36.
5. Petridou E, Trichopoulos D. In: Adami HO, Hunter DH, Trichopoulos D, editors. Textbook of cancer epidemiology. New York: Oxford University Press; 2002. p. 556–72.
6. Huncharek M, Kupelnick B, Klassen H. Maternal smoking during pregnancy and the risk of childhood brain tumors: a meta-analysis of 6566 subjects from twelve epidemiological studies. J Neurooncol 2002;57:51–7.
7. John EM, Savitz DA, Sandler DP. Prenatal exposure to parents' smoking and childhood cancer. Am J Epidemiol 1991;133:123–32.
8. Stjernfeldt M, Berglund K, Lindsten J, Ludvigsson J. Maternal smoking and irradiation during pregnancy as risk factors for child leukemia. Cancer Detect Prev 1992;16:129–35.
9. Severson RK, Buckley JD, Woods WG, Benjamin D, Robison LL. Cigarette smoking and alcohol consumption by parents of children with acute myeloid leukemia: an analysis within morphological subgroups—a report from the Children's Cancer Group. Cancer Epidemiol Biomarkers Prev 1993;2:433–9.
10. Cnattingius S, Zack M, Ekbom A, Gunnarskog J, Linet M, Adami HO. Prenatal and neonatal risk factors for childhood myeloid leukemia. Cancer Epidemiol Biomarkers Prev 1995;4:441–5.
11. Adami J, Glimelius B, Cnattingius S, et al. Maternal and perinatal factors associated with non-Hodgkin's lymphoma among children. Int J Cancer 1996;65:774–7.
12. Brondum J, Shu XO, Steinbuch M, Severson RK, Potter JD, Robison LL. Parental cigarette smoking and the risk of acute leukemia in children. Cancer 1999;85:1380–8.
13. Cnattingius S, Zack MM, Ekbom A, et al. Prenatal and neonatal risk factors for childhood lymphatic leukemia. J Natl Cancer Inst 1995; 87:908–14.

14. Schuz J, Kaatsch P, Kaletsch U, Meinert R, Michaelis J. Association of childhood cancer with factors related to pregnancy and birth. Int J Epidemiol 1999;28:631–9.
15. Pang D, McNally R, Birch JM. Parental smoking and childhood cancer: results from the United Kingdom Childhood Cancer Study. Br J Cancer 2003;88:373–81.
16. Shu XO, Ross JA, Pendergrass TW, Reaman GH, Lampkin B, Robison LL. Parental alcohol consumption, cigarette smoking, and risk of infant leukemia: a Children's Cancer Group Study. J Natl Cancer Inst 1996;88:24–31.
17. Cnattingius S, Ericson A, Gunnarskog J, Kallen B. A quality study of a medical birth registry. Scand J Soc Med Suppl 1990;18:143–8.
18. Cancer incidence in Sweden 1998. Stockholm: Center for Epidemiology, National Board of Health and Welfare, Sweden; 2000.
19. Cnattingius S, Lindmark G, Meirik O. Who continues to smoke while pregnant? J Epidemiol Community Health 1992;46:218–21.
20. Ji BT, Shu XO, Linet MS, et al. Paternal cigarette smoking and the risk of childhood cancer among offspring of nonsmoking mothers. J Natl Cancer Inst 1997;89:238–44.
21. Boffetta P, Tredaniel J, Greco A. Risk of childhood cancer and adult lung cancer after childhood exposure to passive smoke: a meta-analysis. Environ Health Perspect 2000;108:73–82.
22. Klebanoff MA, Clemens JD, Read JS. Maternal smoking during pregnancy and childhood cancer. Am J Epidemiol 1996;144:1028–33.
23. Pershagen G, Ericson A, Otterblad-Olausson P. Maternal smoking in pregnancy: does it increase the risk of childhood cancer? Int J Epidemiol 1992;21:1–5.
24. Everson RB, Randerath E, Santella RM, Cefalo RC, Avitts TA, Randerath K. Detection of smoking-related covalent DNA adducts in human placenta. Science 1986;231:54–7.
25. Everson RB, Randerath E, Santella RM, Avitts TA, Weinstein IB, Randerath K. Quantitative associations between DNA damage in human placenta and maternal smoking and birth weight. J Natl Cancer Inst 1988;80:567–76.
26. Myers SR, Spinnato JA, Pinorini-Godly MT, Cook C, Boles B, Rodgers GC. Characterization of 4-aminobiphenyl-hemoglobin adducts in maternal and fetal blood-samples. J Toxicol Environ Health 1996;47:553–66.
27. Arnould JP, Verhoest P, Bach V, Libert JP, Belegaud J. Detection of benzo[*a*]pyrene-DNA adducts in human placenta and umbilical cord blood. Hum Exp Toxicol 1997;16:716–21.
28. Ammenheuser MM, Berenson AB, Stiglich NJ, Whorton EB Jr, Ward JB Jr. Elevated frequencies of hprt mutant lymphocytes in cigarette-smoking mothers and their newborns. Mutat Res 1994;304:285–94.
29. Finette BA, O'Neill JP, Vacek PM, Albertini RJ. Gene mutations with characteristic deletions in cord blood T lymphocytes associated with passive maternal exposure to tobacco smoke. Nat Med 1998;4:1144–51.
30. Urso P, Zhang W, Cobb JR. Immunological consequences from exposure to benzo(*a*)pyrene during pregnancy. Scand J Immunol Suppl 1992;11:203–6.
31. Holladay SD, Smith BJ. Fetal hematopoietic alterations after maternal exposure to benzo[*a*]pyrene: a cytometric evaluation. J Toxicol Environ Health 1994;42:259–73.
32. Rodriguez JW, Kirlin WG, Wirsiy YG, Matheravidathu S, Hodge TW, Urso P. Maternal exposure to benzo[*a*]pyrene alters development of T lymphocytes in offspring. Immunopharmacol Immunotoxicol 1999;21:379–96.
33. Hardin JA, Hinoshita F, Sherr DH. Mechanisms by which benzo[*a*]-pyrene, an environmental carcinogen, suppresses B cell lymphopoiesis. Toxicol Appl Pharmacol 1992;117:155–64.
34. Cnattingius S, Mills JL, Yuen J, Eriksson O, Salonen H. The paradoxical effect of smoking in preeclamptic pregnancies: smoking reduces the incidence but increases the rates of perinatal mortality, abruptio placentae, and intrauterine growth restriction. Am J Obstet Gynecol 1997;177:156–61.
35. U.S. Surgeon General. Health consequences of tobacco use among women, In: Smoking and women. A report of the Surgeon General; US Department of Health and Human Services, Washington DC. 2001. p. 277–307.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Maternal Smoking and Childhood Leukemia and Lymphoma Risk among 1,440,542 Swedish Children

Lorelei A. Mucci, Frederik Granath and Sven Cnattingius

Cancer Epidemiol Biomarkers Prev 2004;13:1528-1533.

Date of last revision: September 21, 2015

TITLE

An evaluation of glyphosate use and the risk of non-Hodgkin lymphoma major histological sub-types in the North American Pooled Project (NAPP)

AUTHORS AND AFFILIATIONS

Manisha Pahwa[1], Laura Beane Freeman[2], John J. Spinelli[3,4], Paul A. Demers[1,5], Aaron Blair[2], Punam Pahwa[6,7], James A. Dosman[6], John R. McLaughlin[5,8], Shelia Hoar Zahm[2], Kenneth P. Cantor[2], Dennis D. Weisenburger[9], Shelley A. Harris[1,5,10]

1. Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Canada
2. Division of Cancer Epidemiology and Genetics, U.S. National Cancer Institute, Bethesda, U.S.
3. British Columbia Cancer Agency Research Centre, Vancouver, Canada
4. School of Population and Public Health, University of British Columbia, Vancouver, Canada
5. Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
6. Canadian Centre for Health and Safety in Agriculture, University of Saskatchewan, Saskatoon, Canada
7. Community Health and Epidemiology, College of Medicine, University of Saskatchewan, Saskatoon, Canada
8. Public Health Ontario, Toronto, Canada
9. City of Hope, Duarte, U.S.
10. Prevention and Cancer Control, Cancer Care Ontario, Toronto, Canada

TARGET JOURNAL

Occupational and Environmental Medicine

ARTICLE TYPE

Original article

KEY TERMS

Glyphosate; Lymphoma, Non-Hodgkin; Pesticides; Case-Control Studies

CORRESPONDING AUTHOR

Manisha Pahwa, Occupational Cancer Research Centre, Cancer Care Ontario, 620 University Avenue, Toronto, Ontario, Canada, M5G 2L7, Phone: 416-971-9800 ext. 3321, Fax: 416-971-6888, E-mail: manisha.pahwa@occupationalcancer.ca

WORD COUNT

Abstract: limit 250 (count 249)

Text: limit 4500 (count 6222)

Date of last revision: September 21, 2015

TABLE COUNT

In manuscript: limit 5 (count 4)

In supplement: no limit (count 2)

FIGURE COUNT

In manuscript: no limit (count 1)

In supplement: no limit (count 0)

REFERENCE COUNT

Limit 40 (count 33)

WHAT THIS PAPER ADDS

- Exposure to glyphosate, a broad-spectrum and frequently used herbicide, may be associated with non-Hodgkin lymphoma (NHL). Little is known about how risks may differ by glyphosate exposure levels and NHL sub-types.
- To address this research gap, this analysis integrated detailed, self-reported glyphosate use information with assessments of NHL risk overall and by major histological sub-type using pooled data from 1690 NHL cases and 5131 controls from the U.S. Midwest and Canada.
- Subjects who ever used glyphosate had elevated odds ratios for NHL overall and for all subtypes except follicular lymphoma. Significant or nearly significant risks of NHL overall were observed for >2 days per year (OR=2.42, 95% CI: 1.48, 3.96) and >7 lifetime days (OR=1.55, 95% CI: 0.99, 2.44) of glyphosate use, with some differences in risk by sub-type.
- Glyphosate use may be associated with elevated NHL risk. Although the pattern of risks was not clear across exposure categories, these findings from a large dataset offer more precision than results from previous studies.

Date of last revision: September 21, 2015

ABSTRACT (249)

Objectives: Glyphosate is the most frequently used herbicide worldwide. Some epidemiological studies have found positive associations between glyphosate exposure and non-Hodgkin lymphoma (NHL). This study aimed to evaluate NHL risk overall and by major histological sub-type using detailed glyphosate use metrics.

Methods: The NAPP, composed of pooled case-control studies from the U.S. and Canada, includes NHL cases (N=1690) and controls (N=5131) who provided information on pesticide use. Cases (follicular lymphoma [FL], diffuse large B-cell lymphoma [DLBCL], small lymphocytic lymphoma [SLL], other) from cancer registries and hospitals were frequency-matched to population-based controls. Logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI) by ever/never, duration, frequency, and lifetime days of glyphosate use. Models were adjusted for age, sex, location, proxy respondent, family history of lymphohematopoietic cancer, and personal protective equipment.

Results: Cases who ever used glyphosate (N=133) had a significantly elevated risk of NHL overall (OR=1.43, 95% CI: 1.11, 1.83). Subjects who used glyphosate for >3.5 years had increased SLL risk (OR=1.98, 95% CI: 0.89, 4.39) and those who handled glyphosate for >2 days/year had significantly elevated odds of NHL overall (OR=2.42, 95% CI: 1.48, 3.96) and DLBCL (OR=2.83, 95% CI: 1.48, 5.41). There were suggestive increases (p-trend ≤0.02) in risk of NHL overall, FL, and SLL with more days/year of glyphosate use.

Conclusions: Glyphosate use may be associated with increased NHL risk. Although risk differences by histological sub-type were not consistent across glyphosate use metrics, the NAPP's large sample size yielded more precise results than possible in previous studies.

Date of last revision: September 21, 2015

INTRODUCTION

Glyphosate [N-(phosphonomethyl)glycine] is a broad-spectrum herbicide that is one of the most frequently applied pesticides in the world. First developed commercially for agricultural use in the early 1970s, glyphosate quickly became a popular chemical; as of 2012, it was used in more than 750 products with an annual global production volume exceeding 600,000 tonnes (1). In the U.S., the highest levels of agricultural use occur in the mid-west on crops such as corn, soybeans, and wheat (2). These crops are also examples of the many different types of plants that have been genetically engineered to be resistant to glyphosate.

Commented [AB1]: Check to make sure all these crops have genetically modified seed on the market. I do not think that is the case for wheat yet. I think rice was to be available this year.

Glyphosate has been examined as a potential risk factor for lymphatic and hematopoietic cancers including non-Hodgkin lymphoma (NHL). In Canada, NHL ranks as the fifth most incident cancer in males following neoplasms of the prostate, colorectum, lung, and bladder (3). In the American mid-west NHL accounts for an unusually large number of cancers in agricultural areas where populations tend to have lower cancer rates overall (4). The causes of NHL are largely unknown (Hartge P, Wang SS, Bracci PM, Devesa SS, Holly EA. Non-Hodgkin Lymphoma. In Cancer epidemiology and Prevention, 3rd Edition. Shottenfeld D, Fraumeni JF, Jr. (Eds.). Oxford University Press, NY, Ny, 2006), pp. 898-918.). ~~Male NHL~~ has been associated with farming (Blair et al., 1992)~~gender, advanced age, and immune suppression are the best known risk factors. Agricultural exposures are hypothesized to be involved in the development of NHL~~ and this has prompted studies focused on pesticides.

Formatted: Superscript

In the 1980s and 1990s Four~~four~~ population-based case-control studies were conducted in the U.S. mid-west and six Canadian provinces to examine putative associations between agricultural exposures and pesticides and the risk of NHL. Individual study results showed positive associations between self-reported glyphosate use and NHL risk, although there was variation in the magnitude and statistical significance of risks between studies. In an analysis of the Canadian study the odds ratio [OR] for NHL was 1.26 (95% confidence interval [CI]: 0.87, 1.80) for the use of glyphosate with adjustment for age and province (N=51 exposed cases) (5). The OR was slightly higher from~~A similar risk estimate was found in a separate analysis of men who reportedly ever handled glyphosate in Iowa and Minnesota (6) and higher odds were calculated in~~ a pooled analysis that included 36 exposed male cases from Iowa, Minnesota, Kansas, and Nebraska (logistic regression OR=2.1, 95% CI: 1.1, 4.0 adjusted for age, study site, and other pesticides) (7).

Other studies involving glyphosate exposure and NHL risk have been conducted and many were included in a systematic literature review and meta-analysis of epidemiological studies of pesticide exposure and NHL risk (8). This meta-analysis found~~demonstrated~~ that glyphosate exposure was significantly associated with ~~elevated risks of~~ NHL ~~overall~~ (meta risk ratio [mRR]=1.5, 95% CI: 1.1-2.0, 6 papers). The OR for ~~and~~ B cell lymphoma, ~~(mRR=2.0, 95% CI: 1.1-3.6, 2 papers),~~ a commonly diagnosed NHL sub-type in the regions from which included studies were drawn, was (mRR=2.0, 95% CI: 1.1-3.6, 2 papers). ~~However, meta-analyses were based on a small number of included papers and each study contained low numbers of exposed subjects. Only one included study (9) reported risks by NHL sub-type and only three (5, 9, 10) reported risks by glyphosate exposure level.~~

Date of last revision: September 21, 2015

A comprehensive evaluation of glyphosate carcinogenicity was recently undertaken by the International Agency for Research on Cancer (IARC) (11). This review of mechanistic, animal, and epidemiological evidence classifiedled to the evaluation of glyphosate as a "probable" (group 2A) carcinogen for NHL based on limited evidence in humans and sufficient evidence in experimental animals. The assessment of limited evidence from epidemiological studies was based on case-control studiesprimarily focused on evidence from case-control studies of occupational glyphosate exposure in the U.S., Canada, and Sweden that reported increased risks of NHL that persisted after adjustment for other pesticides. No association between NHL and use of glyphosate was seend in the Agricultural Health Study (AHS), a large prospective study of farmers and commercial pesticide applicators in the U.S.(11). In bioassays, gGlyphosate was was associated with renal tubule carcinoma, pancreatic islet-cell adenoma, and skin tumors (11). able to cause different cancers in mice, postulated to occur through initiation and promotion. Mechanistic and other data supported the "probable" carcinogen conclusion by providing strong evidence for genotoxicity and oxidative stress, both of which are mechanisms of action that can take place in humans (11).

There are several research gaps that need to be addressed in order to better understand the role and impact of glyphosate exposure on the development ofcancer risk, specifically NHL. Individual studies often have limited power for glyphosate exposure, lack evaluation of NHL by sub-type, and do not adjust risk estimates for other pesticides and other exposures (8, 11). MAdditionally, most studies do not have quantitative exposure data needed to perform more sensitive epidemiological analyses and few have addressed potential effect modifiers to identify if glyphosate exposure has a different impact on NHL risk under certain circumstances. Schinasi and Leon (8) have suggested pooling studies as an attempt to overcome some of these limitations. ; AGRICOH, a consortium of agricultural cohorts, is a global effort of this kind (12). Other existing studies can be similarly leveraged for enhancing our knowledge and understanding about glyphosate exposure and NHL risk.

The North American Pooled Project (NAPP) is a pooled resource of population-based case-control studies previously conducted in the U.S. and Canada. The primary objective of this effortstudy was to provide larger numbers for more detailed analyses of possible relationships between NHL and pesticide use. In this paper we evaluate the association between glyphosate use and the risk of NHL among men and women in the NAPP. in the North American Pooled Project (NAPP), a pooled resource of population-based case-control studies previously conducted in the U.S. and Canada. NHL risk was assessed overall and by histological sub-type using detailed self-reported glyphosate use information and adjustment for other pesticides and possible risk factors. The secondary aim of this study was to examine the effects of personal protective equipment (PPE) on the association between glyphosate use and NHL risk overall.

METHODS

Study population

The NAPP is a large and newly established resource of pooling ofed data from four previously conducted case-control studies of men and women who were diagnosed with soft tissue sarcoma and lymphatic

Date of last revision: September 21, 2015

and hematopoietic cancers, including NHL, in the U.S. and Canada. NHL cases were recruited from cancer registries and hospitals during the 1980s in four states (Iowa, Minnesota, Kansas, and Nebraska) and between 1991 and 1994 in six provinces (Quebec, Ontario, Manitoba, Saskatchewan, Alberta, and British Columbia). Cases were 19 years of age or older in all jurisdictions (I think the 19 age cut is correct, just check each study to make sure). Controls were selected from the general population in each state or province. Selection procedures varied by study but included ~~by~~ random digit dialing, voters' lists, health insurance records, Medicare listings for those older than 65 years, and from state mortality files for deceased cases. Controls were matched to NHL cases in each state/province on the basis of age (±2 or 5 years). In some states, cases and controls were matched on the additional variables of sex (Nebraska), race (Nebraska), and vital status and year of death for deceased cases (Iowa, Minnesota, Nebraska, Kansas). All states and provinces included men; women were only included in Nebraska. Deceased cases and controls were eligible for inclusion in the U.S. ~~case-control~~ studies. The Canadian study only considered alive cases and controls. The present analysis used data from both men and women and from alive and deceased NHL cases (N=1690) and controls (N=5131).

Data collection

Participants, or surrogates, provided detailed information about demographic characteristics, pesticide use, agricultural exposures, and exposure to other known or suspected NHL risk factors including lifestyle, medical and occupational history. Interviewer-administered questionnaires were conducted by telephone (Kansas and Nebraska) or in person (Iowa and Minnesota) with cases and controls or their surrogates if subjects were deceased or too ill to respond themselves. In Canada, all cases and controls were mailed a questionnaire to complete themselves (or by their surrogates). Participants who indicated that they had used pesticides were subsequently interviewed over the telephone for details about their pesticide exposure. The Canadian questionnaire was modified from the telephone interview questionnaires that were used in Kansas and Nebraska. The questionnaires from all case-control studies were very similar since they shared a common research objective, involved overlapping groups of principal investigators, and were developed during the same time period. This made the data highly amenable to pooling ~~at present~~. The complete methodologies of each case-control study have been described by Cantor et al., 1992 (Iowa and Minnesota) (6), Hoar et al., 1986 (Kansas) (13), Zahm et al., 1990 (Nebraska) (14), and McDuffie et al., 2001 (Canada) (5).

The NAPP contains extensive information about pesticide use and agricultural exposures reported by cases and controls. In general, pesticide classifications are available from ~~data were collected beginning with the~~ broad~~est~~ categories (e.g. occupations with potential pesticide exposure), to ~~followed by~~ major chemical classes (e.g. herbicides), to chemical groups (e.g. phenoxy herbicides), and finally individual compounds (e.g. 2,4-D). For each individual compound reported, information was collected for dichotomous use (ever/never), duration of use (number of years), and frequency of personal handling (number of days/year). Duration data were not collected in Kansas and frequency information was not collected in Iowa, Minnesota, and Kansas ~~and Kansas~~. In Kansas participants were asked to open-endedly recall the details of their pesticide use whereas in all other jurisdictions subjects were prompted by a list of chemicals and their trade names. Participants were also asked to report if they had used any

type of PPE in general (Nebraska and Canada) and with herbicides (Iowa, Minnesota, and Kansas) and specific individual pesticides (Iowa and Minnesota).

Assessment of glyphosate use

Self-reported glyphosate use was examined using several different metrics: dichotomous, duration, frequency, and lifetime days (derived by multiplying number of years used with number of days/year handled). Ordinal categories were created for duration, frequency, and lifetime days analyses based on the median of glyphosate used/handled in controls. Since information about duration of glyphosate use was not collected in Kansas, cases and controls from Kansas were omitted from duration analyses. Similarly, cases and controls from Iowa, Minnesota, and Kansas were excluded from frequency and lifetime days analyses owing to the lack of frequency data collected in these states. Participants who had missing or unknown glyphosate use information, but who were from jurisdictions where glyphosate use information was collected, were coded as "never used" in dichotomous analyses. ~~;~~ ~~f~~For duration and frequency analyses, missing values were assigned based on the median duration or frequency by state/province, age, and NHL sub-type (simple imputation, rounded to the nearest whole number). Subjects who reported that they used glyphosate were coded as "ever used" or used/handled for the number of years and days/year that they had reported. Continuous analyses were also conducted in order to determine possible trends and changes in risk for every 5 years, 5 days/year, and 10 lifetime days of glyphosate use.

NHL classification

NHL cases in these s tudies were diagnosed ~~at different time periods~~ during the 1980s and 1990s. NHL cases were classified in Iowa, Minnesota, and Nebraska according to the Working Formulation (15, 16); in Kansas and Quebec by the International Classification of Diseases for Oncology First Edition (ICD-O-1) (1976) (17); and in Ontario, Manitoba, Saskatchewan, Alberta, and British Columbia by ICD-O-2 (1990) (18). The original histology codes used in each study were revisited to classify NHL cases using a single or similar scheme for the NAPP. We used ICD-O-1 to code NHL overall and sub-types in the NAPP since histological sub-types were classified in all jurisdictions according to ICD-O-1. These sub-types were follicular lymphoma (FL), diffuse large B cell lymphoma (DLBCL), small lymphocytic lymphoma (SLL), and other. The "other" sub-type included all cases whose histologies were unknown or not FL, DLBCL, or SLL. Pathology reviews were conducted on 84% of Canadian cases (5), 87% of Kansas cases (13), and for all interviewed cases in Iowa and Minnesota (6) and Nebraska (14) in order to validate NHL diagnoses.

Power and sample size

A power and sample size analysis was conducted using the U.S. National Cancer Institute's (NCI) Power Version 3.0 program (19, 20) by inputting the following parameters: number of controls = 5131; number of cases = 1690; control:case ratio = 3; type I error (two-sided) = 0.05; type II error = 0.2; probability of NHL at baseline = 0.04 (21).

Of all 5131 controls available in the NAPP, 244 (4.76%) reported that they ever used glyphosate. A 5% prevalence of pesticide exposure in controls corresponds to a~~perfect~~ power of ~~(~~1.00~~)~~ to detect ORs of

2.00 or higher and a~~, but lower~~ power of~~(0.46)~~ to detect an OR of 1.25. Given that approximately 5% of controls reported ever being exposed to glyphosate, at a power level of 0.80, a total of 1103 NHL cases would be required to detect an OR of 1.50 (Appendix 1). The numbers of NHL cases and controls in the NAPP appear to be suitable to~~for~~ detecting low to moderate relative risks associated with glyphosate exposure in this population.

Statistical analyses

Descriptive statistics were used to characterize the study population and identify potentially confounding variables. Based on previously published literature, a priori possible confounders included age, sex, state/province, use of a proxy respondent (5, 6, 22), lymphatic or hematopoietic cancer in a first-degree relative (23), and diagnosis with select medical conditions related to immune suppression (any allergies, food allergies, drug allergies, asthma, hay fever, mononucleosis, arthritis, or tuberculosis; ever received chemotherapy or radiation) (24-26). History of living or working on a farm or ranch was also evaluated as a potential confounder.

It was possible that the use of other pesticides in the NAPP may confound the relationship between glyphosate use and NHL risk. A two-pronged approach was used to identify potentially confounding by other pesticides. First, a correlation matrix of pooled data was produced to determine the presence and extent of correlation between glyphosate and each individual herbicide, insecticide, and fungicide reportedly used by NAPP subjects. Second, previously published articles based on the individual case-control studies comprising the NAPP were searched to identify any positive or significant relationships between individual pesticides and NHL risk, as would be required for confounding to occur. Pesticides that were most strongly correlated with glyphosate (defined in this study as Spearman coefficients ≥0.35 and Cohen's Kappa value ≥0.30) and that were significantly or strongly associated with NHL in previous studies were evaluated as confounders. These were the herbicides 2,4-D (2,4-dichlorophenoxyacetic acid) (5, 6) and dicamba (5, 7), as well as the insecticide malathion (5, 7).

The use of PPE with glyphosate could theoretically modify NHL risk by reducing subjects' exposure to glyphosate. Although such information was sought in some studies, data were on a sizable~~There was a large~~ proportion of the study subjects~~missing data for the more specific variables of PPE used for herbicides and glyphosate~~ and~~. Therefore,~~ effect modification analyses could only be conducted using ~~involving any lifetime PPE use were conducted using data reported by~~ cases and controls from~~in~~ Nebraska and Canada. Any lifetime PPE usage was also included as a confounding variable in models where it was not evaluated as a possible effect modifier.

Unconditional multiple logistic regression was performed using the LOGISTIC procedure of~~on~~ the SAS 9.2 statistical software package (SAS Institute, Cary, North Carolina) to calculate pooled ORs and 95% CIs for associations between glyphosate exposure (dichotomous, duration, frequency, lifetime days, and as a continuous variable) and the risk of NHL overall and by histological sub-type (FL, DLBCL, SLL, and other). Primary logistic regression models (OR[a]) contained the following variables as confounders: age, sex, state/province, lymphatic or hematopoietic cancer in a first-degree relative, use of a proxy respondent, and use of any PPE. Secondary logistic regression models (OR[b]) contained the covariates in the primary

model plus reported use of the pesticides 2,4-D, dicamba, and malathion. Medical conditions and history of living or working on a farm or ranch were found not to~~did not appear to play a role in~~ confound~~ing~~ the relationship between glyphosate use and NHL risk and were not included in the models. Use-response trends for duration, frequency, and lifetime days analyses were deemed to be statistically significant if the two-sided p-value for the ordinal glyphosate use category was ≤0.05. The reference group for all analyses was subjects who never used glyphosate. There was a small proportion of subjects (N=175, 2.57% of all participants) with missing age values; these were imputed based on state/province- and case/control-specific means rounded to the nearest whole number.

Sensitivity tests were conducted by excluding proxy respondents from the main analyses. Proxy respondents were excluded from the analyses of PPE as a potential effect modifier in order to minimize the possibility of bias. For the effect modification analyses, glyphosate use was classified dichotomously and by duration, frequency, and lifetime days and overall NHL risks were calculated using logistic regression models adjusted for age, sex, state/province, lymphatic or hematopoietic cancer in a first-degree relative, and use of 2,4-D, dicamba, and malathion.

Ethics approval

Approval to conduct this analysis was obtained from the University of Toronto Health Sciences Research Ethics Board (#25166) and an ethics exemption was obtained from the U.S. NCI Office of Human Subjects Research (#11351). Individual studies had obtained human subjects approval prior to collection of the data and aAll participants provided informed consent before taking part in the studies included in the NAPP analyses.

RESULTS

Characteristics of NHL cases and controls

A total of 1690 NHL cases and 5131 controls were available in the NAPP for analysis. All participants were included in analyses that encompassed proxy respondents. For assessments involving the duration of glyphosate use, 1520 cases and 4183 controls were available; in frequency and lifetime days analyses, 898 cases and 2938 controls were included. The numbers of cases and controls available for the sensitivity analyses excluding proxy respondents were smaller~~lower~~ (Figure 1).

The most frequently diagnosed histological sub-type was DLBCL (38.28%), followed by FL (27.69%), other (23.91%), and SLL (10.12%) (Table 1). Nebraska yielded the highest proportion of cases (22.78%) and controls (27.91%) compared to other states and provinces. The average ages of cases and controls were 62.72 and 61.66 years, respectively. The majority of subjects were male. A similar proportion of proxy respondents were used by cases and controls. Cases were more than twice as likely to report that a first-degree relative was diagnosed with lymphatic or hematopoietic cancer compared to controls (OR=2.13, 95% CI: 1.69, 2.67). Medical history variables were evaluated as potential confounders but they did not have an appreciable impact on adjusted ORs in the main analyses (OR[a] and OR[b]) and were thus excluded from logistic regression models.

Date of last revision: September 21, 2015

Missing glyphosate use data

There were 7 cases with missing values for the number of years of glyphosate used and 13 cases with missing values for the number of days/year of glyphosate handled in the jurisdictions where duration and frequency of glyphosate use data were collected. The median values for the number of years of glyphosate use in cases all subjects with missing values ranged from 0-2 based on jurisdiction, NHL sub-type, and age. The median value for days/year for subjects with missing information was 0 (zero).

Glyphosate use and NHL risks overall and by major histological sub-type

Overall, 113/1690 cases (6.69%) and 244/5131 (4.76%) controls reported that they had used glyphosate at any point in their lifetime. There was a significant association between glyphosate use and the risk of NHL overall (OR^a=1.43, 95% CI: 1.11, 1.83) (Table 2). Risks were elevated for most NHL sub-types but the magnitude of risk differed by sub-type. The greatest risk was observed in SLL cases (OR^a=1.77, 95% CI: 0.98, 3.22) and the lowest risk was found for FL (OR^a=1.00, 95% CI: 0.65, 1.54). Similar and significant excesses were observed for DLBCL (OR^a=1.60, 95% CI: 1.12, 2.29) and other (OR^a=1.66, 95% CI: 1.04, 2.63) sub-types. Associations were attenuated and no longer statistically significant when the model represented by OR^a was further adjusted for ever use of 2,4-D, dicamba, and malathion (OR^b). The odds of SLL did not change even after adjusting risk estimates for these three pesticides.

When glyphosate use was examined by duration (Table 2), there was a general inverse trend in risks except for cases of SLL, where the odds increased with longer duration of glyphosate use (OR^a=1.98, 95% CI: 0.89, 4.39 for >3.5 years versus OR^a=1.49, 95% CI: 0.63, 3.58 for >0 and ≤3.5 years) and this trend was of borderline statistical significance (p-trend for OR^a=0.08). Additional adjustment for the chemicals 2,4-D, dicamba, and malathion generally resulted in attenuated risk estimates (OR^b) compared to models unadjusted for these pesticides (OR^a) except for SLL, for which the addition of these agents in logistic regression models had no substantial effect on risk (e.g. for >3.5 years of glyphosate use, OR^b=1.94, 95% CI: 0.79, 4.80).

In contrast to duration of glyphosate use, a more consistent pattern of NHL risk emerged in association with frequency of glyphosate personally handled (Table 2). Subjects who handled glyphosate for >2 days/year had NHL risks that were approximately two times the odds observed in participants who handled glyphosate for >0 and ≤2 days/year. This finding was consistent for NHL overall and all sub-types. Elevated risks in the highest category (>2 days/year) were significant for NHL overall (OR^a=2.42, 95% CI: 1.48, 3.96) and DLBCL (OR^a=2.83, 95% CI: 1.48, 5.41) compared to subjects who did not handle glyphosate at all. Significant trends in risk were also found for NHL overall (p-trend for OR^a=0.02) and DLBCL (p-trend for OR^a=0.04). For NHL overall and DLBCL, ORs associated with handling glyphosate for >2 days/year were attenuated but remained statistically significant even after adjusting for the use of 2,4-D, dicamba, and malathion. The pattern of increased risks with more frequent glyphosate handling was still apparent for NHL overall and all sub-types although trends were no longer statistically significant upon adjusting for these three pesticides.

The analysis of lifetime days, derived from the product of number of years used and days/year handled, generally showed risk increases for NHL overall and most sub-types (except "other") in association with

a greater number of lifetime days of glyphosate use (Table 2). These trends were significant for NHL overall (p-trend for OR[a]=0.02), FL (p-trend for OR[a]=0.02), and SLL (p-trend for OR[a]=0.01). There were elevated risks of NHL among participants who had used glyphosate for >7 lifetime days; this was most pronounced for SLL (OR[a]=2.13, 95% CI: 0.76, 5.96). Adjusting for 2,4-D, dicamba, and malathion attenuated risks compared to odds that were unadjusted for these chemicals; however, the general pattern of increased risks remained intact and in some cases (i.e. SLL), was still statistically significant (p-trend for OR[b]=0.03).

Sensitivity analysis

Proxy respondents were used for deceased cases and controls and for alive cases who were too ill to respond to the case-control study questionnaires themselves. The use of proxy respondents might have introduced misclassification of glyphosate use. To account for this possibility, glyphosate use data provided by proxy respondents were excluded from the main analysis presented in Table 2. This generally resulted in reduced ORs compared to risks that included data provided by both self- and proxy respondents, with little effect on the width of confidence intervals and the same general patterns of risks for dichotomous, duration, frequency, and lifetime days analyses (Table 3). For instance, there were significant trends for lifetime days of glyphosate use and the risks of NHL overall (p-trend for OR[a]=0.04), FL (p-trend for OR[a]=0.03), and SLL (p-trend for OR[a]=0.01) (Table 3) that paralleled the trends found in the analysis of data provided by both self- and proxy respondents (Table 2).

However, there were some exceptions to this overall observation. Odds ratios for SLL mostly strengthened with the exclusion of proxy respondents in models both unadjusted for 2,4-D, dicamba, and malathion and models adjusted for these chemicals. For instance, among subjects who ever used glyphosate the risk of SLL excluding data from proxy respondents was 1.89 (OR[a], 95% CI: 1.03, 3.49) which was slightly greater than the risk of SLL based on data provided by self- and proxy respondents (OR[a]=1.77, 95% CI: 0.98, 3.22). Trends of increasing risk of SLL in association with longer duration, greater frequency and lifetime days of glyphosate use were also marginally stronger when data from proxy respondents were excluded.

Effect of PPE

Potential effect modification by PPE usage was evaluated based on data pooled from Canadian and Nebraskan participants. The association between ever glyphosate use and NHL risk overall was generally higher among subjects who reportedly used any type of PPE in their lifetime (OR=0.83, 95% CI: 0.40, 1.73) compared to subjects who never used any type of PPE (OR=0.65, 95% CI: 0.31, 1.35) (Table 4). This pattern of elevated NHL risks in subjects who ever used PPE compared to subjects who never used PPE persisted when glyphosate use was also evaluated by duration, frequency, and lifetime days. Similar to the results in Tables 2 and 3, there were inverse associations between the duration of glyphosate use and NHL risk and positive (increasing) associations between frequency and lifetime days of glyphosate use and NHL risk, regardless of PPE use status. There were many subjects with unknown or missing PPE use information and they were separately modeled in order to reduce the possibility of analyzing

misclassified PPE use data. Risks were high and unstable in this latter group due to the small number of subjects in each glyphosate usage category.

DISCUSSION

The objective of this study was to evaluate potential associations between glyphosate use and NHL risk in the NAPP, a large pooled dataset with detailed information about glyphosate use reported by 1690 NHL cases and 5131 controls. Glyphosate use was associated with elevated NHL risk, a finding that was consistent with previous analyses. Odds somewhat differed by histological sub-type, although there wasn't a consistent pattern across glyphosate use metrics. The novelty of this analysis and increased precision of risk estimates compared to smaller individual studies were major strengths. Yet, the limitations of this study illustrate the need for more research that can better characterize the relationship between glyphosate exposure and the development of NHL.

This report confirms previous analyses indicating increased risks of NHL in association with glyphosate exposure. The odds of NHL for glyphosate use was 1.43 (OR[a], 95% CI: 1.11, 1.83), a value that was situated approximately in between the risks observed in earlier analyses of the Canadian study (OR=1.26, 95% CI: 0.87, 1.80, adjusted for age and province, N=51 exposed cases) (5) and the three pooled U.S. studies (logistic regression OR=2.1, 95% CI: 1.1, 4.0, adjusted for age, study site, and other pesticides, N=36 exposed cases) (7). Further adjusting OR[a] for the pesticides 2,4-D, dicamba, and malathion resulted in an attenuated risk of NHL overall in the NAPP (OR[b]=1.13, 95% CI: 0.84, 1.51). De Roos et al. (2003) (7) used a more conservative approach, a hierarchical regression model, for assessing NHL risk in the three U.S. pooled case-control studies and found that this reduced the odds of NHL overall (OR=1.6, 95% CI: 0.9, 2.8, adjusted for age, study site, and other pesticides). A statistically significant excess of NHL was found in association with more than 2 days per year of use (OR=2.12, 95% CI: 1.20, 3.73) (5) in the Canadian study, a finding that was in agreement with our analogous pooled risk estimate for NHL (OR[a]=2.42, 95% CI: 1.48, 3.96).

Our results are also aligned with findings from epidemiological studies of other populations that found an elevated risk of NHL for glyphosate exposure and with a greater number of days/year of glyphosate use (9), as well as a meta-analysis of glyphosate use and NHL risk (8). From an epidemiological perspective, our results were supportive of the IARC evaluation of glyphosate as a probable (group 2A) carcinogen for NHL (11).

The large sample size of the NAPP was conducive to analyzing NHL risks with different metrics of glyphosate use. Evaluations of dichotomous glyphosate use showed nearly universal increases in risks of NHL overall and by sub-type, but results were more varied upon further examination by duration, frequency, and lifetime days. The odds of NHL, overall and by sub-type, were higher among subjects who reportedly used glyphosate more often in a year or who had greater cumulative use in their lifetime compared to unexposed subjects. Subjects who used glyphosate reported mostly initiating its use in the year 1980. Glyphosate was used by cases and controls for an average of 5 years and handled for an average of 5 days/year. The short duration of use made it challenging to calculate risks associated with longer-term usage, although the mean frequency of handling was typical of how often farmers

reportedly apply glyphosate to agricultural crops (27). For the days/year and lifetime days analyses some trends and risks were statistically significant while others were not, likely due to the lack of sufficient numbers of exposed cases for some sub-types.

There were some differences in risks by sub-type but these were not consistent between the different glyphosate use metrics and were unlikely to be statistically significant. For example, the significant trends observed for lifetime days of glyphosate use and the risks of NHL overall, FL, and SLL were not present for the frequency analysis, where significant trends were only found for NHL overall and DLBCL. In the duration analysis an upward trend was observed for SLL but not for any of the other sub-types or for NHL overall. Despite these uneven results the risks of FL were consistently lower than other sub-types in association with any of the glyphosate use metrics. There was a relatively large number of FL cases in this analysis compared to the numbers available for other sub-types, lessening the likelihood that findings for FL were primarily due to chance. FL is a type of B-cell lymphoma that is the second most common type of NHL, accounting for 22% of all NHLs (28). The observation of lowered FL risks for glyphosate use in this study was a lead for further evaluation. Additionally, the classification of NHL has changed since the case-control studies in the NAPP were conducted. Multiple myeloma is now considered a sub-type of NHL but was not evaluated in this analysis.

A fairly consistent decrease in NHL risk was found when ORs were further adjusted for the pesticides 2,4-D, dicamba, and malathion. This observation suggested that elevated risks of NHL may be attributed, in part, to pesticides other than glyphosate. Formulations of glyphosate reported by NAPP subjects may have contained other active ingredients. In addition or alternatively, glyphosate may have been used in combination with other pesticide active ingredients at the time of application or in the same growing season or year. It is relatively unknown how combinations of pesticides might interact, and we were not able to evaluate this in our analysis. There is a need to further investigate other individual compounds with respect to NHL risk, such as the herbicide 2,4-D, which IARC recently assessed as possibly carcinogenic to humans based on inadequate evidence in humans and limited evidence in animals for NHL (29).

Glyphosate and covariate data provided by self-respondents generally resulted in attenuated risks compared to odds derived from information provided by both self- and proxy respondents. The proportion of proxy respondents used for cases and controls was similar (about one third). Excluding proxies appreciably reduced the numbers of subjects in the sensitivity analysis which might have partly explained differences in risks. There was also the possibility of exposure misclassification by proxy respondents due to inaccurate recall of glyphosate use, which was likely non-differential (27, 30). Non-differential pesticide exposure misclassification was also an issue amongst self-respondents (31). There was less agreement between self-respondents and surrogates for detailed glyphosate use metrics (years and days/year) compared to the dichotomous variable (32). Nevertheless, significant trends of increasing risks in association with greater lifetime days of glyphosate use persisted for NHL overall, FL, and SLL, even when the analysis was limited to self-respondents.

The evaluation of PPE as an effect modifier of the relationship between glyphosate use and overall NHL risk raised some interesting observations. We expected that the use of any PPE such as masks, gloves,

clothing and/or other equipment may confer a protective effect on the development of NHL from glyphosate use by reducing the probability and degree of dermal, respiratory, and oral contact with glyphosate. However, in this study PPE was found to have no effect on the association between glyphosate use and NHL risk overall. This analysis was limited because PPE usage was not specific to glyphosate use or the type or timing of PPE worn. It was also based on pooled data from Canada and Nebraska only and there was a large proportion of missing data. This hypothesis warrants further investigation in larger studies with more information about PPE used with glyphosate in particular.

The exact causes of lymphatic and hematopoietic cancers are not yet known. A suppressed immune system is the most well established risk factor for NHL. It has been hypothesized that pesticides may play a role in modifying immune function (24-26), but there is little evidence to support this hypothesis for glyphosate specifically (11, 25). An alternative or additional explanation is that pesticides may influence the risk of lymphatic and hematopoietic cancers through pathways involving oxidative stress and receptor-mediated mechanisms. The pathway that glyphosate affects in plants is not present in mammals, but there is strong evidence from mechanistic studies that glyphosate causes genotoxicity and the production of reactive oxygen species (11).

The limitations of this study were primarily related to statistical power for some analyses and the possibility of biases and unmeasured confounding. We endeavoured to use data from all subjects for this analysis as reflected by the inclusion of both men and women and alive and deceased subjects. In Canada alone, 50 NHL cases and 133 controls reported ever using glyphosate; pooling resulted in an additional 63 NHL cases and 111 controls who ever used glyphosate in Iowa, Minnesota, Kansas, and Nebraska. Nevertheless, there were small numbers for some categories of duration, frequency, and lifetime days by NHL sub-type due to the absence of duration data collected in Kansas and frequency and lifetime days information from Iowa, Minnesota, and Kansas. Risk estimates based on small numbers may be unstable and could represent chance findings.

To evaluate possible recall bias of self-reported pesticide use, in the study in Kansas, pesticide suppliers were asked to provide information on crops and pesticide purchases for a sample of 130 subjects with farming experience (13, 27). In the Iowa and Nebraska studies, case recall bias was assessed by comparing information on pesticides used that was volunteered versus information that required probing by the interviewer (14, 27, 33). In the Iowa and Minnesota study, interviews were conducted with both farmers and their wives for a sample of subjects (32). There was a moderate level of correspondence between pesticide use information reported by farmers and their pesticide suppliers in Kansas (13, 27). In Iowa and Nebraska, the number of insecticides and herbicides voluntarily identified was similar and suggested the absence of case-response bias, but probing increased the number of positive responses for individual agents (14, 27, 33). In Iowa and Minnesota, surrogate responders were generally a poorer source of information compared to farmers as they had reported a smaller number of pesticides ever used and a greater proportion of "I don't know" answers (32). No similar analysis of recall bias has been conducted in the Canadian case-control study, but the similarity of study designs between the U.S. and Canada make it likely that recall bias is not a major concern in the Canadian study and NAPP as a whole.

Adjusting for several pesticides (2,4-D, dicamba, and malathion) was a useful way to attempt to disentangle the effect of glyphosate from other pesticides on NHL risk. These agents have been shown to be independently associated with NHL in individual case-control studies (5-7). However, they are somewhat correlated with glyphosate exposure in the NAPP and thus their inclusion as confounders may have introduced some degree of collinearity. Unmeasured confounding by other pesticides, agricultural exposures, or unknown factors cannot be ruled out.

While these results are not independent from previous studies, the evaluations by histological sub-type and for detailed glyphosate use metrics are a new and important contribution to the epidemiological literature. NHL is a constellation of heterogeneous cancers that each has its own causes, risk factors, and etiologies. Pesticides, including individual agents such as glyphosate, may exert different effects on these sub-types, and the large size of the NAPP made it possible to parse this out.

The large sample size also resulted in more precise results than possible in previous smaller studies that only had sufficient power to assess risks for dichotomous glyphosate exposure. We were able to model different glyphosate use categories and identify potential trends in NHL risk by sub-type with increasing duration, frequency, and lifetime days of glyphosate use. This made it possible to characterize possible dose-response relationships between glyphosate exposure and lymphoma risk. The effect modification analysis by PPE further allowed an examination of factors that might modify glyphosate exposure (and risk). Both agricultural and non-agricultural uses of glyphosate were reported by cases and controls in this population-based, pooled case-control study, making this evaluation externally valid.

The results of this analysis may be considered in future scientific and regulatory reviews of glyphosate in North America and globally. Stakeholders may also use these results as part of future approaches that communicate the health risks of pesticides using information directly ascertained from the North American population. This will help to inform efforts aimed at mitigating occupational and environmental exposure to pesticides. It will also provide high-quality risk estimates that can be used in future estimations of the burden of cancer from pesticide exposure.

ACKNOLWEDGEMENTS

The authors thank Mr. Joe Barker at IMS, Inc., for his computer programming services to pool the case-control datasets. The authors also thank all of the principal investigators of the individual case-control studies for allowing the data to be pooled.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING

This analysis was conducted with the support of a Prevention Research Grant from the Canadian Cancer Society Research Institute (#703055). There was no involvement in conducting the NAPP, preparing this article, or deciding to submit this paper for publication.

Date of last revision: September 21, 2015

AUTHORS' CONTRIBUTION

MP designed and conducted this analysis and wrote this manuscript. SAH, JJS, and LBF collectively form the NAPP Executive Committee and approved the proposal for this analysis and provided scientific input during the analytic and manuscript preparation phases. AB, SHZ, DDW, and KPC led the original case-control studies in the U.S. JJS, JAM, and JAD were among the principal investigators of the CCSPH in Canada. All co-authors reviewed and approved this manuscript for submission.

DATA SHARING

Unpublished NAPP data is available upon formal request to the NAPP Executive Committee (SAH, JJS, LBF).

REFERENCES

1. Research Report on Global and China Glyphosate Industry, 2013-2017. Available at: http://www.researchandmarkets.com/research/ssn6g8/research_report [Accessed August 18, 2015].

2. United States Geological Survey (USGS). Pesticide use maps – glyphosate. Pesticide National Synthesis Project. 2011. Available at: http://water.usgs.gov/nawqa/pnsp/usage/maps/show_map.php?year=2011&map=GLYPHOSATE&hilo=L&disp=Glyphosate [Accessed October 27, 2014].

3. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2015. Toronto: Canadian Cancer Society, 2015.

4. U.S. National Cancer Institute. GIS Portal: Animated Historical Cancer Atlas. https://gis.cancer.gov/atlas/index.php?geo=United_States&state=99&year=5&cancer=Non-Hodgkin_Lymphoma&gender=m&color=ryb

5. Blair A, Zahm SH, Pearce NE, Heineman EF, Fraumeni JF Jr. Clues to cancer etiology from studies of farmers. Scand J Work Environ Health 1992;18:209-15.

~~4.~~

~~5.~~6. 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 Epidemiology, Biomarkers & Prevention 2001;10:1155-1163.

~~6.~~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 Research 1992;52:2447-2455.

Formatted: No underline, Font color: Auto

Formatted: Level 1, Tab stops: -0.75", Left + -0.5", Left + 0", Left + 0.44", Left + 0.88", Left + 1.38", Left + 1.63", Left + 2", Left + 3", Left + 3.5", Left + 4", Left + 4.5", Left + 5", Left + 5.5", Left + 6", Left + 6.5", Left + 7", Left + 7.5", Left + 8", Left + 8.5", Left + 9", Left + 9.5", Left + 10", Left + 10.5", Left + 11", Left + 11.5", Left + 12", Left + 12.5", Left + 13", Left

Formatted: English (United States)

Date of last revision: September 21, 2015

~~7.~~8. De Roos AJ, Zahm SH, Cantor KP, Weisenburger DD, Holmes FF, Burmeister LF. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. Occupational and Environmental Medicine 2003;60:e11.

~~8.~~9. Schinasi L, Leon ME. Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. International Journal of Environmental Research and Public Health 2014;11:4449-4527.

~~9.~~10. Eriksson M, Hardell L, Carlberg M, Akerman M. Pesticide exposure as a risk factor for non-Hodgkin lymphoma including histopathological subgroup analysis. International Journal of Cancer 2008;123:1657-1663.

~~10.~~11. De Roos AJ, Blair A, Rusiecki JA, Hoppin JA, Svec M, Dosemeci M, Sandler DP, Alavanja MC. Cancer incidence among glyphosate-exposed pesticide applicators in the Agricultural Health Study. Environmental Health Perspectives 2005;113:49-54.

~~11.~~12. International Agency for Research on Cancer (IARC). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 112: Some Organophosphate Insecticides and Herbicides: Diazinon, Glyphosate, Malathion, Parathion, and Tetrachlorvinphos. Lyon: WHO Press, 2015.

~~12.~~13. Leon ME, Beane Freeman LE, Douwes J, Hoppin JA, Kromhout H, Lebailly P, Nordby KC, Schenker M, Schüz J, Waring SC, Alavanja MC, Annesi-Maesano I, Baldi I, Dalvie MA, Ferro G, Fervers B, Langseth H, London L, Lynch CF, McLaughlin J, Merchant JA, Pahwa P, Sigsgaard T, Stayner L, Wesseling C, Yoo KY, Zahm SH, Straif K, Blair A. AGRICOH: A consortium of agricultural cohorts. International Journal of Environmental Research and Public Health 2011;8:1341-1357.

~~13.~~14. Hoar SK, Blair A, Holmes FF, Boysen CD, Robel RJ, Hoover R, Fraumeni JF. Agricultural herbicide use and risk of lymphoma and soft-tissue sarcoma. Journal of the American Medical Association 1986;256:1141-1147.

~~14.~~15. Zahm SH, Weisenburger DD, Babbitt PA, Saal RC, Vaught JB, Cantor KP, Blair A. A case-control study of non-Hodgkin's lymphoma and the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D) in eastern Nebraska. Epidemiology 1990;1:349-356.

~~15.~~16. Dick FR, Van Lier SF, McKeen K, et al. Non-concurrence in abstracted diagnosis of non-Hodgkin's lymphoma. Journal of the National Cancer Institute 1987;78:675-678.

~~16.~~17. Non-Hodgkin's lymphoma pathologic classification project. National Cancer Institute sponsored study of classification of non-Hodgkin's lymphomas: summary and description of a working formulation for clinical usage. Cancer 1982;49:2112-2135.

~~17.~~18. International Classification of Diseases for Oncology, first edition. Geneva, World Health Organization, 1976.

Date of last revision: September 21, 2015

~~18.~~19. International Classification of Diseases for Oncology, second edition. Geneva, World Health Organization, 1990.

~~19.~~20. Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. American Journal of Epidemiology 1990;131:552-566.

~~20.~~21. García-Closas M, Lubin JH. Power and sample size calculations in case-control studies of gene-environmental interactions: Comments on different approaches. American Journal of Epidemiology 1999;149:689-693.

~~21.~~22. Ellison LF, Wilkins K. Cancer prevalence in the Canadian population. Statistics Canada Health Reports 2009;20:1-13.

~~22.~~23. Hohenadel K, Harris SA, McLaughlin JM, Spinelli JJ, Pahwa P, Dosman JA, Demers PA, Blair A. Exposure to multiple pesticides and risk of non-Hodgkin lymphoma in men from six Canadian provinces. International Journal of Environmental Research and Public Health 2011;8:2320-2330.

~~23.~~24. McDuffie HH, Pahwa P, Karunanayake CP, Spinelli JJ, Dosman JA. Clustering of cancer among families of cases with Hodgkin lymphoma (HL), multiple myeloma (MM), non-Hodgkin's lymphoma (NHL), soft tissue sarcoma (STS) and control subjects. BMC Cancer 2009;9:70.

~~24.~~25. Pahwa M, Harris SA, Hohenadel K, McLaughlin JR, Spinelli JJ, Pahwa P, Dosman JA, Blair A. Pesticide use, immunologic conditions, and risk of non-Hodgkin lymphoma in Canadian men in six provinces. International Journal of Cancer 2012;131:2650-2659.

~~25.~~26. Lee WJ, Cantor KP, Berzofsky JA, Zahm SH, Blair A. Non-Hodgkin's lymphoma among asthmatics exposed to pesticides. International Journal of Cancer 2004;111:298-302.

~~26.~~27. Vajdic CM, Fritschi L, Grulich AE, Kaldor JM, Benke G, Kricker K, Hughes AM, Turner JJ, Milliken S, Goumas C, Armstrong BK. Atopy, exposure to pesticides and risk of non-Hodgkin lymphoma. International Journal of Cancer 2007;120:2271-2274.

~~27.~~28. Blair A, Zahm SH. Patterns of pesticide use among farmers: implications for epidemiological research. Epidemiology 1993;4:55-62.

~~28.~~29. Canadian Cancer Society. Follicular lymphoma. Available at: http://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/types-of-nhl/follicular-lymphoma/?region=on [Accessed September 17, 2015].

~~29.~~30. Loomis D, Guyton K, Grosse Y, El Ghissasi F, Bouvard V, Benbrahim-Tallaa L, Guha N, Mattock H, Straif K on behalf of the International Agency for Research on Cancer Monograph Working Group, IARC, Lyon, France. Carcinogenicity of lindane, DDT, and 2,4-dichlorophenoxyacetic acid. Lancet Oncology 2015;16:891-892.

Date of last revision: September 21, 2015

~~30.~~31. Wang D, Gustafson P. On the impact of misclassification in an ordinal exposure variable. Epidemiologic Methods 2014;3:97-106.

~~31.~~32. Blair A, Zahm SH. Methodologic issues in exposure assessment for case-control studies of cancer and herbicides. American Journal of Industrial Medicine 1990;18:285-293.

~~32.~~33. Brown LM, Dosemeci M, Blair A, Burmeister L. Comparability of data obtained from farmers and surrogate respondents on use of agricultural pesticides. American Journal of Epidemiology 1991;134:348-355.

~~33.~~34. Blair A, Stewart PA, Kross B, Ogilvie L, Burmeister LF, Ward MH, Zahm SH. Comparison of two techniques to obtain information on pesticide use from Iowa farmers by interview. Journal of Agricultural Safety and Health 1997;3:229-236.

Human and Ecological Risk Assessment: Vol. 6, No. 1, pp. 47-71 (2000)

The Federal Government's Agricultural Health Study: A Critical Review with Suggested Improvements

George M. Gray,[1,*] Bernard D. Goldstein (Chair),[2] John Bailar,[3] Devra Lee Davis,[4] Elizabeth Delzell,[5] Frank Dost,[6] Raymond S. Greenberg,[7] Maureen Hatch,[8] Ernest Hodgson,[9] Michel A. Ibrahim,[10] James Lamb,[11] Terry Lavy,[12] Jack Mandel,[13] Richard Monson,[14] Mark Robson,[2] Roy Shore,[15] and John D. Graham[1]

[1]Center for Risk Analysis, Harvard School of Public Health; [2]Environmental and Occupational Health Sciences Institute, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School and Rutgers, The State University of New Jersey; [3]Department of Health Studies, University of Chicago; [4]World Resources Institute; [5]Department of Epidemiology, University of Alabama at Birmingham; [6]Professor Emeritus, Oregon State University; [7]Medical University of South Carolina; [8]Department of Community Medicine, Mount Sinai School of Medicine; [9]Department of Toxicology, North Carolina State University; [10]Department of Epidemiology, University of North Carolina; [11]Jellinek, Schwartz & Connolly, Inc.; [12]Department of Agronomy, University of Arkansas; [13]Division of Environmental and Occupational Health, University of Minnesota School of Public Health; [14]Department of Epidemiology, Harvard School of Public Health; [15]Department of Environmental Medicine, New York University

ABSTRACT

The Agricultural Health Study (AHS) has approximately 90,000 pesticide applicators and their spouses enrolled in a number of studies to determine whether exposures to specific pesticides are associated with various cancers and other adverse health outcomes. Although the AHS was intended to be an integrated program of studies, some significant difficulties have emerged. In this report, we examine the design of the AHS, identify important program strengths and flaws, suggest various improvements in the program, and recommend ancillary studies that could be undertaken to strengthen the AHS.

Overall, the AHS is collecting a large amount of information on potential determinants of health status among farmers and farm families. A promising feature of

* Corresponding author: Harvard Center for Risk Analysis, Harvard School of Public Health, 718 Huntington Ave., Boston, MA 02115; Tel: (617) 432–4341; Fax: (617) 432–0190; E-mail: ggray@hsph.harvard.edu.

1080-7039/00/$.50

the AHS is the prospective cohort study of cancers among farmers in which the research design determines exposures prior to the diagnosis of disease. More effort needs to be devoted to reducing selection bias and information bias. Success of the cohort study will depend in part on follow-up surveys of the cohort to determine how exposures and disease states change as the cohort ages. The cross-sectional and case-control studies planned in the AHS are less promising because they will be subject to some of the same criticisms, such as potentially biased and imprecise exposure assessment, that have characterized the existing literature in this field.

Important limitations of the AHS include low and variable rates of subject response to administered surveys, concerns about the validity of some self-reported non-cancer health outcomes, limited understanding of the reliability and validity of self-reporting of chemical use, an insufficient program of biological monitoring to validate the exposure surrogates employed in the AHS questionnaires, possible confounding by unmeasured, nonchemical risk factors for disease, and the absence of detailed plans for data analysis and interpretation that include explicit, *a priori* hypotheses. Although the AHS is already well underway, most of these limitations can be addressed by the investigators if adequate resources are made available. If these limitations are not addressed, the large amounts of data generated in the AHS will be difficult to interpret. If the exposure and health data can be validated, the scientific value of the AHS should be substantial and enduring.

A variety of research recommendations are made to strengthen the AHS. They include reliability and validity studies of farmer reporting of chemical use, biological monitoring studies of farmers and members of farm families, and validity studies of positive and negative self-reports of disease status. Both industry and government should consider expanded research programs to strengthen the AHS.

Key Words: epidemiology, pesticides, farmworkers, health effects

INTRODUCTION

The Agricultural Health Study (AHS) was launched in 1993 by scientists at the National Cancer Institute, the National Institute of Environmental Health Sciences, and the Environmental Protection Agency. The primary impetus for the study is a concern that exposures to chemicals on the farm, particularly certain fungicides, insecticides, and herbicides, may be responsible for a variety of adverse health effects, including cancer, neurological damage, reproductive problems, immunologic defects, nonmalignant respiratory disease, kidney disease, and impairments to the growth and development of children (Alavanja *et al.*, 1996).

As a result of this concern, just over 90,000 farmers, commercial applicators of farm chemicals, and their families in two states, Iowa and North Carolina, have been enrolled in a long-term health study. Most of the data in the study are being obtained from farmers through self-administered questionnaires and telephone interviews. Numerous questions were already asked of enrollees regarding their experiences as a farmer, their patterns of chemical use, their lifestyles, and their current health status. For some diseases, such as cancer, some of the future health information about enrollees will be obtained from state-wide registries.

The AHS is not a single study. Although the population of primary interest is the farmers ("private applicators"), there will also be studies of the health of commercial applicators and the spouses and children of private applicators. The AHS includes studies with at least four different designs and makes use of a variety of data sources.

First, the main prospective cohort study is expected to follow the 90,000 enrollees for many years or until death, to determine whether use of particular chemicals or other features of the farm environment and personal behavior are associated with poor health outcomes. This main study will not be completed until sufficient numbers of the cancers of interest have occurred or, ultimately, until most of the enrollees have died and the collected data on health outcomes have been fully analyzed. Interim reports on the cohort can be expected when the frequency of specific health problems supports a quantitative analysis of the factors associated with these health outcomes. An important design feature of the main cohort study is that much of the information on chemical use is obtained from farmers via survey methods *prior to the diagnosis of disease.* Although some enrollees had chronic diseases when they entered the study, the AHS investigators should consider analyzing the data with and without inclusion of these prevalent cases of disease.

Second, cross-sectional studies are being undertaken to determine the prevalence of certain noncancer health outcomes among farmers and farm families. The three initial cross-sectional studies are investigating (1) history of spontaneous abortion, menstrual function, and fertility in young women; (2) menopausal states, reproductive history, and selected chronic diseases in older women; and (3) neurologic symptoms and visual impairment in farmer-applicators. A cross-sectional design entails comparing the prevalence of reported adverse health outcomes with the reported use of or exposure to specific chemicals. Telephone interviews of subsamples of the cohort are being used to compare those people who responded to take-home questionnaires and those who did not as well as to obtain the information to augment the cross-sectional studies of non-cancer health outcomes (Sandler, 1998).

Third, nested case-control studies are planned for a variety of diseases including non-Hodgkin's lymphoma, leukemia, and cancers of the prostate, brain, ovary, breast, lung, colon, and stomach (Agricultural Health Study, 1993). Farmers in the cohort who develop a particular disease will be compared with controls selected from the cohort. Unlike the main cohort study, the nested case-control studies may entail obtaining some information from farmers or next of kin after a disease has been diagnosed. The investigators will examine whether cases report greater use of agricultural chemicals than selected controls. Cases and controls may also be invited to complete more detailed questionnaires aimed at obtaining a better understanding of possible differences in their exposure to a variety of farm and nonfarm factors.

Finally, some effort is being undertaken to determine how much farmers and their families have been exposed to selected chemicals. Biological monitoring, which typically entails the collection and analysis of urine and/or blood samples for multiple chemicals, is expensive. Biomonitoring was originally proposed to take place at 200 farms. Pilot studies found low participation rates (about 23%) and higher costs than anticipated and thus the program of exposure assessment has been scaled back. The current experimental design calls for samples to be gathered

from people at 125 farms, but unexpected obstacles have surfaced in obtaining funds for even this reduced program of biological monitoring.

The design and implementation of any research program as large and complex as the AHS requires many tradeoffs and compromises. Not every analyst would make the same choices, but on the whole we commend the AHS investigators for making a variety of sound choices in the face of limited resources and a complex challenge. As we emphasize below, we are particularly enthusiastic about the prospective cohort study of cancer outcomes because it responds directly to some of the methodological weaknesses of prior epidemiologic studies of farmers and pesticides. Other aspects of the AHS, such as the cross-sectional studies of disease prevalence, have serious problems. In this report we focus on what the strengths and limitations of the various AHS studies are, how the AHS can be improved, and what steps can be taken by the government and industry to enhance what is being done in the AHS through complementary efforts.

Information about the AHS used in this review was obtained primarily from publicly available documents and information presented at the AHS's annual public Advisory Panel meetings.[1] We recognize that more detailed plans may have been made but are not publicly distributed. Although the cohort has already been defined and enrolled in the study, numerous decisions have yet to be made about how the data will be analyzed and how future surveys of the cohort will be refined and improved. Thus, the emphasis in our report is on two issues: those that can be addressed by the principal investigators of the AHS through expansions or modifications of the workplan and those that need to be understood as inherent limitations when the findings of studies are published and disseminated.

The report is organized as follows. Section 1 addresses "Data Sources, Response Rates and Data Quality". Sections 2 and 3 address "Pesticide Exposure" and "Pesticide Use", respectively. Section 4 examines "Risk Factors Other Than Pesticides". In Sections 5 and 6 we examine the "Study Design Issues" and "Data Analysis Plans". Section 7 summarizes our recommendations on how the study can be improved and what additional studies can be undertaken to advance the field.

DATA SOURCES, RESPONSE RATES, AND DATA QUALITY

The AHS includes four types of data that could play important roles in epidemiologic analyses: health outcome data, pesticide use and exposure data, and data on potential confounders (risk factors) for disease. In this section, possible limitations in the scope or quality of each type of data are identified, and we present some suggestions aimed at enhancing data quality. Since most of the data used in the study are based on surveys of farmers and members of farm families, we begin with a discussion of the response rates obtained for the AHS questionnaires (Tarone *et al.*, 1997).

Response Rates to Questionnaires

The target population for the AHS is all persons required by the states of Iowa and North Carolina to obtain a pesticide applicator license. This includes "private"

applicators (farmers) and "commercial" applicators. Both states require periodic retraining to maintain a license for either type of pesticide applicator. The enrollment questionnaire was given to all attendees at training courses in the two states over a 3-year period. A 3-year cycle for licenses assured that all users had a chance to enroll. In January, 1997, enrollment through training classes was completed.

Not all applicators at training sessions agreed to participate. Some special recruitment efforts were undertaken to increase participation rates. In Iowa, the response rate for the enrollment questionnaire was 81.9% for private applicators and 42.2% for commercial applicators. In North Carolina, 84.8% of private applicators enrolled and the study design did not include commercial applicators. Overall, enrollment questionnaire data are available from about 53,000 private applicators and 5,000 commercial applicators (out of about 76,000 possible). Questionnaire data have also been collected from about 32,000 spouses of farmers (about 73% of those eligible).[2]

After pesticide applicators filled out the enrollment questionnaire at the training session, they were given three supplemental questionnaires (applicator; spouse; female and family health) to complete at home and return. The AHS uses the supplemental questionnaires to enroll spouses and other family members. The response rates for the supplemental questionnaires are low. Overall, about 44% of *enrolled* applicators completed and returned the additional questionnaire (33.5% of all eligible applicators). The Spouse Questionnaire, or a telephone administered version, was completed by 73% of eligible spouses. The Female/Family Health questionnaire was returned by about 39% of female applicators or spouses of enrolled farmers (64.6% of enrolled spouses).

The questionnaires are the primary source of data for the AHS. The enrollment questionnaire, which is used to define membership in the cohort, gathers personal identifiers on the applicator and his or her spouse. It also asks about work on and off of the farm, frequency of use of 22 pesticide compounds (*e.g.*, ever/never used and frequency of application) and ever/never used information on 28 more, one question about application methods and another about protective equipment, whether a doctor has ever diagnosed any of 16 diseases, and several questions on some lifestyle activities (including smoking) and the specific crops or livestock raised on the farm. These data are available for all applicators in the cohort except when there are missing responses.

The supplemental questionnaires are intended to gather more detailed information from the applicator and his or her spouse about pesticide use, family history of cancer, personal history of infectious and chronic diseases, over-the-counter medicine use, and diet. The Spouse Questionnaire, for the wife or husband of the applicator, asks for information about pesticide use and farm activities, along with information about factors such as laundering and vacuuming and information about the home that might influence pesticide exposure. Information about dietary and cooking practices is also collected. A self-reported medical history elicited from each subject includes about 55 diseases or disease symptoms. The Female and Family Health questionnaire is intended for female applicators or female spouses of pesticide applicators. This questionnaire collects information about the woman's reproductive cycle, pregnancies, and children. Identifiers, birthweight, nursing history, and whether the child ever worked on a farm are recorded for each child.

The low and variable response rates to the supplemental questionnaires seriously affect the quality of the AHS. Steps have been taken to increase response rates but the rate of non-response remains substantial. We encourage more efforts to increase the response rate, to reduce the potential for selection bias and increase statistical power. An evaluation of the potential for selection bias to influence risk estimates should be undertaken.

In the prospective cohort study, low response rates to questionnaires designed to obtain information on subject identifiers, exposures, and baseline disease status will clearly diminish statistical power and may create bias. The success of the cohort study also depends upon acceptable response rates to future follow-up surveys of the cohort. Periodic follow-up surveys are necessary to determine how exposures and disease states change as the cohort ages, thereby maintaining the prospective character of the study. If low response rates occur with the follow-up questionnaires, the potential for bias will increase, partly from misclassification of subjects (and person-years) with regard to chemical exposure and partly from residual confounding stemming from inaccurate measurement of risk factors other than pesticides. According to the AHS protocol (Agricultural Health Study, 1993), follow-up questionnaires will be administered every 5 years. Since no follow-up has yet been administered, response rates are unknown.

Selection bias should be reduced in the prospective cohort study if persons who already have the disease(s) of interest are identified and excluded from the cohort at the beginning. Identification of diseases diagnosed at the time of enrollment into the cohort may be done well for conditions, such as some cancers, that have an easily defined point of diagnosis but is more difficult for certain neurological conditions and for renal, respiratory, and cardiovascular diseases. For instance, bias will occur if persons who are at risk of cancer and are exposed are more likely to participate by returning questionnaires. There are plans for cohort studies of kidney, neurologic, respiratory and cardiovascular disease that might be biased by the erroneous inclusion of subjects with disease onset before enrollment, if the probability of study participation depends on exposure status. Furthermore, if response rates are low for questionnaires designed to obtain information on medical conditions occurring during the follow-up period, the likelihood of bias is high.

In cross-sectional and case-control studies, low response rates have most of the same potentially detrimental effects on precision and accuracy as mentioned above. In addition, poor response raises the likelihood that selection bias will occur because it is likely that participation will depend both on exposure status and on "disease" status in a manner that could bias estimates of prevalence ratios or odds ratios (*e.g.*, through underrepresentation of exposed persons without disease).

Health Outcomes

Accurate ascertainment of the presence or absence of disease among farmers and members of farm families is critical to the success of the AHS. Some of the diseases of interest in the study are relatively rare and only a small number of cases of these diseases can be expected. Thus, it is appropriate to consider the quality of the health-outcome data being collected in the AHS, looking at the potential for both false-positive and false-negative errors.

Cancer

The statewide cancer registries in Iowa and North Carolina will be used by the AHS investigators to determine which subjects develop various types of cancer. The Iowa registry is well established and is believed to provide accurate and reasonably complete data on incidence of cancer in the state. The North Carolina registry is newer but should provide data of sufficient accuracy.

Use of the cancer registries will be hampered if personal identifiers such as name, birth date, Social Security number, and gender are not available from both the cancer registries and the AHS cohort. Such identifiers are critical to linking subjects in the AHS cohort to registry records. In February 1997 it was reported that in Iowa the four identifiers mentioned above were available for 94.0% of commercial applicators, 86.3% of private applicators, and 53.4% of enrolled spouses of married private applicators. In North Carolina the four identifiers were available for 86.6% of private applicators and 76.5% of spouses. Linkage with registries may be acceptable with current identifiers but AHS investigators are making efforts to increase the completeness and quality of data needed for record linkage. There are other ways to determine whether enrollees have developed cancer, but they are generally more expensive.

Non-Cancer Health Outcomes

Mortality from kidney, neurologic, respiratory, cardiovascular, and other diseases can also be assessed through objective measures that do not entail self-reporting by subjects in the cohort. For example, mortality from specific causes can be monitored through periodic follow-up through the National Death Index and state and local vital statistics records. Yet even for data from objective sources, potential validity problems need to be identified and addressed.

In order to accelerate the opportunity to cover a wide range of non-cancer outcomes, the AHS relies on self-reporting of health states by farmers and members of farm families on both the enrollment and supplemental questionnaires. The self-reporting occurs either through return of written questionnaires or responses to telephone interviews. Telephone surveys of special subgroups of the cohort are being employed to reduce the potential for selection bias in the cross-sectional studies, but it is possible for a modest amount of selection bias to have a substantial effect on results. Diseases of particular interest to the AHS investigators include kidney disease, neurotoxicity and neurological disease, reproductive and developmental impairments, and immunologic effects. Several questions ask about possible acute toxicity episodes associated with pesticide use.

Section IV of the main enrollment questionnaire has two questions regarding health. Question #28 inquires whether "a doctor has ever told you that you had any of the following conditions": A list of 16 conditions is supplied (asthma, tuberculosis, other chronic lung disease, pneumonia, melanoma of skin, other skin cancer, leukemia, Hodgkin's disease, non-Hodgkin's lymphoma, other cancer, heart disease, diabetes, Parkinson's disease, kidney disease, nervous disorder, and depression), each to be answered yes or no. For the cancer outcomes, it will ultimately be

feasible to compare the self-reports of subjects to the data obtained through the statewide cancer registries. A strategy for addressing discordant data is needed.

Question #29 inquires whether "your parents, brothers, sisters, or children related to you by blood ever had any of the following?" A list of 14 conditions is supplied, again with yes or no responses. If the subject has multiple blood relatives, the implicit understanding is that the question refers to any of them.

In addition to these questions about diseases, questions #14 and #15 provide additional information about acute health effects that may be related to pesticides. These questions do not ask about a medical diagnosis, and no effort is being made to validate the answers.

Question #14 asks "How often, if ever, have you had the following symptoms that you think may be related to your using pesticides?" There are seven listed symptoms: "been excessively tired", "had headaches/dizziness", "had nausea or vomiting", "had skin irritation", "had eye irritation", "had chest discomfort", and "felt nervous or depressed". For each symptom, the respondent is asked to respond on a scale of never/rarely, sometimes, frequently, almost always. This set of questions seems to combine elements of symptom frequency and causal attribution. It is not clear how the respondent is expected to judge whether such symptoms were "related to your using pesticides" unless the effects were immediate and unambiguous. It may be preferable to ask separate questions about the frequency of these symptoms and the respondent's view about whether they are associated with pesticide use, although questions about validity might remain. It is also not clear what would be meant by "frequently/almost always", since no frequency context is suggested to the respondent. The response may represent symptom frequency in absolute terms or as a percentage of the total number of pesticide applications. Given the ambiguous nature of this question, the meaning of the information that is elicited will be uncertain.

Question #15 asks subjects: "As a result of USING PESTICIDES (emphasis in original), how often have you: a. seen a doctor, b. been hospitalized." The possible responses are never, once, twice, or three or more times. Again, this question presumes that the respondent knows something about the causative role of pesticides in particular situations, perhaps because he or she experiences unusual symptoms in short order after the chemical is applied. Some visits may be after exposure but before symptoms appear. In ambiguous situations involving common symptoms and longer time lags, the respondent may not realize that the chemical exposure was responsible for the symptom or may attribute to the chemical a symptom that was not caused by the exposure. If the question is intended to provide a surrogate measure of exposure to chemicals, it needs to be used with caution if it is used at all.

Although there is limited information on noncancer health outcomes in the main enrollment questionnaire, the supplemental questionnaire includes a fairly detailed self-reported "medical history" from each subject. The low overall rate of response to the supplemental questionnaire, despite efforts to increase response, will prevent full understanding of the cohort's exposure and health states.

Questions #87 through #102 in the supplemental questionnaire ask about numerous aspects of the applicator's health status. For example, Question #87 asks about

each of 41 listed diagnoses (*i.e.*, whether a doctor has ever told the subject that he or she has that condition). Question #90 asks how frequently, during the last 12 months, the subject has experienced each in a list of 23 symptoms ranging from dizziness and headaches to feeling tense or depressed. Questions #96 through #102 focus on the respondent's vision and use of eyeglasses. Responses to these questions need to be validated.

It may be that any biases will cancel out because potential cases and non-cases interpret questions in roughly the same manner, as may be expected in a prospective cohort study, but it will be very difficult to know for sure the overall or net impact of any resulting biases.

Applicators who are women, and the spouses of male applicators, are also asked to complete a "Female and Family Health Questionnaire" that includes numerous questions on the subject's reproductive and pregnancy history, and about the health status of children. The AHS is also using a specialized "Women's Health Questionnaire" and a separate "Young Women's Health Questionnaire" to obtain specific pesticide use information and more detailed health information on subgroups of women who have enrolled in AHS. The former questionnaire has a special section on menopause while the latter questionnaire emphasizes menstrual functioning and pregnancy history.

Epidemiologists do not expect perfect concordance between self-reports and medical records. Although subjects may supply inaccurate data, medical records are themselves not free from error. The accuracy of self-reports presumably vary by type of health endpoint, questionnaire design, period of recall, and population studied. For many reproductive endpoints, the results of reliability and validity studies are reassuring, while for others there is concern (Bean *et al.*, 1979; Wilcox and Horney, 1984; Olson *et al.*, 1997). For some endpoints, such as menstrual function, there is no practicable gold-standard to compare with self-reports. It is important for the investigators to address how they will incorporate uncertainty about self-reports into their analyses and interpretation of results.

The AHS is collecting a large amount of self-reported health information on non-cancer health outcomes. Most of the specific questions on non-cancer health outcomes used in the questionnaires have not been assessed for validity or reliability and there appear to be no plans to initiate such studies by the AHS team. Apparently, follow-up questionnaires will not repeat questions about past health outcomes, preventing assessment of reliability. Some of these questions have already been used in previous studies and may have been subjected to some reliability and validity checks but study context can influence responses. More such studies would help users understand the quality of the non-cancer outcome information that will be analyzed in the AHS. It is crucial that reports of both the presence and absence of specific outcomes be validated in order to ascertain false-positive and false-negative errors.

Bias can occur when subjects know the purpose of a study and when they also know their exposure status and disease status. For example, "exposed" subjects (*e.g.*, heavy users of chemicals) with disease may be more willing to participate in the AHS cross-sectional studies than nonexposed subjects who also have the same disease. The prospective cohort design provides an important protection against such bias,

as long as a subject's exposure truly precedes the onset or diagnosis of the disease being investigated. The prospective cohort study also provides a good opportunity to obtain valid information on exposure and disease status. This strength may be enhanced through various analyses designed to detect and diminish information bias and other problems with the AHS data. It would probably be necessary to gather additional data to support such methodological substudies. It is preferable to minimize the potential for bias by obtaining valid information from all subjects.

PESTICIDE EXPOSURE

Although the primary goal of the AHS is to assess the relationship between human exposure to pesticides and a variety of adverse health outcomes, direct measurement of human exposure to pesticides will be limited by cost considerations. Most of the analyses will be based on surrogates for exposure.

As of early 1998, the U.S. USEPA team planned to select a sample of 125 farms and evaluate total exposure for several chemicals through measurement of environmental media, personal exposure (*e.g.*, through patches on clothing), and samples of urine and blood, taken soon after application. These samples will then be analyzed for a limited number of chemicals of greatest interest. Unexpected funding problems may prevent implementation of USEPA's plan.

Since no direct measures of pesticide exposure will be available on most of the 90,000 members of the AHS cohort, the investigators will rely primarily or exclusively on surrogates for pesticide exposure derived from the questionnaires administered to farmers and members of farm families. For example, previous studies have considered as surrogate factors such measures as frequency of application per year, number of years of application, and application practices that may be related to exposure (*e.g.*, method of application and type of protective equipment used) (Hoar *et al.*, 1986; Zahm *et al.*, 1990). It is not known how well any of these surrogates indicate biologically significant exposures or whether any is appropriate. A case can be made that exposure surrogates should be validated before initiating a major epidemiologic study, or at least before exposure-response analyses are undertaken.

A key goal of the USEPA portion of the AHS is calibration of reported work practices with actual farmer exposures, using the information obtained from the measurements gathered on the sample of farms. Ideally, this information would allow at least a ranking of exposure potential by method of application and protective equipment used. For example, some pesticides are formulated as liquids, and gloves may provide a great deal of protection. Others are formulated as dusts or sprays and thus gloves may make little difference, while a respirator or mask may greatly reduce exposure. Still others are large granules and neither type of protective equipment may have much influence on exposure.

Because of its limited size, the USEPA study is unlikely to provide a rigorous validation of the numerous exposure surrogates derived from the AHS questionnaire data. A larger sample of farms, pesticides, and work practices would be useful in validating the surrogates against the background of other significant determinants of exposure such as the subject's age and role in pesticide use. There are also

questions about the representativeness of sampled farms. The USEPA has had difficulty obtaining the participation of farmers. In a pilot study in North Carolina, fewer than 10% of farmers asked agreed to participate (U.S. Environmental Protection Agency, 1997). It seems unlikely that the farmers who agree to participate will be representative of all of the farmers in the AHS. The timing of the USEPA exposure study is also a source of concern. USEPA's exposure study is just getting underway but the AHS enrollment questionnaires have already been administered to the 90,000 enrollees. If the USEPA study raises questions about the validity of the exposure surrogates contained in the enrollment questionnaire, the data that have already been collected from farmers on work practices will be of diminished utility. If done in a timely fashion, it may be feasible to revise future follow-up questionnaires in a way that will benefit from the insights generated from the USEPA's exposure study.

Previous studies have relied on the assumption that total lifetime exposure to one or more pesticides is determined by annual frequency of application and number of years of application. Although this assumption may seem logical, there is no plan to validate it. It is possible that those farmers who apply pesticides frequently and have done so for many years do so with particular experience and care, which might suggest that their absorbed dose per application is less than the exposure of farmers who apply chemicals less frequently or have fewer years of experience in farming. Of course, bias may also run in the opposite direction if some applicators become careless or even contemptuous of risks as the substances and application practices become familiar. A particular task, such as mixing, may lead to much greater exposure than frequent application. If rare but serious mishaps or spills have a powerful influence on total lifetime exposure, number of applications may be a poor surrogate for total exposure, since the probability of mishap/spill may be smaller among high-frequency applicators. The USEPA study may not be large enough to detect these rare yet serious incidents. Thus, it is not obvious that total exposure to pesticides in a farmer's lifetime, on average, will be a straightforward multiple of the number of applications in a farmer's lifetime.

The use of inappropriate or imperfect exposure surrogates may compromise the validity of the study by producing erroneous measures of association. Errors due to misclassification can produce bias toward the null (attenuation of the magnitude of a true positive or inverse association) or bias away from the null (exaggeration of the strength of a true weak or true null association). In large prospective follow-up studies of relatively common exposures and diseases, exposure misclassification tends to be nondifferential with regard to disease status. Nondifferential exposure misclassification will produce bias toward the null if exposure is classified dichotomously (*e.g.*, exposed vs. unexposed, high vs. low exposure). If more than two categories of exposure are evaluated, however, nondifferential misclassification has an unpredictable impact and can produce bias away from the null (Correa-Villasenor, A., Stewart, W. F., Franco-Marina, F., and Seacat H. (1995); Thomas, 1995). In small studies or studies in which exposure is rare or disease rates are low, the impact of misclassification, again, is unpredictable. There is no guarantee that exposure misclassification will be nondifferential even if objective exposure assessment pro-

cedures are used. Misclassification will reduce the power of the study to detect any genuine cause-effect relationships and will also reduce the validity of findings. Reductions in power are a serious issue because they will undermine the ability of government and industry to regulate harmful exposures and to reassure farmers with "negative" results.

Biomonitoring studies of farmers who mix and apply pesticides with different frequencies might help resolve this matter, but such studies would need to be large in size and would be logistically complex. Such studies may induce behavioral changes (*e.g.*, extra safety precautions) on the part of some farmers that are not typical of their normal behavior.

Although it will be difficult to validate whether number of applications is a strong predictor of total exposure, it may be more feasible to study the impact of work practices and method of application on the amount of actual pesticide exposure. A farmer's personal habits can have an enormous influence on pesticide dose, as measured by urinary excretion, even when the same protective equipment is used (Lavy, Walstad, Flynn, and Mattice, 1982; Forbess *et al.*, 1982; Leng, Ramsey, Braun, and Lavy, 1982). It will be difficult to characterize this source of variability in the small sample of farmers being evaluated by the USEPA. Broader studies of the type planned by the USEPA, with a focus on the AHS pesticides and work practice and protective equipment questions, would be very useful. Some information on the role of work practices and protective equipment is already available in USEPA's Pesticide Use Handlers Database and our understanding is that the AHS investigators have begun to exploit this source of data. We encourage more efforts in this direction. The Department of Defense has conducted large programs of research on the efficiency, safety, and comfort of protective gear, and some of the results (*e.g.*, points of leakage or tolerance by the protected person) may be directly applicable to pesticide applicators.

There are also practical and technical concerns associated with any urine biomonitoring program. The USEPA investigators are aware of many potential pitfalls but still may have difficulty dealing with them. One of the biggest problems is time. If a pesticide is rapidly excreted, measurements must be made quickly after a single application to be useful for exposure assessment. If, however, the material is cleared slowly from the body, the amount of the chemical measured in urine will be highly dependent on the frequency of applications and the time interval between applications. There are significant differences in pharmacokinetics across compounds that will influence the relationship between frequency/pattern of use and exposure. Thus, a serious biomonitoring program must have a protocol that tailors the measurement regime to the behavior of the compounds under study. Yet the USEPA plans to sample only a fraction (perhaps as few as 10) of the 50+ chemicals being assessed in the AHS, and funding obstacles are jeopardizing even this modest level of effort.

Another key assumption of the AHS is that exposure of farm family members to pesticides is associated with the farmer's patterns and frequencies of use. Little is known about the nature of this relationship or how it varies for different compounds and farm types (Lowenherz *et al.*, 1997). The existing studies are small in size and are quite limited in the number and type of pesticides evaluated. Assuming partici-

pation obstacles can be overcome, biomonitoring could be used productively to better understand the presence and magnitude of indirect exposures to farm families that are assumed in the Spouse Questionnaire and the Female and Family Health Questionnaire. USEPA has limited plans in this area that will need to be expanded considerably if they are to be useful in the AHS.

The NCI also plans a biomarker component, collecting buccal DNA samples from a subsample of the AHS cohort, to store for later analysis of genetic polymorphisms potentially related to susceptibility to pesticide-induced disease. Although this effort is of considerable scientific interest, it is not likely to assist in validation of the exposure surrogates to be used in epidemiologic analyses.

In general, a major limitation of the current design of the AHS study is that so few direct measurements of human exposure to chemicals will be available. The information that USEPA plans to collect may be useful in its own right but, for the reasons stated above, is not likely to be as useful as it could be for use in the epidemiologic analyses to be performed in the AHS. Pesticide exposure studies that are linked to epidemiologic investigations are urgently needed if a major advance is to be made in our understanding of the relationship between pesticides and human disease. The significant cost associated with such an effort is noted, but the scientific value of this major epidemiologic study is questionable without a valid exposure assessment.

PESTICIDE USE

In the AHS, the questionnaires filled out by subjects elicit information on various aspects of pesticide use rather than on exposure directly. This approach is sensible because the respondent is in a better position to report accurate information on whether and how a chemical is used than information on the amount of exposure to chemicals. However, there are still serious questions about the quality of the pesticide use data that are being collected in the AHS. Since these data are likely to be critical to the interpretation of the epidemiologic analyses, the associated quality concerns need to be considered carefully.

In the AHS enrollment questionnaire, the primary questions (Qo, #11A-D) ask: "During your lifetime have you ever personally mixed or applied this pesticide?; how many years did you mix or apply this pesticide?; in an average year when you personally used this pesticide, how many days did you use it?; and when did you first personally use this pesticide?" (Paraphrased). These questions are posed for 22 named pesticides. For an additional 28 compounds, there is a simple question about whether that pesticide had ever been used.

In order to answer these questions, respondents must remember with some accuracy when they first used products and their frequency of use of each pesticide product, and they must be able to compute averages in their head involving multiple years of use. For older subjects who have many years of farm experience, accurate responses will be difficult to supply. Moreover, some pesticides are sold and applied as mixtures and thus the exact ingredients may not be known to farmers. It can reasonably be expected that there will be inaccuracies in these data.

In the AHS enrollment questionnaire, there are two important questions about work practices. Question 16 asks: "how do you personally apply pesticides?" The offered answers include 20 options that are not differentiated by livestock or crop farming, by specific crop, or by pesticide used. Question 17 asks "what type of protective equipment do you generally wear when you personally handle pesticides?" The offered answers include 8 options, again making no distinction between farm types or pesticides used. Since most farmers will have had different practices for different crops or pesticide products, it is not clear how they will answer these questions in a meaningful way since multiple answers do not appear to be allowed.

There are, of course, real concerns about the ability of farmers to recall use of specific pesticides, let alone their frequency of use, when confronted with a long list of compounds. Many farmers know pesticides by trade names, not technical names. The AHS questionnaires list some trade names for all chemicals but the list is not exhaustive. In addition, farmers now often use formulations that contain several pesticides. A respondent who knows only one of the compounds or trade names could underreport the use of other pesticides in the mixture. Errors of recall may occur differentially between controls and diseased persons.

Due to a change in enrollment procedures, the AHS investigators do have duplicate enrollment questionnaires from 1223 applicators from Iowa (Alavanja, 1998). Reliability was reported as both the percent agreement (the fraction of applicators giving the identical answer to a question on both questionnaires) and kappa statistic, often used as a measure of reliability. For example, smoking had an agreement of about 90% and a kappa of 0.88. Reports of ever/never use of specific pesticides had agreement around 80% with kappas around 0.60. The agreement of frequency of use questions was not reported. Some questions, especially those about vegetable and fruit consumption, had quite low agreements (30 to 40%) and kappas (about 0.50). Of course, this analysis does not address the validity of the responses. It may be useful to include some more important use questions on future follow-up surveys to gauge reliability in the whole cohort.

A weakness of the AHS is that adequate information is not being collected on excipients such as solvents, stabilizers, diluting agents, preservatives and other chemical substances that are used with pesticide products. Confusion may occur about whether reported health effects are attributable to active ingredients or excipients. For regulators and firms interested in the design of pesticide products, it is crucial to know what precisely is causing a reported health effect.

There is no reason to believe that large numbers of subjects were deliberately dishonest in the enrollment questionnaire about their patterns of pesticide use. However, the questions about use of protective equipment may have induced some "socially desirable" but inaccurate answers, especially when questionnaires were administered at training sessions. It is also quite possible that pesticide products near the bottom of the lists of 22 and 28 were checked less frequently by respondents who became weary filling out this rather arduous aspect of the questionnaire. This problem could be smoothed out in the future follow-up surveys by rotating the order of the products.

A study of the magnitude of the AHS requires good understanding of the validity and reliability of each major data set. The AHS will obtain pesticide use data from

responses to a written questionnaire of farmers. Data will be collected both at the beginning of the study and with follow-up questionnaires of unspecified frequency, for either the whole cohort or a select subsample, in later years. On the subject of validity, purchase records have been used in the past to ascertain whether written answers to a "yes/no" question on use of specific products are accurate. One study reported a 60% agreement rate between purchase records and reported use of specific products (yes/no) (Blair and Zahm, 1993). Agreement between farmer's recall of years of use and the records of their suppliers ranged from 38% to 68% depending on type of pesticide and crop. Measures of frequency of use in a year have never been subjected to a validation study.

When social scientists find it difficult to validate questionnaire data, it is typical to at least conduct reliability studies, such as repeated administrations of the same (or similar) questionnaire(s) to respondents, to determine whether answers to the same question are stable. Few reliability studies of self-reported pesticide use, particularly the quantitative responses, have been published in the literature (Johnson *et al.*, 1993). In addition, since reliability is influenced by the particular wording of questions and response choices, there probably would be limited generalizability from reliability studies of other questionnaires.

The questions of reliability and validity regarding the reported data could be addressed in several ways. In addition to the small study already mentioned, a comparison of the responses of farmers to selected questions that have been included on both the enrollment and supplemental questionnaires will provide some ideas about reliability. Studies comparing self-reported use to purchase records for a subsample of the AHS farmers could provide an idea of the validity of self-reported use estimates. Even if recent purchase data can be obtained, it is likely that purchase records for earlier years will be less complete. Thus, it will be more difficult to verify the accuracy of self-reports of pesticide use in the past. Another opportunity to check self-reports might come from the Extension Service recommendations for each crop in Iowa and North Carolina. Consistency between self-reports and the recommendations of the Extension Service is one possible measure of accuracy. However, if such recommendations are widely known, farmers may be reluctant to report actual use patterns that deviate significantly from these recommendations.

The chemicals, formulations, and application methods used on farms have changed significantly over time. Herbicides once applied at rates of pounds of active ingredient per acre are now applied in ounces per acre. Formulations have been developed to reduce exposure by making the pesticide in large granules or as packets that are dropped into an application tank, with no need for mixing or loading. These changes in patterns of pesticide use mean that data gathered about farming practice today are not a valid reflection of what was done in the past. The amount of exposure per application is probably smaller today than it was years ago, further complicating any calculation of cumulative exposure.

These details are important because if pesticides cause chronic diseases such as cancer and neurological disease, the biologically meaningful measure of exposure may be a cumulative dose figure that accounts for farming practices years or even decades ago. For chronic diseases diagnosed over the next 5 years or so, the exposure of interest probably occurred many years ago. Yet information about

changes in farming practices over time is not being gathered in the AHS. In addition, the extent of pesticide use information to be collected in follow-up surveys of the cohort is not clear. If most of the pesticide use assessment in the AHS proves to be retrospective, the AHS will have little advantage over previous studies.

RISK FACTORS OTHER THAN PESTICIDES

Numerous factors other than pesticide use are known or suspected to contribute to the development of various diseases and health impairments under study in AHS. These factors are important because they may confound (exaggerate or attenuate) the effects of pesticides, they may interact with the effects of pesticides, or they may prove to be of much greater quantitative importance than pesticides even if they are not confounders or interacting variables.

Confounding Variables

In epidemiologic analysis, a confounding variable is a risk factor for the disease of interest that is associated with the exposure of interest (in this case, pesticides). For example, in an analysis to determine whether frequent application of a particular pesticide is a risk factor for a particular type of skin cancer, exposure to sunlight is a potentially confounding (or interacting) variable. The ultraviolet radiation from exposure to sunlight is known to be a cause of skin cancer and farmers who engage in frequent application of pesticides may have more exposure to the sun than other farmers. If exposure to sunlight is a confounding variable and is omitted from the epidemiologic analysis, the estimated risks associated with pesticide exposure will be biased. This bias can be reduced or eliminated by collecting information on the confounder and including such information in a multivariate analysis of the disease in question.

Concern about possible confounding may arise if certain patterns of pesticide misuse (*e.g.*, failure to use protective equipment) are used as a surrogate for pesticide exposure without consideration of the farmer's lifestyle. Farmers who do not use protective equipment (or engage in risky application practices) may be more likely to engage in a wide range of risky behaviors at work and at home than farmers who use protective equipment (or engage in low-risk application practices). Some of those risky personal actions may be linked to the health outcomes under study.

The AHS collects data on numerous variables that might confound the relationship between pesticide use/exposure and disease outcomes. Yet we know of no effort to identify such confounding variables and include them in the AHS study plans. Information about some risk factors other than pesticides is being collected in the AHS study (*e.g.*, aspects of the diets of farmers) but it is not clear whether such variables are correlated with pesticide exposure and are likely to cause the same types of tumors that chemicals may cause. In addition, since these data are collected in the supplemental questionnaires, they are not available for the entire cohort.

Interacting Variables

The effects of pesticide exposure on human health may be magnified or attenuated by other behavioral and/or environmental factors. For example, it has been

shown that the risk of lung cancer due to radon exposure among uranium miners is much larger among smokers than nonsmokers (Hornung, Deddens, and Roscoe, 1995). We do not know of any interaction effects to be expected in the AHS data, but if others know or suspect of such interactions, they should be postulated explicitly prior to data analysis and then tested rigorously in the statistical analyses.

Other Important Risk Factors

Although pesticide exposures are certainly worthy of study, these exposures are not necessarily the most biologically plausible determinants of disease in farmers or farm families and they may not prove to be as quantitatively important as a variety of risk factors such as smoking, diet, and obesity. Even accepting that chemicals are a major priority for study, more effort might be devoted to understanding farmer exposures to such agents as veterinary pharmaceuticals, engine oils, consumer products, animal viruses, and the crops themselves.

If modified appropriately, the AHS could be used to generate comparative information that might help farm families develop a sense of perspective about the relative risks associated with different risk factors in farm life. In order to serve this function, future surveys of the cohort planned by the AHS investigators need to devote more attention to risk factors other than pesticides and compare their relative significance to those of pesticides based on rigorous epidemiologic analysis. Nevertheless, a significant focus on pesticides is worthwhile.

STUDY DESIGN ISSUES

From a methodological perspective, the AHS employs several different study designs in various phases of the epidemiologic inquiry. They include a prospective cohort design, a nested case-control design, and a cross-sectional design. These different study designs have inherent strengths and weaknesses that need to be understood when the findings of the study are interpreted and compared to the findings of other investigators.

Prospective Cohort Study

A typical prospective cohort study follows subjects from the time of enrollment in a study until a particular disease is diagnosed or some other event occurs and/or death. The subjects' frequency and/or degree of exposure to the chemical or physical agents of interest are typically documented at the time of enrollment and throughout the follow-up period. An advantage of this study design is that exposure determinations are made by the investigators before anyone (including the investigators and the subjects) knows which subjects will develop a particular disease or die prematurely. A disadvantage of the prospective design is that accurate measurement of exposure to pesticides and other disease determinants requires that the cohort be questioned or monitored at intervals during the study period, not just at the beginning. For cancers diagnosed during the first 5 years of study, the exposure assessment in the cohort study is based on recollections of pesticide use patterns from years or even decades ago.

Determining exposure status prior to knowledge of health outcome is particularly critical in the epidemiology of pesticides. Previous findings in the literature, which were based primarily on the case-control design, have been criticized on the grounds that those farmers who developed disease (or their next of kin) may have been motivated (for a variety of reasons) toward more complete and accurate reporting of pesticide use and/or exposure than those farmers who did not develop the diseases of interest (Ibrahim et al., 1991). If such differential misclassification of exposure occurs, it will tend to create a spurious positive association between exposure and disease. The prospective cohort design selected by the AHS investigators reduces, but does not eliminate, the chances that bias from differential exposure misclassification will occur because use and exposure are determined prior to knowledge of health outcome. It is critical that follow-up surveys of the cohort be administered on a regular basis to document how exposure and disease states change as subjects age.

The major disadvantage of the prospective cohort design is that, for some chronic illnesses, it takes a long time for sufficient numbers of subjects to fall ill or for the data to be useful for analysis. It is also an inefficient approach to studying relatively rare tumors such as soft-tissue sarcoma and leukemia. Overall, though, we are very enthusiastic about the decision of the AHS team to invest in the prospective cohort design and encourage the investigators to make every feasible effort to achieve acceptable response rates in the follow-up surveys of the cohort and address potential biases in the study.

Nested Case-Control Study

A typical case-control study will enroll "cases" who are known to have the disease in question and compare them to a random subset of "controls" who do not have the disease in question. If cases and controls are both selected from subjects enrolled in a particular cohort study, the study is referred to as a "nested" case-control study. The strength of this design is that the cases are included in the cohort studied. If exposures to a particular agent cause the disease in question, then the life histories of the cases should exhibit different (and presumably greater) exposures than the life histories of controls. Exposures to the agents of interest are typically assessed retrospectively for cases and controls (*i.e.*, after the death has occurred or the disease determination has been made), sometimes via interviews with next of kin or through reconstruction of job histories and practices. Like the prospective study, the nested case-control aspect of the AHS would be constrained by the time to development of disease and the numbers of persons in the cohort. We do not discuss this design in detail here, because it is currently being given low priority in the AHS and its strengths and weaknesses have been addressed elsewhere (*e.g.*, Monson 1990).

The Cross-Sectional Design

A typical cross-sectional study collects information on exposure and disease simultaneously from a sample of subjects. The association between reported exposure and disease is then investigated within the sample. If exposure causes disease,

it is anticipated that subjects that report more exposure will be more likely to report the health outcome of interest. A major weakness in the cross-sectional approach is the potential for bias. When exposure and disease data are gathered at the same time, it may also be unknown whether the temporality is correct, that is, that the exposure to the presumed cause actually preceded the disease, especially with diseases with no easily identified time of onset. There may be bias if persons change exposure status because of disease (*e.g.*, people with disease may stop working with pesticides). Although this research design has some utility for generating hypotheses, it is not considered useful in defining most cause-effect relationships. There are also concerns about the quality of the data gathered for the AHS cross-sectional studies. Without medical verification of self-reported disease, any associations found in the cross-sectional studies will be a weak basis for planning future studies.

DATA ANALYSIS PLANS

While the AHS study team presumably has some well-defined primary hypotheses, they are not specified in the *Environmental Health Perspectives* article (Alavanja et al., 1996) or in the more detailed study plans that have been made available to the public. By well-defined primary hypotheses, we refer to *a priori* hypotheses regarding specific chemicals, specific tumor types or health outcomes, and specific surrogate measures of exposure. Specific hypotheses and detailed plans for analysis help focus the gathering of data on both exposure and disease outcomes. They may also help investigators to avoid overinterpretation of the random oddities that occur in any large and complex data set.

Given the many possible comparisons of pesticides, methods of use, work practices, and health outcomes, a formal statement of why a particular pesticide/outcome combination should be analyzed seems desirable. Without any precommitment to specific hypotheses, the proper interpretation of any associations that are found will be less clear. Although it is appropriate for the AHS team to explore many possibilities when the data are analyzed, it should be clear to readers and decision makers which results confirm prior evidence or concerns and which are found only in the AHS data.

The large amount of questionnaire data developed by the AHS provides rich scientific opportunities but also particular challenges for analysis and interpretation. For example, information is gathered from respondents on numerous health outcomes (approximately 25 outcomes in the private applicator enrollment questionnaire, 70 outcomes in the farmer applicator and spouse questionnaires, and 35 outcomes in the female and family health questionnaire — a total of 130). For cancer, there will be numerous tumor types available for analysis from registry data. In addition to numerous health outcomes, information is gathered on numerous pesticide products (approximately 50 in the enrollment questionnaire and another 100 in the farmer applicator questionnaire). For exposure (dose)-response analysis, it appears that more than 35 different surrogates of exposure can be constructed from the responses to the questions about pesticide use, application methods and

work practices (*e.g.*, average days of use per year, number of products used, years of use, different types of protective equipment and methods of application).

One can confidently predict that some of the multitude of exposure-response combinations will be statistically significant in the absence of any real effect. Without clearly stated *a priori* hypotheses, the investigators will have to exercise considerable discretion in data analysis and may exercise insufficient or excessive caution in their interpretation. The exercise of this discretion can be evaluated by the scientific community only if a small number of completely specified primary hypotheses are developed prior to any inspection of results. "Completely specified" means that the method of analysis must be given in detail for each primary hypothesis. The benefit of this approach is the increase in plausibility of any "positive" findings among the primary hypotheses; the cost is that all other hypotheses lose some support, though some may still be compelling and others may be examined in subsequent studies.

Important questions arise about the role of conventional measures of statistical significance in the reporting and interpretation of results. Should numerical adjustments be made to published p-values to account for multiple comparisons? Given that many possible associations may be explored prior to publication of final results, what degree of documentation should be provided by the investigators of exploratory analyses? If the documentation requirements are minimal, how will the scientific community understand the importance of the associations that are reported? The importance that may be placed on findings of no association between a specific pesticide and health outcomes raises the question of the reporting requirements for analyses that failed to find an association. Parallel consideration must be given to reporting requirements for "inverse" associations (*e.g.*, relative risks less than 1.0 for a particular exposure). At the same time, it would be helpful if the AHS investigators would publish all data and analytical results in some accessible format. Key findings would especially benefit from documentation of their consistency within the AHS database. Widely accessible electronic media such as the World Wide Web makes this feasible.

A detailed analysis plan and careful interpretation can reduce or eliminate these concerns. Examination of internal consistency can provide information about the plausibility of a particular association. A reasonably consistent dose-response gradient is an important criterion. One implication of this criterion is that statistically significant dose-response trends caused primarily by one dose group, especially if it is an intermediate dose group, should be interpreted cautiously. On the exposure side, a finding that the strength of an association increases with particular use practices that are expected to yield higher exposures (and decreases with increasing farmer care), could be valuable evidence in buttressing study results. Sensitivity analyses involving different exposure surrogates and exposure groupings can also demonstrate whether findings are robust.

A key form of evidence to inform hypotheses and corroborate (or refute) analytic findings is biological plausibility. Pesticides, in addition to prescription drugs, are among the most thoroughly studied of all chemicals from a toxicologic perspective. Pesticides are diverse in mode of action and in excipients, raising doubts about attempts to group pesticides for analysis except under very specific conditions (*e.g.*,

examining insecticides with similar mechanisms of *toxic* action and with similar excipients). In the interpretation of epidemiologic results, dose considerations from toxicology can play an important role in determining the plausibility of the response. Associations with exposures far below those causing effects in animals may be less credible than those demonstrated at higher exposure levels. The nature and limited amount of exposure information in the AHS makes this important use of toxicology difficult. More generally, toxicological reasoning has not yet played a significant role in the design or execution of the AHS but should be an important part of a detailed plan of analysis.

SUMMARY OF RESEARCH RECOMMENDATIONS

The AHS is a major undertaking with the potential to add significantly to our knowledge of possible associations between pesticide use and other factors and the health of farmers. The weight that will be accorded to results from this major study requires care in assuring the accuracy of the findings. Several of the most important limitations of the AHS could be addressed through additional research with the cohort or through complementary studies on different groups. The priorities should be to (1) assess the validity of self-reported health outcomes; (2) explore the reliability and validity of pesticide use data; (3) understand the relationship between exposure surrogates and exposure; (4) examine the biological plausibility of any hypotheses; and (5) develop explicitness on analysis and statistical issues.

Assessing the Validity of Self-Reported Health Outcomes

Many of the early analyses from the AHS will be based on self-reported health data. The validity of these data is crucial to interpretation of the results. There are studies in the literature that raise serious questions about self-reports of disease (Harlow and Linet, 1989; Paganini-Hill and Chao, 1993; Kehoe et al., 1994; The Italian Longitudinal Study on Aging Working Group, 1997). Clinical verification of key self-reported health outcomes, where feasible, is essential. It is important that validity be assessed for both those members of the cohort reporting disease and those who claim none. These studies could also help address some concerns about recall bias in the noncancer studies as well as concerns about whether the disease was indeed preceded by exposure.

Exploring the Reliability and Validity of Pesticide Use Data

Since pesticide use data will be the basis for categorizing potential pesticide exposure in the AHS, the validity of these data is also critical. A simple and pertinent step would be to readminister the questionnaire to a sample of respondents to see how much the answers change. Other studies to validate reported pesticide use, for example, by comparison with purchase records, are also essential. A relatively simple check would consist of questions about number of acres for each specific crop for which a specific pesticide was used. This would allow comparison to label instruc-

tions or Extension Service recommendations to help gauge the validity of use reports. Results of validation studies would suggest the amount of confidence that we could place in the questionnaire data as well as pinpoint ways to enhance the design of follow-up questionnaires. Validation studies will be able to address only relatively recent use since use records from the past are likely to be less complete. Given that many of the pesticides of concern for cancer were used more heavily in the past, and that a substantial period occurs between exposure and detection of disease, there may be significant questions about the validity of self-reported pesticide use in earlier years.

Understanding the Relationship between Exposure Surrogates and Exposure

Complementary studies are needed to assess the accuracy of the assumptions in the AHS that link specific use patterns and work practices with different levels of exposure. Biomonitoring studies could provide critical information to link pesticide use information to actual exposure by measuring pesticide levels in the blood or urine. Biomonitoring studies to correlate farmer exposure and dose to pesticide use patterns and work practices would be extraordinarily valuable in linking chemical use data to exposure categories. Similarly, biomonitoring studies of spouses and children of farmers could help determine whether conditions of pesticide use are associated with family exposures that are frequent enough and high enough to lead to possible adverse effects. This effort would help focus attention and resources on the most critical of possible adverse effects.

Assessing the Biological Plausibility of Any Associations

A key research need is the careful enumeration, in advance of analysis, of the biological effects expected at relevant doses for specific pesticides. This undertaking will help avoid the criticism that identified associations are supported only by toxicologic explanations that are *post hoc* and hence unreliable. This effort should rely on both the existing epidemiologic literature and the immense toxicologic database on pesticide products. Dose-response information must play a key role. Identification of chemicals expected to be capable of affecting health at anticipated exposures can corroborate findings and help focus analysis efforts.

Analysis and Statistical Issues

It is critical that a detailed analysis plan for the AHS be developed. Specifics to be addressed should start with a small number of precise hypotheses about pesticide/disease relationships, including in detail the analytic method. Potential confounders, interacting variables, and other risk factors should be identified in a systematic way, where possible, with a focus on causation of specific diseases. There is a need to specify an analytic framework, including specific statistical procedures, that encompasses decision rules for analysis and reporting.

The general study plan of the AHS is not yet detailed enough to support a confident evaluation of the technical strengths and weaknesses of this major undertaking, and we recommend substantial efforts toward developing such a plan. The level of effort and detail we are suggesting here would be typical of a major

investigator-initiated proposal that is peer reviewed and judged to be worthy of funding by the National Institutes of Health.

ACKNOWLEDGMENTS

Preparation of this report was a collaborative effort involving Drs. John D. Graham and George M. Gray of the Harvard Center for Risk Analysis and members of the Center's Advisory Committee on Agricultural Health Risks. Financial support for this work was provided to the Center by the American Crop Protection Association, a trade association whose members have commercial interests in the production of agricultural chemicals that are used widely on farms throughout the world. Preparation of the report was facilitated by a cooperative relationship between the Center and the principal investigators of the federal government's Agricultural Health Study. We are particularly thankful for the information and assistance provided by AHS team members Drs. Michael C. R. Alavanja, Dale P. Sandler, Shelia Hoar Zahm, Aaron Blair and David Mage. The AHS Advisory Panel, chaired by Dr. James Felton, provided useful information at their annual public advisory panel meetings. We would also like to thank the following scientists from the agricultural chemical industry who supplied useful information and encouragement throughout the project: Drs. John McCarthy, John Acquavella, Jon Amsel, James Gibson, Gerry Ott, George Rolofson, James Stevens, and Abe Tobia. The findings and opinions in this report should be attributed to the authors of the report and are not necessarily a reflection of the viewpoints of the Harvard Center for Risk Analysis, the principal investigators of the Agricultural Health Study, or the American Crop Protection Association.

REFERENCES

Agricultural Health Study Protocol, Revision of December 16, 1993.

Alavanja, M. C. R. Agricultural Health Study Information Packet for Advisory Panel Meeting, March 4–5, 1998 — NCI Progress Report.

Alavanja, M. C. R., Sandler, D. P., McMaster, S. B., Zahm, S. H., McDonnell, C. J., Lynch, C. F., Pennybacker, M., Rothman, N., Dosemeci, M., Bond, A. E., and Blair, A. 1996. The agricultural health study. *Environ. Health Perspect.* **104**, 362–369.

Bean, J. A., Leeper, J. D., Wallace, R. B., Sherman, B. M., and Jagger, H. 1979. Variations in the reporting of menstrual histories. *Am. J. Epidemiol.* **109**, 181–185.

Blair, A. and Zahm, S. H. 1993. Patterns of pesticide use among farmers: implications for epidemiologic research. *Epidemiology* **4**, 55–62.

Correa-Villasenor, A., Stewart, W. F., Franco-Marina, F., and Seacat H. 1995. Bias from nondifferential misclassification in case-control studies with three exposure levels. *Epidemiology.* **6**, 276–81.

Forbess, R. C., Morris, J. R., Lavy, T. L., Talbert, R. E., and Flynn, R. R. 1982. Exposure measurements of applicators who mix and spray paraquat in grape vineyards. *Horticult. Sci.* 956–957.

Harlow, S. D. and Linet, M. S. 1989. Agreement between questionnaire data and medical records. *Am. J. Epidemiol.* **129**, 233–248.

Hoar, S. K. *et al.* 1986. Agricultural herbicide use and risk of lymphoma and soft tissue sarcoma. *J. Am. Med. Assoc.* **256**, 1141–1147.

Hornung R. W., Deddens, J., and Roscoe R. 1995. Modifiers of exposure-response estimates for lung cancer among miners exposed to radon progeny. *Environ. Health Perspect.* **103**(Suppl. 2), 49–53.

Ibrahim, M. A. *et al.* 1991. Weight of the evidence on the human carcinogenicity of 2,4-D. *Environ. Health Perspect.* **96**, 213–222.

Johnson, R. A., Mandel, J. S., Gibson, R. W., Mandel, J. H., Bender, A. P., Gunderson, P. D. and Renier, C. M. 1993) Data on prior pesticide use collected from Self- and proxy respondents. *Epidemiology* **4**, 157–164.

Kehoe, R. *et al.* 1994. Comparing self-reported and physician-reported medical history. *Am. J. Epidemiol.* **139**, 813–818.

Lavy, T. L., Walstad, J. D., Flynn, R., and Mattice, J. D. 1982. (2,4-Dichlorophenoxy)acetic acid exposure received by aerial application crews during spray operations. *J. Agric. Food Chem.* **30**, 373–381.

Leng, M. L., Ramsey, J. C., Braun, W. H., and Lavy, T. L. 1982. Review of Studies with 2,4-Dichlorophenoxyacetic Acid in Humans Including Applicators Under Field Conditions. ACS Symposium Series No. 182. (Plimmer, J. R., Ed.), American Chemical Society.

Lowenherz, C., Fenske, R. A., Simcox, N. J., Bellamy, G., and Kalman, D. 1997. Biological monitoring of organophosphorous pesticide exposure among children of agricultural workers in central Washington state. *Environ. Health Perspect.* **105**, 1344–1353.

Monson, R. R. 1990. *Occupational Epidemiology*, 2nd ed. CRC Press, Boca Raton, FL.

Olson, J., Shu, X., Ross, J., Pendergrass, T., and Robison, L. 1997. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. *Am. J. Epidemiol.* **145**, 58–67.

Paganini-Hill, A. and Chao, A. 1993. Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. *Am. J. Epidemiol.* **138**, 101–106.

Sandler, D. P. Agricultural Health Study Information Packet for Advisory Panel Meeting, March 4 to 5, 1998 — NIEHS Progress Report.

Tarone, R. E., Alavanja, M. C. R., Zahm, S. H., Lubin, J. H., Sandler, D. P., McMaster, S. B., Rothman, N., and Blair, A. 1997. The agricultural health study: factors affecting completion and return of self-administered questionnaires in a large prospective cohort study of pesticide applicators. *Am. J. Indust. Med.* **31**, 233–242.

The Italian Longitudinal Study on Aging Working Group 1997. Prevalence of chronic diseases in older italians: comparing self-reported and clinical diagnoses. *Int. J. Epidemiol.* **26**, 995–1000.

Thomas D. C. 1995. re: When Will Nondifferential Misclassification of an Exposure Preserve the Direction of a Trend? *Am. J. Epidemiol.* **142**, 782–4.

USEPA (U.S. Environmental Protection Agency). Agricultural Health Study Information Packet for Advisory Panel Meeting, February 27 to 28, 1997 — USEPA Summary.

Wilcox, A. and Horney, L. 1984. Accuracy of spontaneous abortion recall. *Am. J. Epidemiol.* **120**, 727–733.

Zahm, S. H. *et al.* 1990. A case-control study of Non-Hodgkin's Lymphoma and the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D) in eastern Nebraska. *Epidemiology* **1**, 349–356.

ENDNOTES

1. Agricultural Health Study Information Packet for Advisory Panel Meeting, January 18–19, 1996 — Meeting Overview, NCI Summary, Biomarkers Studies,

1-2002

Reliability of Reporting on Life-Style and Agricultural Factors by a Sample of Participants in the Agricultural Health Study from Iowa

Aaron Blair
National Cancer Institute

Robert Tarone
National Cancer Institute

Dale Sandler
National Institute of Environmental Health

Charles F. Lynch
University of Iowa

Andrew Rowland
National Institute of Environmental Health Sciences

See next page for additional authors

Reliability of Reporting on Life-Style and Agricultural Factors by a Sample of Participants in the Agricultural Health Study from Iowa

Abstract
Repeat interviews from 4,088 Iowa pesticide applicators participating in the Agricultural Health Study provided the opportunity to evaluate the reliability of self-reported information on pesticide use and various demographic and life-style factors. Self-completed questionnaires were administered 1 year apart when participants returned to county agricultural extension offices for pesticide certification or training. Percentage agreement for ever-/never-use of specific pesticides and application practices was quite high, generally ranging from 70% to more than 90%, and did not vary by age, educational level, or farm size. Agreement was lower (typically 50–60%) for duration, frequency, or decade of first use of specific pesticides. Level of agreement regarding pesticide use in this population is similar to that generally found for factors typically used in epidemiologic studies such as tobacco use and higher than typically reported for diet, physical activity, and medical conditions.

Disciplines
Agriculture | Entomology | Environmental Studies

Comments
This article is published as Blair, Aaron, Robert Tarone, Dale Sandler, Charles F. Lynch, Andrew Rowland, Wendy Wintersteen, William C. Steen, Claudine Samanic, Mustafa Dosemeci, and Michael CR Alavanja. "Reliability of reporting on life-style and agricultural factors by a sample of participants in the Agricultural Health Study from Iowa." *Epidemiology* 13, no. 1 (2002): 94-99. Posted with permission.

Authors
Aaron Blair, Robert Tarone, Dale Sandler, Charles F. Lynch, Andrew Rowland, Wendy Wintersteen, and William C. Steen

Reliability of Reporting on Life-Style and Agricultural Factors by a Sample of Participants in the Agricultural Health Study from Iowa

Aaron Blair,[1] *Robert Tarone*,[1] *Dale Sandler*,[2] *Charles F. Lynch*,[3] *Andrew Rowland*,[2] *Wendy Wintersteen*,[4] *William C. Steen*,[5] *Claudine Samanic*,[1] *Mustafa Dosemeci*,[1] *and Michael C. R. Alavanja*[1]

Abstract: Repeat interviews from 4,088 Iowa pesticide applicators participating in the Agricultural Health Study provided the opportunity to evaluate the reliability of self-reported information on pesticide use and various demographic and life-style factors. Self-completed questionnaires were administered 1 year apart when participants returned to county agricultural extension offices for pesticide certification or training. Percentage agreement for ever-/never-use of specific pesticides and application practices was quite high, generally ranging from 70% to more than 90%, and did not vary by age, educational level, or farm size. Agreement was lower (typically 50–60%) for duration, frequency, or decade of first use of specific pesticides. Level of agreement regarding pesticide use in this population is similar to that generally found for factors typically used in epidemiologic studies such as tobacco use and higher than typically reported for diet, physical activity, and medical conditions. (EPIDEMIOLOGY 2002;13:94–99)

Key words: pesticides, reliability, agriculture.

The Agricultural Health Study is a long-term prospective cohort study designed to evaluate cancer and other diseases among farmers and their families in relation to agricultural exposures and life-style factors.[1] Farmers in Iowa and North Carolina, as well as commercial applicators in Iowa, were asked to participate in the study when they sought pesticide licenses or training at county agricultural extension offices. Self-completed questionnaires were used to obtain information on agricultural exposures and other factors necessary to evaluate disease risks.

Although questionnaires have long been used by the Economic Research Service of the U.S. Department of Agriculture (USDA) to obtain information on agricultural practices from farmers,[2,3] and farmers can provide considerable detail about their pesticide practices,[4,5] few studies have evaluated the reliability of information obtained on agricultural practices in epidemiologic investigations.[6] We took advantage of a special situation in Iowa to assess the reporting consistency for agricultural and life-style factors on a sample of the cohort that completed two questionnaires approximately 1 year apart.

Subjects and Methods

Participants in the Agricultural Health Study enrolled by completing a self-administered questionnaire when they came to the county agricultural extension offices to seek pesticide certification and training. At the beginning of the study, applicators in Iowa were required to take an examination in which a passing mark gained them certification for 3 years. Enrollment and completion of the questionnaires occurred from 1994 through 1996. After initiation of the study, the Iowa legislature changed procedures regarding pesticide certification for private applicators by allowing annual training as an alternative to the examination. This change meant that some individuals who chose pesticide training and completed the enrollment questionnaire one year would

From the [1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; [2]Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC; [3]Department of Epidemiology, University of Iowa, Iowa City, IA; [4]Iowa State University Extension Service, Ames, IA; and [5]U.S. Environmental Protection Agency, Athens, GA.

Address correspondence to: Aaron Blair, Division of Cancer Epidemiology and Genetics, National Cancer Institute, EPS Room 8118, Bethesda, MD 20892; blaira@mail.nih.gov

This work was supported by intramural funds from the National Cancer Institute.

Submitted May 8, 2001; final version accepted September 16, 2001.

TABLE 1. Comparison of Dichotomous Responses on Pesticide Use between First and Second Questionnaires

Factor	% Exact Agreement	Kappa Statistic	95% CI	No. of Subjects
Ever mixed or applied pesticides	95	0.15	0.08–0.22	3,634
Ever mixed or applied				
Atrazine	86	0.62	0.58–0.64	3,802
Glyphosate	82	0.54	0.52–0.58	3,763
Trifluarlin	87	0.71	0.68–0.73	3,760
2,4-D	87	0.48	0.44–0.52	3,786
Terbufos	83	0.66	0.63–0.68	3,712
Fonofos	84	0.63	0.60–0.66	3,674
Chlorpyrifos	81	0.61	0.59–0.64	3,728
Permethrin	88	0.59	0.55–0.62	3,665
Malathion	81	0.54	0.52–0.57	3,805
Carbaryl	79	0.57	0.55–0.60	3,643
DDT	87	0.63	0.60–0.66	3,599
Method of application				
Do not apply	90	0.19	0.14–0.24	4,112
Airblast	99	0.26	0.11–0.40	4,112
Boom	76	0.37	0.34–0.40	4,112
Hand spray gun	72	0.39	0.36–0.42	4,112
Backpack	85	0.48	0.44–0.51	4,112
Mist blower	94	0.46	0.40–0.52	4,112
Aerial	98	0.19	0.09–0.30	4,112
In furrow	76	0.51	0.48–0.54	4,112
Seed treatment	78	0.42	0.38–0.45	4,112
Pour fumigant	97	0.11	0.04–0.18	4,112
Gas canister	98	0.20	0.10–0.30	4,112
Powder	88	0.31	0.26–0.35	4,112
Inject animals	78	0.50	0.48–0.53	4,112
Dip animals	85	0.37	0.33–0.41	4,112
Spray animals	75	0.50	0.47–0.53	4,112
Ear tags	84	0.56	0.53–0.59	4,112
Dust/pour on animals	77	0.45	0.42–0.48	4,112

DDT = dichlorodiphenyltrichloroethane; 2,4-D = 2,4 dichlorophenoxyacetic acid.

return for training the following year, which provided us the opportunity to compare information from questionnaires completed 1 year apart. Individuals returning for training who had already enrolled in the Agricultural Health Study were given a questionnaire containing a subset of the questions originally asked on selected pesticide practices and life-style factors.

The abbreviated questionnaire was administered at the county agricultural extension office in the same fashion as the original enrollment version.[1] It was completed by 2,895 applicators (2,842 men and 53 women). In addition, a second group of 1,193 applicators who completed the enrollment questionnaire and returned for training the following year inadvertently completed the full enrollment questionnaire a second time. An analysis of the separate and combined data revealed that these two groups were similar with respect to age, gender, marital status, education, and other factors (data not shown). We combined these two groups for analysis, which resulted in 4,088 respondents (4,008 men and 80 women). Most of these were private applicators (3,829), but 259 were commercial applicators. We compared responses on the first and second questionnaires by calculating percentage exact agreement, percentage agreement within one category (for quantitative or ordinal categories), and kappa statistic.[7] We calculated weighted kappas for multiple-response questions such as years or days per year of pesticide application.

Results

The reliability subgroup was very similar to the full cohort for various factors including age, gender, marital status, and farm size (data not shown).

Comparability of reported use of pesticides and method of application is shown in Table 1. Although agreement on ever mixed or applied any pesticide was high (95%), kappa was considerably lower (0.15) because the proportion of subjects who had never used pesticides was very small. The reliability of reported use for specific chemicals was high and quite consistent (from 79% for carbaryl to 88% for permethrin) with no apparent differences by pest class, chemical class, or prevalence of use. Kappas ranged from 0.48 to 0.71. For the method of pesticide application, agreement of responses from the two questionnaires ranged from 72% to 99%. The range of values for kappa was from 0.11 to 0.56.

We evaluated the number of paired responses falling above and below the diagonal of exact agreement, *ie*, individuals providing different responses on the two questionnaires. For ever mixed or applied specific pesticides, the numbers were typically equally distributed above and below the diagonal of agreement, with the possible exception of malathion, carbaryl, and dichlorodiphenyltrichloroethane, for which considerably more positive responses were found in the original than in the second questionnaire (either follow-up abbreviated questionnaire or duplicated enrollment questionnaire).

The numbers of pairs falling above and below the diagonal were also similar for method of pesticide application. Sign tests of distributions were not statistically significant for ever-use of pesticides or method of application. Thus, there was no obvious tendency to get "yes" or "no" responses from one or the other of the questionnaires.

For years and days per year of mixing or applying any pesticides, exact agreement percentages were 55% and 45%, respectively, and weighted kappas were 0.56 and 0.45, respectively (Table 2). For each factor, 89% and 90% of the subjects were within one category of agreement.

TABLE 2. Comparison of Multiresponse Questions on Pesticide Use between First and Second Questionnaires

Factor	% Exact Agreement	Kappa Statistic*	95% CI	No. of Subjects
Years mixed or applied pesticides†	55	0.56	0.54–0.58	3,550
Days per year pesticides mixed or applied‡	45	0.45	0.42–0.47	3,494
% time mixed§	72	0.39	0.36–0.42	3,479
% time applied‖	79	0.47	0.44–0.50	3,471
Years mixed or applied†				
Atrazine	50	0.52	0.49–0.54	2,651
Glyphosate	53	0.70	0.66–0.75	2,379
Trifluralin	53	0.52	0.50–0.55	2,286
2,4-D	50	0.54	0.52–0.56	2,919
Terbufos	55	0.47	0.44–0.50	1,334
Fonofos	58	0.48	0.43–0.53	798
Chlorpyrifos	55	0.45	0.41–0.48	1,255
Permethrin	77	0.37	0.30–0.44	407
Days per year mixed or applied‡				
Atrazine	52	0.74	0.71–0.78	2,613
Glyphosate	52	0.71	0.67–0.75	2,342
Trifluralin	57	0.80	0.76–0.83	2,257
2,4-D	50	0.48	0.45–0.50	2,860
Terbufos	59	0.82	0.78–0.86	1,310
Fonofos	55	0.78	0.72–0.85	780
Chlorpyrifos	57	0.50	0.47–0.54	1,226
Permethrin	50	0.37	0.29–0.44	403
Decade first applied				
Atrazine	64	0.59	0.57–0.62	2,448
Glyphosate	62	0.37	0.34–0.40	2,098
Trifluralin	63	0.54	0.51–0.56	2,104
2,4-D	62	0.63	0.61–0.65	2,548
Terbufos	60	0.40	0.36–0.45	1,205
Fonofos	60	0.44	0.39–0.50	729
Chlorpyrifos	60	0.37	0.25–0.35	1,128
Permethrin	61	0.34	0.20–0.37	375

* Kappa values weighted across multiple response categories.
† Years of use categories are: 1 or less, 2–5, 6–10, 11–20, 21–30, and more than 30.
‡ Days per year of use categories are: less than 5, 5–9, 10–19, 20–39, 40–59, 60–150, and more than 150.
§ Refers to the percentage of all pesticides mixed on the farm that are mixed by this operator.
‖ Refers to the percentage of all pesticides applied on the farm that are applied by this operator. 2,4-D = 2,4 dichlorophenoxyacetic acid.

Exact agreement for estimated percentage of the time subjects mixed or applied pesticides was 72% and 79%, respectively, with kappas of 0.39 and 0.47 (Table 2). Agreement within one category of exact agreement was 98% and 99%. In addition, exact agreement for years, days per year, and decade of use of specific pesticides was generally in the 50–70% range, which was lower than for dichotomous outcomes such as ever-/never-use (Table 1). Ninety per cent of the subjects gave responses within one category of agreement on the two questionnaires. Kappas were between 0.37 and 0.63.

Responses regarding frequency of reported symptoms after using pesticides were similar in the two questionnaires, as shown in Table 3. Exact agreement ranged from 76% to 92% with kappas from 0.31 to 0.45. Agreement within one response category was nearly 100%.

We evaluated level agreement by age, education, and farm size. For ever-handled any pesticide, the level of exact agreement was 94% or higher, with kappas ranging from 0.10 to 0.18. Exact agreement for ever-handled specific pesticides was between 80% and 90%, with kappas ranging from 0.42 to 0.73 (Table 4). Level of agreement for methods of application did not differ by age, amount of education, or farm size (data not shown).

We also compared responses for tobacco use and reported disease histories. Agreement was very high (over 90%) for smoking cigarettes and quite high for number of cigarettes per day (76%). Kappas were also high (0.71 for number of cigarettes per day and 0.87 for ever-smoked cigarettes). Percentage agreement for diseases the subject reported for themselves and their relatives was more than 90%. Kappas were more variable with values of 0.71 for asthma, 0.65 for pneumonia, 0.34 for kidney disease, and 0.10 for Parkinson's disease. The kappa for Parkinson's disease was low, despite high exact agreement, because there were a relatively small number of persons reporting this disease. Kappas for cancers among relatives were 0.64 for lung, 0.70 for breast, and 0.41 for the lymphatic and hematopoietic system.

Percentage agreement and kappas calculated for agricultural and life-style factors for commercial applicators separately were essentially identical to those of private applicators (data not shown).

About one-third of the study group completed the full enrollment questionnaire twice (N = 1,193). We used these data to compare reliability of responses to questions not included in the abbreviated follow-up questionnaire, including alcohol drinks per day (71% exact agreement; kappa = 0.63), vegetable servings per day (35% exact agreement; kappa = 0.43), and fruit servings per day (40% exact agreement; kappa = 0.49).

TABLE 3. Comparison of Responses on Frequency of Reported Symptoms from Pesticide Use

Symptom Reported*	% Exact Agreement	Kappa Statistic	95% CI	No. of Subjects
Excessive tiredness	82	0.38	0.34–0.42	3,678
Headaches/dizziness	76	0.45	0.42–0.48	3,748
Nausea/vomiting	92	0.31	0.25–0.36	3,619
Skin irritation	79	0.40	0.37–0.44	3,683
Eye irritation	82	0.35	0.31–0.39	3,628
Chest discomfort	91	0.38	0.32–0.42	3,633
Nervousness or depression	87	0.41	0.37–0.45	3,628

* Response categories for symptoms from pesticide use were: never or rarely, sometimes, and frequently/almost always.

TABLE 4. Comparison of Dichotomous Responses on Pesticide Use between First and Second Questionnaires by Age, Education, and Farm Size

	Age, Years				Education				Acres			
	<50		≥50		≤High School		>High School		<500		≥500	
Factor	% Exact Agreement	Kappa	% Exact Agreement	Kappa	% Exact Agreement	Kappa	% Exact Agreement	Kappa	% Exact Agreement	Kappa	% Exact Agreement	Kappa
Ever mixed or applied pesticides	95	0.18	95	0.10	94	0.14	97	0.15	96	0.12	97	0.10
Ever mixed or applied												
Atrazine	85	0.62	87	0.60	85	0.58	88	0.67	85	0.61	87	0.57
Glyphosate	82	0.54	81	0.56	80	0.55	84	0.50	81	0.57	84	0.50
Trifluralin	87	0.71	87	0.70	87	0.70	88	0.72	87	0.73	88	0.65
2,4-D	85	0.51	88	0.42	85	0.46	89	0.51	87	0.54	87	0.41
Terbufos	83	0.65	83	0.66	82	0.64	84	0.68	82	0.64	83	0.66
Fonofos	85	0.61	84	0.65	84	0.62	85	0.63	84	0.62	85	0.63
Chlorpyrifos	81	0.61	81	0.62	79	0.58	82	0.64	81	0.62	81	0.61
Permethrin	87	0.58	91	0.58	90	0.60	86	0.57	88	0.59	88	0.58
Malathion	82	0.57	81	0.51	79	0.54	84	0.55	81	0.57	81	0.50
Carbaryl	80	0.58	78	0.56	78	0.54	80	0.61	79	0.58	78	0.56
DDT	93	0.42	87	0.55	85	0.62	89	0.65	86	0.64	87	0.63

DDT = dichlorodiphenyltrichloroethane; 2,4-D = 2,4 dichlorophenoxyacetic acid.

Discussion

The USDA has used questionnaires to assemble information on pesticide use by farmers for many years.[2,3] Use of interviews to obtain information on pesticide use and exposure in epidemiologic research is a more recent phenomenon.[4] There are differences between the approach used by the USDA and the approach required for epidemiologic research. The USDA typically obtains information on the past year. The long latency associated with most chronic disease, however, requires that epidemiologic studies obtain data on pesticide use from several years in the past, underscoring the need to evaluate the reliability and validity of information on pesticides obtained by interview.[8]

In the present analyses, the time frame for questions on pesticide use and other factors covered the subject's entire farming history of use up to the interview, so for many this required quite lengthy recall. Several general patterns were observed. First, agreement for self-reported smoking, selected diseases, and other factors in this population was consistent with other reports in the literature, *ie*, in the 90% range.[9–11,14,19–21] Second, the reported agreement for ever-/never-use of specific pesticides is also quite high, *ie*, mostly in the 70–90% range; these compare favorably with the reliability reported in other studies for factors such as smoking and alcohol use and are better than those reported for diet, physical activity, and health conditions. Third, the level of agreement on pesticide reporting decreased as the amount of detail sought increased, such as the number of years a person applied specific pesticides instead of ever-/never-use. This is similar to other factors such as tobacco use, in which agreement for the number of cigarettes smoked per day is lower than reporting of ever having smoked. Fourth, for pesticide factors, as well as for life-style factors and disease, the disagreements between the enrollment and follow-up interviews were symmetrical, *ie*, there was no general tendency for a higher prevalence of positive reporting in one or the other questionnaire administration. For example, in situations in which subjects reported at one interview that they used a particular pesticide but not at the other, the number of positive reports was about equivalent for the enrollment and follow-up interviews. This suggests that the additional year of farming experience before the completion of the follow-up questionnaire had little impact on the amount of disagreement. If dramatic changes occurred in the farming operation from one year to the next, we might have expected a disproportionate number of positive or negative responses to specific responses on the follow-up questionnaire. Still, some of the disagreements could be due to changes in pesticide application activities by study participants between the first and second questionnaires. Fifth, for questions with quantitative or ordinal responses, percentage agreement within one response category was quite high, typically 80% or higher. This is especially important for epidemiologic studies because responses are often grouped into a few categories. Sixth, percentage agreement did not differ by age, level of education, or farm size, which suggests a relatively consistent reliability of reporting among the various subgroups of the cohort. Finally, it is interesting that the values for number of cigarettes per day, years of pesticide use, and days per year of pesticide use from the abbreviated questionnaire were virtually identical to those on the original enrollment questionnaire, which indicates that these reliability results are applicable to the entire cohort.

The kappas for interview-reinterview for pesticides generally ranged from 0.20 to 0.50. Although perfect

agreement would generate a value of 1.0, values much lower can represent good agreement. This is because kappas are highly dependent on prevalence of the characteristic in the population, as well as on the sensitivity and specificity of the measure.[12] Thompson and Walter[12] have shown that for factors with a true prevalence of 0.2 to 0.8 and sensitivities and specificities that are quite high (in the 70–90% range), kappas fall into a range 0.3–0.6. Most values we observed are in this range. The few exceptions of kappas outside this range are for reporting on rare diseases or activities performed by nearly all applicators. Kappas can be quite low, and level of exact agreement is high for situations in which the factor is very prevalent or extremely rare. This was the situation when we observed low kappas.

The dependence of kappa on response prevalence is most obvious for questions with dichotomous response options to which few subjects gave a particular response (*eg*, "Did you pour fumigants?" or "Have you been diagnosed with Parkinson's disease?") or questions to which almost all subjects gave a particular response (*eg*, "Have you ever personally handled pesticides?"). In such situations the percentage exact agreement will always be near 100% (*ie*, almost all subjects give the majority response on both questionnaires), but kappas may be quite low (*eg*, if the few minority responses tend to come from different subjects for the second questionnaire than for the first questionnaire).

Few reports are available on the reliability of reported pesticide use specifically among farmers. Van Der Gulden *et al*[13] found 82% agreement and kappas of 0.55 for reported occupational exposure to pesticides from a reliability study in the Netherlands. Farrow *et al*[14] found kappas of about 0.29 for weed killers and 0.53 for pesticides/insecticides in general for women completing self-reported questionnaires before and after a miscarriage. These are somewhat lower than we found for specific pesticides (range of 0.48–0.70), but this might be expected because women in the miscarriage study were not from farms where pesticide use had a very important economic component, which might facilitate recall.

Several reports have compared reported pesticide use among farmers and surrogates,[4,5,15–18] and these provide a framework for considering results in this study. Agreement between farmers and pesticide suppliers regarding farmers' use of pesticides was about 50–60%.[4] Agreement between farmers and surrogates (primarily wives) on reported use of specific pesticides was about 50–70%.[5,17] A comparison of self-assessed and expert-assessed exposure to pesticides and fertilizers found an agreement of 91% with a kappa of 0.53 in a case-control study from Montreal.[18] The level of agreement we observed from repeat interviews is generally better than that from comparisons between subjects and surrogates.

Although the reliability of reported pesticide use among Iowa farmers is as good as for many other factors assessed by questionnaires in epidemiologic research and better than for some variables,[9–11,14,19–21] it is important to assess effects of potential misclassification on estimates of relative risk. If the level of agreement between the first and second interviews is considered a measure of nondifferential exposure misclassification, we can calculate effects on relative risks.[22] For example, if the true relative risk was 4.0 and nondifferential misclassification for ever-/never-handled individual pesticides is as in Table 1 (from 79% to 88% agreement), the calculated relative risks would range from 2.0 to 2.6. If the true relative risk was 2.0, calculated relative risks for individuals pesticides would be from 1.1 to 1.6. Even though the level of agreement is quite high, the impact of misclassification in this range on the relative risks can be substantial and diminish the opportunity to detect real associations. It is important to note that nondifferential misclassification, *ie*, misclassification that does not differ by presence or absence of disease, would only diminish estimates of relative risk for dichotomous classifications in a prospective investigation such as the Agricultural Health Study. It could, however, result in an increase or decrease in calculated relative risks in multiple response situations for the middle exposure categories, but not for the upper exposure category. In the upper exposure category, nondifferential misclassification would always diminish the relative risk.[23] Although these data suggest that pesticide use is reliably reported by farmers in this cohort, it is important to underscore that they do not provide information on the validity of these reports.

In summary, agreement for self-reported use of pesticides by farmers is similar to that found for other factors routinely evaluated by questionnaire in epidemiology studies such as smoking and alcohol reporting, and better than others such as consumption of fruits and vegetables and physical activity. Because epidemiologic studies have successfully related disease risk to these factors, it seems likely that information on pesticide use from interviews can also be used successfully to address exposure-disease relationships.

References

1. Alavanja MCR, Sandler DP, McMaster SB, *et al*. The Agricultural Health Study. *Environ Perspect* 1996;104:362–369.
2. Blake HT, Andrilenas PA, Jenkins RP, Eichers TR, Fox AS. Farmers' Pesticide Expenditures in 1966. Agricultural Economic Report No. 192. Washington DC: Economic Research Service, 1970.
3. Blake HT, Andrilenas PA. Farmers' Use of Pesticides in 1971. Agricultural Economic Report No. 296. Washington DC: Economic Research Service, U.S. Department of Agriculture, 1975.

4. Blair A, Zahm SH. Patterns of pesticide use among farmers: implications for epidemiologic research. *Epidemiology* 1993;4:55–62.
5. Blair A, Stewart PA, Kross B, *et al*. Comparison of two techniques to obtain information on pesticide use from Iowa farmers by interview. *J Agric Safety Health* 1997;3:229–236.
6. Blair A, Zahm SH. Methodologic issues in exposure assessment for case-control studies of cancer and herbicides. *Am J Ind Med* 1990;18:285–293.
7. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70:213–220.
8. Blair A, Zahm SH, Cantor KP, Stewart PA. Estimating exposure to pesticides in epidemiological studies of cancer. In: Wang RGM, Franklin CA, Honeycutt RC, Reinert JC, eds. Biologic Monitoring for Pesticide Exposure. ACS Symposium Series, No. 382. Washington DC: American Chemical Society, 1989;38–46.
9. Persson P-G, Norell SE. Retrospective *vs* original information on cigarette smoking. *Am J Epidemiol* 1989;130:705–712.
10. Kelly JP, Rosenberg L, Kaufman DW, Shapiro S. Reliability of personal interview data in a hospital-based case-control study. *Am J Epidemiol* 1990;131:79–90.
11. Jain M, Howe GR, Rohan T. Dietary assessment in epidemiology: comparison of a food frequency and diet history questionnaire with a 7-day food record. *Am J Epidemiol* 1996;143:953–960.
12. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol* 1998;41:949–958.
13. Van Der Gulden JWJ, Jansen IW, Verbeek ALM, Kolk JJ. Repeatability of self-reported data on occupational exposure to particular compounds. *Int J Epidemiol* 1993;22:284–287.
14. Farrow A, Farrow SC, Little R, Golding J, ALSPAC Study Team. The repeatability of self-reported exposure after miscarriage. *Int J Epidemiol* 1996;25:797–806.
15. Brown LM, Dosemeci M, Blair A, Burmeister L. Comparability of data obtained from farmers and surrogate respondents on use of agricultural pesticides. *Am J Epidemiol* 1991;134:348–355.
16. Boyle CA, Brann EA, and the Selected Cancers Cooperative Study Group. Proxy respondents and the validity of occupational and other exposure data. *Am J Epidemiol* 1992;136:712–721.
17. Johnson RA, Mandel JS, Gibson RW, *et al*. Data on prior pesticide use collected for self and proxy respondents. *Epidemiology* 1993;4: 157–164.
18. Fritschi L, Siemiatycki J, Richardson L. Self-assessed *vs* expert-assessed occupational exposures. *Am J Epidemiol* 1996;144:521–527.
19. Herrmann N. Retrospective information from questionnaires. II. Intrarater reliability and comparison of questionnaire types. *Am J Epidemiol* 1985;121:948–953.
20. Blair SN, Dowda M, Pate RR, *et al*. Reliability of long-term recall of participation in physical activity by middle-aged men and women. *Am J Epidemiol* 1991;133:266–275.
21. Hahn RA, Truman BI, Barker ND. Identifying ancestry: the reliability of ancestral identification in the United States by self, proxy, interviewer, and funeral director. *Epidemiology* 1996;7:75–80.
22. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488–495.
23. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746–748.

A Review of Pesticide Exposure and Cancer Incidence in the Agricultural Health Study Cohort

Scott Weichenthal, Connie Moase, and Peter Chan

Health Canada, Ottawa, Ontario, Canada

Objective: We reviewed epidemiologic evidence related to occupational pesticide exposures and cancer incidence in the Agricultural Health Study (AHS) cohort.

Data sources: Studies were identified from the AHS publication list available at http://aghealth.nci.nih.gov as well as through a Medline/PubMed database search in March 2009. We also examined citation lists. Findings related to lifetime-days and/or intensity-weighted lifetime-days of pesticide use are the primary focus of this review, because these measures allow for the evaluation of potential exposure–response relationships.

Data synthesis: We reviewed 28 studies; most of the 32 pesticides examined were not strongly associated with cancer incidence in pesticide applicators. Increased rate ratios (or odds ratios) and positive exposure–response patterns were reported for 12 pesticides currently registered in Canada and/or the United States (alachlor, aldicarb, carbaryl, chlorpyrifos, diazinon, dicamba, *S*-ethyl-*N*,*N*-dipropylthiocarbamate, imazethapyr, metolachlor, pendimethalin, permethrin, trifluralin). However, estimates of association for specific cancers were often imprecise because of small numbers of exposed cases, and clear monotonic exposure–response patterns were not always apparent. Exposure misclassification is also a concern in the AHS and may limit the analysis of exposure–response patterns. Epidemiologic evidence outside the AHS remains limited with respect to most of the observed associations, but animal toxicity data support the biological plausibility of relationships observed for alachlor, carbaryl, metolachlor, pendimethalin, permethrin, and trifluralin.

Conclusions: Continued follow-up is needed to clarify associations reported to date. In particular, further evaluation of registered pesticides is warranted.

Key words: Agricultural Health Study, cancer, pesticides, review. *Environ Health Perspect* 118:1117–1125 (2010). doi:10.1289/ehp.0901731 [Online 5 May 2010]

Agricultural workers are exposed to a variety of chemical, physical, and biological hazards in the process of cultivating and harvesting crops and/or raising livestock (Litchfield 1999; Popendorf and Donham 1991; Shaver and Tong 1991; White and Cessna 1989). In addition to pesticides, occupational exposure to solvents, metals, engine exhaust, welding fumes, and grain dusts are prevalent in agriculture (Coble et al. 2002; Shaver and Tong 1991). However, the potential health effects of agricultural pesticide exposures are of particular interest, as these chemicals are designed to have adverse biological effects on target organisms. To address this concern, the Agricultural Health Study (AHS) was initiated in 1993 to explore the potential health effects of pesticide exposures in commercial pesticide applicators, farmers, and their families in Iowa and North Carolina, USA. The AHS is a collaborative research project including the U.S. National Cancer Institute, the U.S. National Institute of Environmental Health Sciences, and the U.S. Environmental Protection Agency (EPA).

Details of the AHS design have been described previously (Alavanja et al. 1996). Briefly, participants were recruited between 1993 and 1997 from pesticide applicator licensing facilities in Iowa and North Carolina, and an enrollment questionnaire was used to collect data on the duration and frequency of pesticide use. In the AHS, self-reported pesticide use serves as a surrogate measure of pesticide exposure, and a cumulative pesticide exposure index (termed intensity-weighted exposure-days) is used to weigh lifetime-days (LDs) of pesticide use based on mixing conditions, application methods, and use of personal protective equipment (Dosemeci et al. 2002). Applicators who completed the enrollment questionnaire were asked to complete a take-home questionnaire that collected detailed information on factors including occupational exposures, pesticide use, lifestyle, medical history, and diet. Two additional take-home questionnaires were provided for private applicators: a spouse questionnaire and a female/family health questionnaire. Commercial applicators were asked to complete a female health questionnaire if they were female, but spouses and children of commercial applicators were not included in the AHS. A total of 52,395 private pesticide applicators (farmers or nursery workers), 32,347 spouses of private applicators, and 4,916 commercial pesticide applicators were enrolled in the AHS. Applicators are primarily male (> 95%); spouses are predominantly female (99.3%) (Alavanja et al. 2005). Private applicators, commercial applicators, and spouses are all predominantly Caucasian (94.6–98.6%) (Alavanja et al. 2005). At enrollment, 65% of private applicators reported pesticide exposure for > 11 years relative to 32% and 18% for commercial applicators and spouses of private applicators, respectively (Alavanja et al. 2005). Twelve percent of private applicators, 2.5% of commercial applicators, and 3.6% of spouses reported > 30 years of pesticide use at enrollment (Alavanja et al. 2005). Information provided at enrollment (phase 1) was updated in phase 2 (1999–2003) and phase 3 (2003–2010) of the AHS.

Approximately 80% of eligible applicators completed the enrollment questionnaire (Tarone et al. 1997). Forty percent of eligible applicators completed both the enrollment questionnaire and the take-home questionnaire (Tarone et al. 1997). Seventy-four percent of eligible spouses were enrolled in the AHS (Engel et al. 2005). Applicators who completed the take-home questionnaire tended to be older than nonrespondents but were otherwise similar to respondents with respect to pesticide use practices, medical history, and other characteristics such as education, smoking, alcohol consumption, and diet (Tarone et al. 1997). The reliability of self-reported pesticide use did not vary substantially by age, education, or farm size, and respondents generally provided plausible information regarding the duration and year of first pesticide use (Blair et al. 2002; Hoppin et al. 2002b).

In this review we focused on epidemiologic studies of pesticide exposure and cancer incidence in the AHS cohort. Studies of physical injury (Sprince et al. 2002, 2003a, 2003b, 2003c, 2007), mortality (Blair et al. 2005a, 2005b; Lee et al. 2007a), respiratory disorders (Hoppin et al. 2002a, 2006a, 2006b, 2007a, 2007b, 2008; Valcin et al. 2007), neurologic symptoms (Kamel et al. 2005, 2007a), retinal degeneration (Kamel et al. 2000; Kirrane et al. 2005), diabetes (Montgomery et al. 2008; Saldana et al. 2007), menstrual cycle characteristics (Farr et al. 2004, 2006), hearing loss (Crawford et al. 2008), Parkinson's disease (Kamel et al. 2007b), changes in serum androgen levels (Martin et al. 2002), arthritis (De Roos et al. 2005b), depression (Beseler et al. 2006, 2008), and immune

Address correspondence to S. Weichenthal, Health Canada, 269 Laurier Avenue West, Ottawa, Ontario Canada, K1A 0K9. Telephone: (613) 957-2617. Fax: (613) 948-8482. E-mail: scott.weichenthal@hc-sc.gc.ca

Supplemental Material is available online (doi:10.1289/ehp.0901731 via http://dx.doi.org/).

The findings and conclusions in this report are those of the authors and do not necessarily reflect the views of Health Canada.

The authors declare they have no actual or potential competing financial interests.

Received 25 November 2009; accepted 5 May 2010.

responses (Cooper et al. 2004) have also been conducted as part of the AHS but are not discussed in this review.

Methods

We identified studies from the publication list on the AHS Web site (Agricultural Health Study 2009) and by searching the Medline/PubMed database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed) using the key words "Agricultural Health Study" and "pesticides" in combination with various cancer types including "leukemia," "prostate cancer," "colon cancer," and others. We also examined citation lists. Studies were included in this review if they were published before March 2009 and examined the relationship between pesticides and cancer in the AHS cohort. Findings related to LDs and/or intensity-weighted lifetime-days (IWLDs) of pesticide use are the primary focus of this review, because these measures allow for the evaluation of potential exposure–response relationships.

We identified 28 studies that examined the relationship between pesticide exposures and cancer incidence in the AHS cohort. A list of the pesticides and cancer types examined in these studies is provided in Supplemental Material (doi:10.1289/ehp.0901731). Prevalent cancer cases were excluded in all studies reviewed, and incident cases were identified by matching cohort members to state cancer registry files. Cohort members were matched to state death registries and the National Death Index to ascertain vital status, and current address records of the Internal Revenue Service, motor vehicle records, and pesticide license registries were used to identify cohort members who were alive at the end of follow-up but no longer resided in Iowa or North Carolina. Follow-up was censored at the time of cancer diagnosis, participant death, or movement out of state. In general, only a small fraction (< 5%) of study participants was lost to follow-up. Average follow-up times ranged from approximately 4 to 9 years, and when presented, standardized incidence ratios (SIRs) reflect cancer incidence in the AHS cohort relative to the populations of Iowa and North Carolina using age, sex, and race-specific incidence data. The primary pesticide exposure measures were LDs (the product of days of use per year and years of use) and IWLDs (a weighted measure of lifetime exposure-days that accounts for mixing conditions, application methods, and the use of personal protective equipment) (Dosemeci et al. 2002). Where appropriate, a number of potential confounding factors were included in Poisson/logistic regression models, including age, smoking history, alcohol consumption, education, race, sex, applicator type, state of residence, LDs of any pesticide exposure, pesticides highly correlated with the pesticide of interest, family history of cancer, year of enrollment, body mass index, sun exposure, susceptibility to sunburn, aspirin intake, physical activity, nonfarm occupational exposures, diet, menopausal status, maternal age at first birth, age at menarche, and age at menopause. Findings were reported as rate ratios (RRs) or odds ratios (ORs) and their associated 95% confidence intervals (CIs). Linear trend tests (alpha = 0.05) were used to evaluate exposure–response relationships by treating median values for each pesticide exposure category as a quantitative score or by using continuous values for LDs or IWLDs of exposure. Criteria for including a given cancer type in chemical-specific analyses (i.e., studies of one pesticide and multiple cancer types) differed between studies and ranged from a minimum total of 5 to 30 exposed cases. Alternatively, some authors specified that a given cancer type was included in chemical-specific analyses only if ≥ 4 cases were available in each category of exposure. In general, pesticides with < 5 exposed applicators were excluded from cancer-specific analyses (i.e., studies of one cancer type and multiple pesticides). Values for the median number of exposed cases included in the studies we reviewed are listed in Supplemental Material (doi:10.1289/ehp.0901731).

Results

Study findings are summarized below for individual cancer types examined in the AHS cohort to date. Pesticides associated with cancer are listed in Table 1. A number of studies examined pesticide exposures according to both LDs and IWLDs of exposure. RRs (or ORs) for both exposure measures are included in Table 1 if at least one of these measures was associated with a significantly increased risk of cancer ($p < 0.05$). ORs were reported by Alavanja et al. (2003, 2004), Andreotti et al. (2009), and Lee et al. (2007b), and all other studies in Table 1 reported RRs. All RRs and ORs in Table 1 refer to nonexposed applicators except those reported for alachlor and all lymphohematopoietic cancers (Lee et al. 2004b), dicamba and colon (Samanic et al. 2006), and lung cancer (Alavanja et al. 2004), which refer to applicators in the lowest category of exposure.

Study Summaries

All cancers. Overall cancer incidence was increased among applicators in the highest exposure categories for diazinon (Beane Freeman et al. 2005) and *S*-ethyl-*N*,*N*-dipropylthiocarbamate (EPTC) (van Bemmel et al. 2008) relative to nonexposed applicators, and significant exposure–response trends were observed for both pesticides with both exposure measures. None of the other pesticides examined were associated with a significant increase in overall cancer incidence (alachlor, atrazine, captan, carbaryl, carbofuran, chlorothalonil, cyanazine, dicamba, dichlorvos, fonofos, glyphosate, heptachlor, imazethapyr, malathion, metolachlor, pendimethalin, permethrin, phorate, and trifluralin) (Bonner et al. 2005, 2007; De Roos et al. 2005a; Greenburg et al. 2008; Hou et al. 2006; Kang et al. 2008; Koutros et al. 2008, 2009; Lee et al. 2004b; Lynch et al. 2006; Mahajan et al. 2006a, 2006b, 2007; Mozzachio et al. 2008; Rusiecki et al. 2004, 2006, 2009; Samanic et al. 2006).

Lung cancer. Applicators with the highest LDs of exposure to chlorpyrifos (Lee et al. 2004a), diazinon (Beane Freeman et al. 2005), dieldrin (Alavanja et al. 2004; Purdue et al. 2006), metolachlor (Alavanja et al. 2004), and pendimethalin (Alavanja et al. 2004; Hou et al. 2006) had increased lung cancer incidence relative to nonexposed applicators. Lung cancer risk was also significantly increased among applicators in the highest category of intensity-weighted dieldrin exposure days (Purdue et al. 2006), but diazinon and pendimethalin were not associated with lung cancer when exposures were analyzed according to this measure. Applicators with the highest LDs of dicamba exposure had increased lung cancer incidence relative to low-exposed applicators but not relative to nonexposed applicators (OR = 1.6; 95% CI, 0.7–3.4) (Alavanja et al. 2004). Significant exposure–response trends were observed for chlorpyrifos (Lee et al. 2004a), diazinon (Beane Freeman et al. 2005), dicamba (Alavanja et al. 2004), dieldrin (Alavanja et al. 2004; Purdue et al. 2006), metolachlor (Alavanja et al. 2004), and pendimethalin (Alavanja et al. 2004). However, the most recent study of pendimethalin exposure did not observe a significant exposure–response trend for lung cancer (Hou et al. 2006). Fourteen other pesticides were examined but were not associated with increased lung cancer incidence in pesticide applicators (atrazine, captan, carbaryl, chlorothalonil, cyanazine, dichlorvos, EPTC, fonofos, glyphosate, imazethapyr, malathion, permethrin, phorate, and trifluralin) (Alavanja et al. 2004; Bonner et al. 2007; De Roos et al. 2005a; Kang et al. 2008; Koutros et al. 2008, 2009; Greenburg et al. 2008; Mahajan et al. 2006a, 2006b, 2007; Mozzachio et al. 2008; Lynch et al. 2006; Rusiecki et al. 2004, 2009; Samanic et al. 2006; van Bemmel et al. 2008).

Pancreatic cancer. Applicators in the highest categories of intensity-weighted EPTC and pendimethalin exposure-days had an increased risk of pancreatic cancer relative to nonexposed applicators, and we observed significant exposure–response trends for both pesticides (Andreotti et al. 2009). Eleven other pesticides were examined but were not associated

Table 1. Pesticides associated with cancer in the AHS cohort.

Cancer type	Pesticide(s)	Chemical family	Categorical exposure cutoff value	RR or OR[a] (95% CI)	*p*-Value for trend	References
All cancers	Diazinon	OP	> 109 LD[b]	1.58 (1.10–2.28)	0.007	Beane Freeman et al. 2005
			Highest IWLD[b]	1.41 (1.03–1.95)	0.033	
	EPTC	Thiocarbamate	> 50 LD[c]	1.28 (1.09–1.50)	< 0.01	van Bemmel et al. 2008
			> 112 IWLD[c]	1.16 (1.01–1.35)	0.02	
Lung	Chlorpyrifos	OP	> 56 LD[b]	2.18 (1.31–3.64)	0.002	Lee et al. 2004a
			> 417 IWLD[b]	1.80 (1.00–3.23)	0.036	
	Diazinon	OP	> 109 LD[b]	3.46 (1.57–7.65)	0.001	Beane Freeman et al. 2005
			Highest IWLD[b]	1.55 (0.65–3.72)	0.22	
	Dicamba	Benzoic acid	> 224 LD[b]	3.10 (1.20–7.70)	0.04	Alavanja et al. 2004
	Dieldrin	OC	> 50 LD[b]	5.30 (1.50–18.6)	0.005	
			> 9 LD[d]	2.80 (1.10–7.20)	0.02	Purdue et al. 2006
	Metolachlor	Chloroacetanilide	Highest IWLD[d]	3.50 (1.60–7.70)	0.002	
			> 457 LD[b]	4.10 (1.60–10.4)	0.015	Alavanja et al. 2004
	Pendimethalin	Dinitroaniline	> 224 LD[b]	3.50 (1.10–10.5)	0.005	
			>116 LD[b]	2.40 (1.10–5.30)	0.29	Hou et al. 2006
			> 539 IWLD[b]	1.10 (0.50–2.60)	0.94	
Pancreas	EPTC	Thiocarbamate	> 118 IWLD[d]	2.50 (1.10–5.40)	0.01	Andreotti et al. 2009
	Pendimethalin	Dinitroaniline	> 117 IWLD[d]	3.00 (1.30–7.20)	0.01	
Colon	Aldicarb	Carbamate	> 56 LD[b]	4.10 (1.30–12.8)	0.001	Lee et al. 2007a
	Dicamba	Benzoic acid	> 116 LD[b]	3.29 (1.40–7.73)	0.02	Samanic et al. 2006
			> 739 IWLD[b]	2.57 (1.28–5.17)	0.002	
	EPTC	Thiocarbamate	> 50 LD[c]	2.09 (1.26–3.47)	< 0.01	van Bemmel et al. 2008
			> 112 IWLD[c]	2.05 (1.34–3.14)	< 0.01	
	Imazethapyr	Imidazolinone	> 311 IWLD (proximal)[b]	2.73 (1.42–5.25)	0.001	Koutros et al. 2009
			> 311 IWLD (distal)[b]	1.21 (0.55–2.68)	0.75	
	Trifluralin	Dinitroaniline	> 224 LD[b]	1.48 (0.78–2.80)	0.12	Kang et al. 2008
			> 1176 IWLD[b]	1.76 (1.05–2.95)	0.036	
Rectum	Chlordane	OC	> 9 LD[d]	2.70 (1.10–6.80)	0.03	Purdue et al. 2006
			Highest IWLD[d]	2.10 (0.90–5.30)	0.04	
	Chlorpyrifos	OP	> 56 LD[b]	3.25 (1.60–6.62)	0.035	Lee et al. 2004a
			> 417 IWLD[b]	3.16 (1.42–7.03)	0.057	
			>109 LD[b]	2.70 (1.20–6.40)	0.008	Lee et al. 2007a
	Pendimethalin	Dinitroaniline	> 116 LD[c]	4.30 (1.50–12.7)	0.007	Hou et al. 2006
			> 539 IWLD[c]	3.60 (1.20–11.3)	0.02	
	Toxaphene	OC	> 56 LD[b]	4.30 (1.20–15.8)	0.123	Lee et al. 2007a
Leukemia	Chlordane/Heptachlor	OC	> 9 LD[d]	2.60 (1.20–6.00)	0.02	Purdue et al. 2006
			Highest IWLD[d]	2.10 (0.80–5.50)	0.10	
	Chlorpyrifos	OP	> 56 LD[b]	2.15 (0.96–4.81)	0.36	Lee et al. 2004a
			> 417 IWLD[b]	3.01 (1.35–6.69)	0.15	
	Diazinon	OP	> 39 LD[c]	3.36 (1.08–10.5)	0.026	Beane Freeman et al. 2005
			Highest IWLD[c]	2.88 (0.92–9.03)	0.053	
	EPTC	Thiocarbamate	> 50 LD[c]	2.36 (1.16–4.84)	0.02	van Bemmel et al. 2008
			> 112 IWLD[c]	1.87 (0.97–3.59)	0.05	
	Fonofos	OP	> 609 IWLD[c]	2.67 (1.06–6.70)	0.04	Mahajan et al. 2006a
All LH	Alachlor	Chloroacetanilide	> 116 LD[c]	2.04 (0.89–4.65)	0.02	Lee et al. 2004b
			> 710 IWLD[c]	2.42 (1.00–5.89)	0.03	
	Chlorpyrifos	OP	> 56 LD[b]	1.43 (0.86–2.36)	0.26	Lee et al. 2004a
			> 417 IWLD[b]	1.99 (1.22–3.26)	0.09	
	Diazinon	OP	> 39 LD[c]	1.84 (0.89–3.82)	0.094	Beane Freeman et al. 2005
			Highest IWLD[c]	2.01 (1.02–3.94)	0.049	
	Permethrin	Pyrethroid	> 50 LD[c]	1.64 (1.07–2.52)	0.35	Rusiecki et al. 2009
			> 220 IWLD[c]	1.31 (0.84–2.04)	0.60	
NHL	Lindane	OC	> 22 LD[d]	2.10 (0.80–5.50)	0.12	Purdue et al. 2006
			Highest IWLD[d]	2.60 (1.10–6.40)	0.04	
Multiple myeloma	Permethrin	Pyrethroid	> 50 LD[c]	5.72 (2.76–11.8)	< 0.01	Rusiecki et al. 2009
			> 220 IWLD[c]	5.01 (2.41–10.4)	< 0.01	
Bladder	Imazethapyr	Imidazolinone	> 311 IWLD[b]	2.37 (1.20–4.68)	0.01	Koutros et al. 2009
Prostate	Fonofos	OP	> 56 LD[c]	1.77 (1.03–3.05)	0.02	Mahajan et al. 2006a (for applicators
			> 315 IWLD[c]	1.83 (1.12–3.00)	0.01	with a family history of prostate cancer)
	Methylbromide	Halogenated alkane	Highest IWLD[e]	3.47 (1.37–8.76)	0.004	Alavanja et al. 2003
Brain	Chlorpyrifos	OP	> 56 LD[b]	2.58 (0.73–9.17)	0.076	Lee et al. 2004a
			> 417 IWLD[b]	4.03 (1.18–13.8)	0.036	
Melanoma	Carbaryl	Carbamate	> 175 LD[b]	4.11 (1.33–12.7)	0.07	Mahajan et al. 2007
			Highest intensity score[b]	1.54 (0.61–3.86)	0.92	
	Toxaphene	OC	> 25 LD[d]	2.90 (1.10–8.10)	0.03	Purdue et al. 2006
			Highest IWLD[d]	1.80 (0.70–5.10)	0.24	

Abbreviations: LH, lymphohematopoietic cancers; OC, organochlorine; OP, organophosphate. ORs were reported by Alavanja et al. (2003, 2004), Andreotti et al. (2009), and Lee et al. (2007b); all others are RRs.

[a]All RRs and ORs were estimated relative to nonexposed applicators except those reported for alachlor and all LH (Lee et al. 2004b) and dicamba and colon (Samanic et al. 2006) and lung cancer (Alavanja et al. 2004), which are in reference to applicators in the lowest category of exposure. [b]Highest quintile. [c]Highest quartile. [d]Highest tertile. [e]Highest sixth.

with pancreatic cancer [alachlor, atrazine, chlorpyrifos, cyanazine, 2,4-dichlorophenoxy acetic acid (2,4-D), dicamba, glyphosate, imazethapyr, metolachlor, terbufos, and trifluralin] (Andreotti et al. 2009; De Roos et al. 2005a; Rusiecki et al. 2004).

Colon and rectal cancer. Applicators with the highest LDs of exposure to aldicarb (Lee et al. 2007b), dicamba (Samanic et al. 2006), EPTC (van Bemmel et al. 2008), imazethapyr (Koutros et al. 2009), and trifluralin (Kang et al. 2008) had increased colon cancer incidence relative to nonexposed applicators. For imazethapyr, excess colon cancer incidence was limited to the proximal colon (Koutros et al. 2009); other studies did not examine colon cancer according to location. Significant exposure–response relationships were observed for aldicarb, dicamba, EPTC, imazethapyr, and trifluralin, but few cases were available in the first (n = 7) and second tertiles (n = 4) of intensity-weighted EPTC exposure days. Applicators with the highest lifetime exposure-days for chlordane (Purdue et al. 2006), chlorpyrifos (Lee et al. 2004a; 2007b), pendimethalin (Hou et al. 2006), and toxaphene (Lee et al. 2007b) had increased rectal cancer risk relative to nonexposed applicators. Rectal cancer risk was also significantly increased when chlorpyrifos and pendimethalin exposures were analyzed according to IWLDs (Hou et al. 2006; Lee et al. 2004a), and we observed significant exposure–response trends for chlordane (Purdue et al. 2006), chlorpyrifos (Lee et al. 2004a; 2007b), and pendimethalin (Hou et al. 2006). Seventeen other pesticides were examined but were not associated with colon or rectal cancer incidence (2,4-D, alachlor, aldrin, atrazine, captan, carbaryl, carbofuran, chlorothalonil, cyanazine, diazinon, dichlorvos, fonofos, glyphosate, malathion, metolachlor, permethrin, and phorate) (Beane Freeman et al. 2005; Bonner et al. 2005, 2007; De Roos et al. 2005a; Greenburg et al. 2008; Koutros et al. 2008; Lee et al. 2004b, 2007b; Lynch et al. 2006; Mahajan et al. 2006a, 2007; Mozzachio et al. 2008; Purdue et al. 2006; Rusiecki et al. 2004, 2006, 2009).

All lymphohematopoietic cancers. Applicators in the highest category of intensity-weighted alachlor exposure-days had an increased incidence of all lymphohematopoietic cancers relative to low-exposed applicators (Lee et al. 2004b). Applicators in the highest categories of intensity-weighted exposure-days for chlorpyrifos (Lee et al. 2004a) and diazinon (Beane Freeman et al. 2005) had an increased incidence of all lymphohematopoietic cancers relative to nonexposed applicators, but RRs were not significantly increased when exposures were analyzed according to LDs of use. Applicators in the highest category of permethrin exposure-days had an increased incidence of all lymphohematopoietic cancers but the RR for the highest category of intensity-weighted permethrin exposure-days was not significantly increased (Rusiecki et al. 2009). Significant exposure–response trends were observed for alachlor (Lee et al. 2004b) and diazinon (Beane Freeman et al. 2005). Fifteen other pesticides were examined but were not associated with lymphohematopoietic cancers (captan, carbaryl, carbofuran, cyanazine, dicamba, dichlorvos, EPTC, fonofos, glyphosate, imazethapyr, malathion, metolachlor, pendimethalin, phorate, and trifluralin) (Bonner et al. 2005, 2007; De Roos et al. 2005a; Greenburg et al. 2008; Hou et al. 2006; Kang et al. 2008; Koutros et al. 2008, 2009; Lee et al. 2004b; Lynch et al. 2006; Mahajan et al. 2006a, 2006b, 2007; Rusiecki et al. 2006; Samanic et al. 2006).

Leukemia. Applicators with the highest LDs of exposure for heptachlor/chlordane (Purdue et al. 2006), diazinon (Beane Freeman et al. 2005), and EPTC (van Bemmel et al. 2008) had increased leukemia incidence relative to nonexposed applicators. When exposures were analyzed according to intensity-weighted exposures-days, applicators in the highest categories of exposure for fonofos (Mahajan et al. 2006a) and chlorpyrifos (Lee et al. 2004a) also had increased leukemia incidence relative to nonexposed applicators. Significant dose–response relationships were observed for heptachlor/chlordane, diazinon, EPTC, and fonofos. Seven other pesticides were examined but were not associated with leukemia in pesticide applicators (alachlor, atrazine, carbaryl, glyphosate, imazethapyr, malathion, and trifluralin) (Bonner et al. 2007; De Roos et al. 2005a; Kang et al. 2008; Koutros et al. 2009; Mahajan et al. 2007; Rusiecki et al. 2004, 2009).

Non-Hodgkin lymphoma. Applicators in the highest category of intensity-weighted exposure-days for lindane had increased non-Hodgkin lymphoma (NHL) incidence relative to nonexposed applicators, and a significant exposure–response trend was observed (Purdue et al. 2006). However, we did not observe a significant trend for lindane when we analyzed exposures according to lifetime exposure-days. Sixteen other pesticides were examined but were not associated with NHL in pesticide applicators (alachlor, atrazine, carbaryl, carbofuran, chlorpyrifos, cyanazine, diazinon, dicamba, EPTC, glyphosate, imazethapyr, malathion, metolachlor, pendimethalin, permethrin, and trifluralin) (Beane Freeman et al. 2005; Bonner et al. 2005, 2007; De Roos et al. 2005a; Hou et al. 2006; Kang et al. 2008; Koutros et al. 2009; Lee et al. 2004a, 2004b; Lynch et al. 2006; Mahajan et al. 2007; Rusiecki et al. 2004, 2006, 2009; Samanic et al. 2006; van Bemmel et al. 2008).

Multiple myeloma. Applicators in the highest categories of permethrin exposure had an increased incidence of multiple myeloma relative to nonexposed applicators, and we observed significant exposure–response patterns (Rusiecki et al. 2009). However, ≤ 3 cases were available in the first and second tertiles of exposure; further evaluation of a potential exposure–response pattern is required once more cases have accrued. Four other pesticides were examined but were not associated with multiple myeloma (alachlor, atrazine, chlorpyrifos, and glyphosate) (De Roos et al. 2005a; Lee et al. 2004a, 2004b; Rusiecki et al. 2004).

Breast cancer. Engel et al. (2005) examined breast cancer incidence among farmers' wives. Breast cancer incidence was decreased among women who reported ever applying pesticides relative to the general population (SIR = 0.87; 95% CI, 0.89–1.24), and strong associations were not detected for specific pesticides. Ever use of pesticides in this study included use on crops and livestock as well as use in the home or garden. Although few women personally applied many of the pesticides examined, breast cancer incidence was increased among women whose husbands reported ever use of aldrin (RR = 1.9; 95% CI, 1.3–2.7), carbaryl (RR = 1.4; 95% CI, 1.0–2.0), chlordane (RR = 1.7; 95% CI, 1.2–5.5), dieldrin (RR = 2.0; 95% CI, 1.1–3.3), heptachlor (RR = 1.6; 95% CI, 1.1–2.4), lindane (RR = 1.7; 95% CI, 1.1–2.5), malathion (RR = 1.4; 95% CI, 1.0–2.0), 2,4,5-trichlorophenoxypropionic acid (2,4,5-TP) (RR = 2.0; 95% CI, 1.2–3.2) or captan (RR = 2.7; 95% CI, 1.7–4.3). RRs varied by menopausal status, with increased breast cancer incidence observed among premenopausal women ever exposed to chlorpyrifos (RR = 2.2; 95% CI, 1.0– 4.9), dichlorvos (RR = 2.3; 95% CI, 1.0–5.3), or terbufos (RR = 2.6; 95% CI, 1.1–5.9) but not among postmenopausal women. Potential exposure–response patterns were not examined in this study because pesticide exposure information was limited to ever/never use data. Other studies of pesticide exposure and breast cancer incidence in the AHS cohort have not been conducted to date.

Bladder cancer. Applicators in the highest category of intensity-weighted imazethapyr exposure-days had an increased incidence of bladder cancer relative to nonexposed applicators, and a significant exposure–response pattern was observed (Koutros et al. 2009). Nine other pesticides were examined but were not associated with bladder cancer in pesticide applicators (alachlor, atrazine, carbaryl, dicamba, EPTC, glyphosate, malathion, permethrin, and trifluralin) (Bonner et al. 2007; De Roos et al. 2005a; Kang et al. 2008; Lee et al. 2004b; Mahajan et al. 2007; Rusiecki et al. 2004, 2009; Samanic et al. 2006; van Bemmel et al. 2008).

Prostate cancer. Applicators in the highest categories of fonofos exposure who also had a family history of prostate cancer had increased prostate cancer incidence relative to nonexposed applicators (Mahajan et al. 2006a). Significant exposure–response patterns were observed for both exposure measures, and the reported findings suggest that a family history of prostate cancer may modify prostate cancer risk in applicators exposed to fonofos. Applicators in the highest categories of fonofos exposure without a family history of prostate cancer did not have increased prostate cancer incidence (RR_{LD} = 0.86; 95% CI, 0.60–1.24; RR_{IWLD} = 0.96; 95% CI, 0.70–1.31). Applicators in the highest category of intensity-weighted exposure-days for methyl bromide had increased prostate cancer risk relative to nonexposed applicators, and a significant exposure–response trend was observed (Alavanja et al. 2003). Twenty-three other pesticides were examined but were not associated with prostate cancer (alachlor, aldrin, atrazine, captan, carbofuran, carbaryl, chlorothalonil, chlorpyrifos, cyanazine, DDT, diazinon, dicamba, dichlorvos, EPTC, glyphosate, heptachlor, imazethapyr, malathion, metolachlor, pendimethalin, permethrin, phorate, and trifluralin) (Alavanja et al. 2003; Beane Freeman et al. 2005; Bonner et al. 2005, 2007; De Roos et al. 2005a; Greenburg et al. 2008; Hou et al. 2006; Kang et al. 2008; Koutros et al. 2008, 2009; Lynch et al. 2006; Mahajan et al. 2006b, 2007; Mozzachio et al. 2008; Rusiecki et al. 2004, 2006, 2009; Samanic et al. 2006; van Bemmel et al. 2008).

Brain cancer. Lee et al. (2004a) examined brain cancer incidence in pesticide applicators exposed to chlorpyrifos. Applicators in the highest category of intensity-weighted chlorpyrifos exposure-days had increased brain cancer incidence relative to nonexposed applicators, and a significant exposure–response pattern was observed. However, findings were based on small numbers of exposed cases ($2 \le n \le 7$) and the exposure–response trend was not monotonic, as the second-highest exposure group had a lower risk of brain cancer (RR = 1.25; 95% CI, 0.26–6.10) than applicators in the lowest category of exposure (RR = 3.32; 95% CI, 0.98–11.24). Elevated RRs were reported for the two highest categories of chlorpyrifos exposure-days, but these estimates were not significantly increased and a significant exposure–response trend was not observed. Other studies of pesticide exposure and brain cancer incidence in the AHS cohort have not been conducted to date.

Melanoma. Applicators in the highest categories of lifetime carbaryl (Mahajan et al. 2007) and toxaphene (Purdue et al. 2006) exposure-days had an increased incidence of melanoma relative to nonexposed applicators. Significant exposure–response patterns were not observed for carbaryl when exposures were analyzed according to lifetime exposure-days or intensity-weighted exposure-days (Mahajan et al. 2007). We observed a significant exposure–response trend with increasing LDs of toxaphene exposure but not when exposures were analyzed according to intensity-weighted exposure-days. Ten other pesticides were examined but were not associated with melanoma incidence in pesticide applicators (atrazine, diazinon, dicamba, EPTC, fonofos, glyphosate, imazethapyr, malathion, pendimethalin, and permethrin) (Beane Freeman et al. 2005; Bonner et al. 2007; De Roos et al. 2005a; Hou et al. 2006; Koutros et al. 2009; Mahajan et al. 2006a; Rusiecki et al. 2004, 2009; Samanic et al. 2006; van Bemmel et al. 2008).

Other cancers. **Kidney cancer.** Six studies examined the relationship between pesticide exposure and kidney cancer in pesticide applicators, but strong associations were not observed for any of the pesticides examined (atrazine, chlorpyrifos, glyphosate, imazethapyr, malathion, and trifluralin) (Bonner et al. 2007; De Roos et al. 2005a; Kang et al. 2008; Koutros et al. 2009; Lee et al. 2004a; Rusiecki et al. 2004). Increased RRs were reported for applicators in the highest categories of trifluralin exposure relative to nonexposed applicators (RR_{LD} = 2.06; 95% CI, 0.75–5.65; RR_{IWLD} = 1.77; 95% CI, 0.73–4.30), but these estimates were not significantly increased, and significant exposure–response trends were not observed for either exposure measure (Kang et al. 2008).

Childhood cancer. Flower et al. (2004) examined cancer incidence in children of male farmers in Iowa. Overall cancer incidence was increased among children of pesticide applicators (SIR = 1.36; 95% CI, 1.03–1.79), and more lymphoma (SIR = 2.18; 95% CI, 1.13–4.19) and Hodgkin lymphoma (SIR = 2.56; 95% CI, 1.06–6.14) cases were observed than expected based on childhood cancer rates in the Iowa population. However, SIRs for specific cancers were based on small numbers of cases ($2 \le n \le 11$). Cancer risk was increased among children whose fathers did not use chemically resistant gloves when mixing pesticides relative to children with fathers who wore gloves (OR = 1.98; 95% CI, 1.05–3.76). In addition, relative to children of nonexposed men, children whose fathers used aldrin during the prenatal period also had an increased cancer risk (OR = 2.66; 95% CI, 1.08–6.59). Potential exposure–response patterns were not examined for specific pesticides, however, and 15 other pesticides examined in this study were not associated with childhood cancer (alachlor, atrazine, chlorpyrifos, cyanazine, 2,4-D, dichlorvos, dicamba, EPTC, glyphosate, malathion, metolachlor, metribuzin, phorate, trifluralin, and terbufos).

Miscellaneous. Several studies examined the relationship between pesticide exposures and oral cavity cancers (De Roos et al. 2005a; Koutros et al. 2009; Rusiecki et al. 2004, 2006), stomach cancer (Lee et al. 2004b), esophagus cancer (Lee et al. 2004a; Rusiecki et al. 2004), and thyroid cancer (Lee et al. 2004b), but none of the six pesticides examined were associated with increased risk of these types of cancers (alachlor, atrazine, chlorpyrifos, glyphosate, imazethapyr, and metolachlor).

Discussion

Through March 2009, 27 studies examined the relationship between LDs or IWLDs of pesticide exposure and cancer incidence in the AHS cohort. Thirty-two different pesticides were included in these studies, and most study participants personally applied pesticides for 11– 30 years before enrollment (Alavanja et al. 2005). When appropriate, all studies adjusted for the use of pesticides correlated with the specific pesticide of interest; however, it is possible that this adjustment did not completely isolate the independent effects of each individual pesticide or eliminate the impact of multiple exposures. Nonetheless, findings from Coble et al. (2002) suggest that the magnitude of bias due to confounding from exposure to multiple agents is likely to be minimal based on the proportion of farmers reporting exposure to agents including cleaning solvents and diesel exhaust. Findings of chemical cohort analyses (i.e., studies of a single pesticide and multiple cancer types) and cancer site analyses (i.e., studies of a single cancer type and multiple pesticide exposures) were generally consistent with respect to the magnitude and direction of the observed associations. Specifically, chemical cohort and cancer site analysis were consistent for carbaryl and colon cancer (no association) (Lee et al. 2007b; Mahajan et al. 2007), chlorpyrifos and rectal cancer (significantly increased risk) (Lee et al. 2004a, 2007b), pendimethalin, permethrin, and colorectal cancer (no association) (Hou et al. 2006; Lee et al. 2007b; Rusiecki et al. 2009), carbofuran, chlorpyrifos, diazinon, dieldrin, pendimethalin, and lung cancer (significantly increased risk) (Alavanja et al. 2004; Beane Freeman et al. 2005; Bonner et al. 2005; Hou et al. 2006; Lee et al. 2004a; Purdue et al. 2006), dicamba and lung cancer (no association with nonexposed reference group) (Alavanja et al. 2004; Samanic et al. 2006), atrazine, glyphosate, and pancreatic cancer (no association) (Andreotti et al. 2009; De Roos et al. 2005a; Rusiecki et al. 2004), and atrazine, captan, carbofuran, permethrin, and prostate cancer (no association) (Alavanja et al. 2003; Bonner et al. 2005; Greenburg et al. 2008; Rusiecki et al. 2004, 2009). Findings of Lee et al. (2007b) and Purdue et al. (2006)

were somewhat inconsistent with respect to the relationship between chlordane and rectal cancer. Specifically, Purdue et al. (2006) reported an increased RR for applicators with > 9 LDs of chlordane exposure, whereas Lee et al. (2007b) did not observe increased rectal cancer incidence among applicators with > 56 days of exposure. However, in the analysis conducted by Lee et al. (2007b), only two exposed cases were available in the highest exposure group (> 56 LDs), and rectal cancer incidence was increased among applicators with 20–56 days of chlordane exposure. Therefore, both studies provide evidence of an association between chlordane and rectal cancer even though Lee et al. (2007b) did not observe an increased OR in the highest category of exposure. Chemical cohort analysis for metolachlor (Rusiecki et al. 2006) did not confirm the previously observed association with lung cancer (Alavanja et al. 2004). One explanation for this discrepancy may be that the association reported by Alavanja et al. (2004) occurred by chance, as findings were based on fewer exposed cases and were less precise. Alternatively, differences in cutoff points used for the highest exposure groups may explain this inconsistency; a value of 457 LDs was used by Alavanja et al. (2003), whereas a value of 116 LDs was used by Rusiecki et al. (2006). Therefore, it is possible that an increased RR was not reported by Rusiecki et al. (2006) because the highest exposure group included applicators with exposure levels below those likely to result in increased cancer risk. This possibility is supported by the fact that Alavanja et al. (2003) did not observe an association between metolachlor and lung cancer among applicators exposed for 116–457 LDs. Finally, increased risks of colon cancer observed in chemical specific analyses for dicamba (Samanic et al. 2006) and trifluralin (Kang et al. 2008) do not agree with findings reported in cancer-specific analyses (Lee et al. 2007b). However, analyses for dicamba and trifluralin were limited to ever/never exposure classification in the study by Lee et al. (2007b), whereas Samanic et al. (2006) and Kang et al. (2008) reported increased RRs for colon cancer when exposures were analyzed according to intensity-weighted exposure days. Therefore, differences in exposure classification may account for discrepancies observed between these studies.

In total, 19 pesticides were associated with a significantly increased risk of at least one type of cancer. Clear similarities by type (i.e., insecticide, herbicide, or fungicide), chemical structure, or chemical family were not apparent among these pesticides (Table 1). Seven of these 19 pesticides are no longer registered for use in Canada or the United States (chlordane, dieldrin, fonofos, heptachlor, lindane, methyl bromide, and toxaphene), and three additional pesticides, alachlor, aldicarb, and metolachlor, are registered for use in the United States but are not registered in Canada. Of the remaining pesticides currently registered for use in Canada or the United States, statistically significant exposure–response trends were observed for alachlor (all lymphohematopoietic), aldicarb (colon), carbaryl (melanoma), chlorpyrifos (lung, rectal), diazinon (all cancers, all lymphohematopoietic, leukemia, lung), dicamba (colon, lung), EPTC (all cancers, colon, pancreas), imazethapyr (bladder, colon), metolachlor (lung), pendimethalin (lung, pancreas, rectal), permethrin (multiple myeloma), and trifluralin (colon) (Alavanja et al. 2004; Andreotti et al. 2009; Beane Freeman et al. 2005; Hou et al. 2006; Kang et al. 2008; Koutros et al. 2009; Lee et al. 2004a, 2004b, 2007b; Mahajan et al. 2007; Rusiecki et al. 2006, 2009; Samanic et al. 2006; van Bemmel et al. 2008). These pesticides are listed in Table 2 along with animal toxicologic evidence of carcinogenicity noted by the U.S. EPA (2007), Canadian Pest Management Regulatory Agency (PMRA) (Health Canada 2003, 2005, 2007, 2008), and the International Agency for Research on Cancer (IARC 2010). The U.S. EPA classification terms "likely" and "not likely" in Table 2 do not correspond to quantifiable probabilities of carcinogenicity but instead reflect the weight of animal toxicologic evidence for or against such a relationship (i.e., a classification of "likely" does not mean that a given pesticide is a confirmed carcinogen, but only that such an effect is plausible give current toxicologic evidence).

The IARC has not evaluated most pesticides listed in Table 2 (alachlor, chlorpyrifos, diazinon, dicamba, EPTC, imazethapyr, metolachlor, pendimethalin, and trifluralin) and considers the remaining pesticides (aldicarb, carbaryl, and permethrin) not classifiable with respect to human carcinogenicity (group 3). Evidence of carcinogenicity was noted by the U.S. EPA and/or PMRA in animal toxicity studies for alachlor, carbaryl, metolachlor, pendimethalin, permethrin, and trifluralin, thus supporting the biological plausibility of associations observed for these pesticides. The remaining registered pesticides for which exposure–response relationships were observed are not considered carcinogenic by the U.S. EPA or PMRA (aldicarb, chlorpyrifos, diazinon, dicamba, EPTC, and imazethapyr).

Risk estimates were imprecise for most registered pesticides that displayed an exposure–response pattern with at least one type of cancer because of small numbers of exposed cases. Specifically, ≤ 12 cases were available in the highest categories of exposure

Table 2. Evidence of carcinogenicity noted as of March 2009 by the U.S. EPA, PMRA, and IARC for registered pesticides that displayed a significant exposure–response relationship with at least one type of cancer.

Pesticide	Type	Cancer type(s) with exposure–response in the AHS cohort	Organization		
			U.S. EPA	PMRA	IARC
Alachlor (Lee et al. 2004b)	Herbicide	All LH	Likely (high doses) / not likely (low doses)	Not registered in Canada	Not evaluated
Aldicarb (Lee et al. 2007a)	Insecticide	Colon	Group E[a]	Not registered in Canada	Group 3[b]
Carbaryl (Mahajan et al. 2007)	Insecticide	Melanoma	Likely	Under re-evaluation (positive)[c]	Group 3[b]
Chlorpyrifos (Lee et al. 2004a, 2007b)	Insecticide	Lung, rectum	Group E[a]	Negative[d]	Not evaluated
Diazinon (Alavanja et al. 2004; Beane Freeman et al. 2005)	Insecticide	All cancers, all LH, leukemia, lung	Not likely	Negative[d]	Not evaluated
Dicamba (Alavanja et al. 2004; Samanic et al. 2006)	Herbicide	Colon, lung	Not likely	Negative[d]	Not evaluated
EPTC (Andreotti et al. 2009; van Bemmel et al. 2008)	Herbicide	All cancers, colon, leukemia, pancreas	Not likely	Negative[d]	Not evaluated
Imazethapyr (Koutros et al. 2009)	Herbicide	Bladder, colon	Not likely	Under re-evaluation (negative)[d]	Not evaluated
Metolachlor (Alavanja et al. 2004)	Herbicide	Lung	Group C[e]	Not registered in Canada	Not evaluated
Pendimethalin (Alavanja et al. 2004; Andreotti et al. 2009; Hou et al. 2006)	Herbicide	Lung, rectum, pancreas	Group C[e]	Positive[c]	Not evaluated
Permethrin (Rusiecki et al. 2006, 2009)	Insecticide	Myeloma	Likely	Positive[c]	Group 3[b]
Trifluralin (Kang et al. 2008)	Herbicide	Colon	Group C[e]	Positive[c]	Not evaluated

[a]Evidence of noncarcinogenicity in humans. [b]Not classifiable as to carcinogenicity to humans. [c]Evidence of carcinogenicity noted in animal toxicology database. [d]No evidence of carcinogenicity in animal toxicology database. [e]Possible human carcinogen.

for aldicarb (colon), carbaryl (melanoma), chlorpyrifos (rectal), diazinon (all LH, leukemia, lung), dicamba (lung), EPTC (pancreas), imazethapyr (colon, bladder), metolachlor (lung), pendimethalin (lung, pancreas, rectal), and permethrin (multiple myeloma). As a result, little can be concluded at this time regarding the causal nature of these associations, and further analyses are required once more cases have accrued. RRs for chlorpyrifos and lung cancer and EPTC and colon cancer were more precise, each indicating that cancer incidence doubled in the highest exposure groups relative to nonexposed applicators (Lee et al. 2004a; van Bemmel et al. 2008). However, the weight of biological evidence reviewed by the U.S. EPA and PMRA does not suggest that chlorpyrifos and EPTC are carcinogenic (Health Canada 2003, 2008; Smegal 2002; U.S. EPA 1999). One explanation for discrepancies between AHS findings and toxicologic evidence may be that animal toxicity tests typically reflect exposure to a single pesticide active ingredient and not the end-use product or multiple products. Therefore, real-life exposures in the field are not equivalent to what is tested in animal toxicity studies. Alternatively, some AHS findings may have occurred by chance because of the large number of multiple comparisons or bias from uncontrolled confounding or other source of bias. Nevertheless, as many of the studies reviewed are the first to examine the reported associations, the findings are useful for generating hypotheses that require confirmation in future studies. Going forward, adjustment for multiple comparisons will be important to avoid spurious associations. Likewise, bias analysis may be helpful in characterizing overall uncertainty in AHS findings; a recent study by Lash (2007) illustrates how conventional frequentist methods may understate uncertainty in effect measures by quantifying only random error and may result in bias away from the null.

Epidemiologic evidence outside the AHS cohort remains limited with respect to associations observed for specific pesticides and cancer types listed in Table 2. Three studies examined cancer incidence in a cohort of alachlor manufacturing workers in Iowa (Acquavella et al. 1996, 2004; Leet et al. 1996). Lymphohematopoietic tumors were increased in one of these studies relative to expected values in the Iowa population (SIR = 3.6; 95% CI, 1.2–8.5) (Leet et al. 1996); however, this estimate was based on only five exposed cases, and potential confounding factors were not included in the analyses. Two or fewer lymphohematopoietic cancer cases were available in studies conducted by Acquavella et al. (1996, 2004), and in general, none of the three studies of alachlor manufacturing workers provides strong evidence of an important relationship between alachlor and cancer. Studies of carbaryl and melanoma have not been conducted outside the AHS cohort, but Zheng et al. (2001) reported an association between small-cell lymphoma among participants exposed to carbaryl in a pooled analysis of three case–control studies in the United States (Cantor et al. 1992; Hoar et al. 1986; Zahm et al. 1990). However, carbaryl exposure was not associated with NHL in a Canadian case–control study after adjusting for a number of potential confounding factors (McDuffie et al. 2001), and carbaryl was not associated with NHL in the AHS cohort (Mahajan et al. 2007). A case–control study examined lung cancer mortality in Florida pest control workers exposed to chlorpyrifos and diazinon, but neither pesticide was associated with a significantly increased risk of lung cancer mortality (Pesatori et al. 1994). Chlorpyrifos and diazinon were identified as possible risk factors for NHL in a pooled analysis of three case–control studies in Iowa, Minnesota, Kansas, and Nebraska (Waddell et al. 2001); however, these pesticides were not associated with NHL in the AHS cohort (Beane Freeman et al. 2005; Lee et al. 2004a). One previous case–control study observed an association between dicamba and NHL (McDuffie et al. 2001), but dicamba was not associated with NHL in the AHS cohort. Hoar et al. (1986) noted an association between NHL and trifluralin exposure based on only three exposed cases, but a more recent pooled analysis did not observe a significant relationship between trifluralin and NHL (De Roos et al. 2003). Previous studies of chlorpyrifos and rectal cancer and diazinon and leukemia were not identified. Likewise, studies of aldicarb, dicamba, and colon cancer, EPTC and colon or pancreatic cancer, imazethapyr and bladder or colon cancer, metolachlor and lung cancer, pendimethalin and lung, rectal, or pancreatic cancer, permethrin and myeloma, and trifluralin and colon cancer were not identified outside the AHS cohort, and epidemiologic evidence in general is limited for these pesticides.

Exposure assessment is a challenge in large-scale epidemiologic studies, as it is often not possible to obtain quantitative exposure data at etiologically relevant time periods for large numbers of study participants. In the AHS, self-reported LDs and IWLDs of pesticide exposure are used as the primary exposure measures. To date, four studies have examined the validity of the intensity-weighted exposure algorithm used in the AHS (Acquavella et al. 2006; Coble et al. 2005; Thomas et al. 2009). Two of these studies were conducted using data from the Pesticide Exposure Assessment Study conducted in Canada (Acquavella et al. 2006; Coble et al. 2005), and the remaining two studies were conducted using members of the AHS cohort (Hines et al. 2008; Thomas et al. 2009). In general, validation studies have observed low to moderate correlations between exposure intensity algorithm scores and urinary biomarkers of 2,4-D, 4-chloro-2-methylphenoxyacetic acid (MCPA), captan, glyphosate, and chlorpyrifos (Acquavella et al. 2006; Coble et al. 2005; Hines et al. 2008; Thomas et al. 2009). Low to moderate correlations were also reported between algorithm scores and quantitative levels of 2,4-D and chlorpyrifos measured in hand-wipe samples, dermal patches, and personal air samples (Thomas et al. 2009); however, correlations between algorithm scores and quantitative measures of chlorpyrifos exposure varied by application method, with stronger correlations observed for liquid spray applications relative to granular in-furrow applications. For captan, intensity-weighted algorithm scores were predictive of exposure levels measured on dermal patch samples located on the thighs of pesticide applicators but were not significant predictors of captan levels measured in personal air samples, hand rinses, and forearm patches (Hines et al. 2008). Weighted kappa values for categorical agreement between algorithm scores and biomarker levels were low to moderate ($0.07 < \text{kappa} < 0.37$) in validation studies conducted to date, and considerable overlap in urinary biomarker levels was apparent between exposure categories based on algorithm scores (Acquavella et al. 2006; Thomas et al. 2009). However, algorithm scores were able to detect significant trends in urine and hand-wipe concentrations of 2,4-D (Coble et al. 2005; Thomas et al. 2009) and urine concentrations of MCPA (Coble et al. 2005), and in general, biomarker levels tended to be greatest among participants labeled as having the highest exposures (Acquavella et al. 2006; Thomas et al. 2009).

Exposure misclassification undoubtedly had an impact on AHS findings reported to date. As participants reported exposures prior to disease onset, the process of exposure misclassification in the AHS cohort is likely to be nondifferential; however, this does not guarantee bias toward the null in any individual study (Dosemeci et al. 1990; Jurek et al. 2005; 2008; Pearce et al. 2007; Sorahan and Gilthorpe 1994; Thomas 1995). What seems apparent from validation studies is that the exposure intensity algorithm is capable of differentiating subjects with the highest and lowest exposure levels but is less capable of valid exposure classification across an exposure gradient. Unfortunately, this limits the ability to detect exposure–response patterns, as substantial exposure misclassification is expected to occur across categories of exposure. Further validation of the exposure intensity algorithm in biomonitoring studies for an expanded group of pesticides may help to characterize

this uncertainty. Similarly, the applicability of intensity-weighted exposure measures in studies of lung cancer requires further evaluation, as the algorithm weighs dermal exposures most heavily (Dosemeci et al. 2002) and may not offer an improvement over lifetime-exposure days for outcomes related to inhalation exposures. Finally, it is not clear how current exposure levels compare with those during etiologically relevant time periods, as the first years of pesticide use often occurred decades prior to enrollment in the AHS.

Conclusions

We reviewed 28 studies of pesticide exposure and cancer incidence in the AHS cohort. Most of the 32 pesticides examined were not strongly associated with cancer, but increased RRs/ORs and positive exposure–response relationships were observed for 12 pesticides currently registered in Canada and/or the United States. However, RRs and ORs were often imprecise because of small numbers of exposed cases, and further follow-up is required once more cases have accrued. Epidemiologic evidence outside the AHS cohort remains limited with respect to most of the observed associations, but animal toxicity data support the possible carcinogenicity of alachlor, carbaryl, metolachlor, pendimethalin, permethrin, and trifluralin. Although the exposure intensity algorithm developed for the AHS offers an improvement over ever/never exposure classification often employed in environmental health studies, exposure misclassification remains a concern. In particular, analysis of exposure–response trends is limited by expected exposure misclassification across categories of LDs and IWLDs. Further validation of the exposure intensity algorithm for an expanded group of pesticides will help to characterize uncertainty resulting from exposure misclassification. In addition, continued follow-up of the AHS cohort as a whole will help to clarify associations reported to date. In doing so, particular attention should be paid to registered pesticides that displayed evidence of a possible association with cancer.

References

Acquavella JF, Alexander BH, Mandel JS, Burns CJ, Gustin C. 2006. Exposure misclassification in studies of agricultural pesticides: insights from biomonitoring. Epidemiology 17:69–74.

Acquavella JF, Delzell E, Cheng H, Lynch CF, Johnson G. 2004. Mortality and cancer incidence among alachlor manufacturing workers 1968-99. Occup Environ Med 61:680–685.

Acquavella JF, Riordan SG, Anne M, Lynch CF, Collins JJ, Ireland BK, et al. 1996. Evaluation of mortality and cancer incidence among alachlor manufacturing workers. Environ Health Perspect 104:728–733.

Agricultural Health Study. 2009. Agricultural Health Study Homepage. Available: http://aghealth.nci.nih.gov [accessed 1 March 2009].

Alavanja MCR, Dosemeci M, Samanic C, Lubin J, Lynch CF, Knott C, et al. 2004. Pesticides and lung cancer risk in the Agricultural Health Study cohort. Am J Epidemiol 160:876–885.

Alavanja MCR, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, et al. 2003. Use of agricultural pesticides and prostate cancer in the Agricultural Health Study cohort. Am J Epidemiol 157:800–814.

Alavanja MCR, Sandler DP, Lynch CF, Knott C, Lubin JH, Tarone R, et al. 2005. Cancer incidence in the Agricultural Health Study. Scand J Work Environ Health 31(suppl 1):39–45.

Alavanja MCR, Sandler DP, McMaster SB, Zahm SH, McDonnell CJ, Lynch CF, et al. 1996. The Agricultural Health Study. Environ Health Perspect 104:362–369.

Andreotti G, Beane-Freeman L, Hou L, Coble J, Rusiecki J, Hoppin JA, et al. 2009. Agricultural pesticide use and pancreatic cancer risk in the Agricultural Health Study cohort. Int J Cancer 124:2495–2500.

Beane Freeman LE, Bonner MR, Blair A, Hoppin JA, Sandler DP, Lubin JH, et al. 2005. Cancer incidence among male pesticide applicators in the Agricultural Health Study cohort exposed to diazinon. Am J Epidemiol 162:1070–1079.

Beseler C, Stallones L, Hoppin JA, Alavanja MCR, Blair A, Keefe T, et al. 2006. Depression and pesticide exposures in female spouses of licensed pesticide applicators in the Agricultural Health Study cohort. J Occup Environ Med 48:1005–1013.

Beseler CL, Stallones L, Hoppin JA, Alavanja MCR, Blair A, Keefe T, et al. 2008. Depression and pesticide exposure among private pesticide applicators enrolled in the Agricultural Health Study. Environ Health Perspect 116:1713–1719.

Blair A, Sandler D, Tarone R, Lubin J, Thomas K, Hoppin JA, et al. 2005a. Mortality among participants in the Agricultural Health Study. Ann Epidemiol 15:279–285.

Blair A, Sandler D, Thomas K, Hoppin JA, Kamel F, Coble J, et al. 2005b. Disease and injury among participants in the Agricultural Health Study. J Agric Saf Health 11:141–150.

Blair A, Tarone R, Sandler D, Lynch CF, Rowland A, Wintersteen W, et al. 2002. Reliability of reporting on lifestyle and agricultural factors by a sample of participants in the Agricultural Health Study from Iowa. Epidemiology 13:94–99.

Bonner MR, Coble J, Blair A, Beane-Freeman L, Hoppin JA, Sandler DP, et al. 2007. Malathion exposure and the incidence of cancer in the Agricultural Health Study. Am J Epidemiol 166:1023–1034.

Bonner MR, Lee WJ, Sandler DP, Hoppin JA, Dosemeci M, Alavanja MCR. 2005. Occupational exposure to carbofuran and the incidence of cancer in the Agricultural Health Study. Environ Health Perspect 113:285–289.

Cantor KP, Blair A, Everett G, Gibson R, Burmeister LF, Brown LM, et al. 1992. Pesticides and other agricultural risk factors for non-Hodgkin's lymphoma among men in Iowa and Minnesota. Cancer Res 52:2447–2455.

Coble J, Arbuckle T, Lee W, Alavanja M, Dosemeci M. 2005. The validation of a pesticide exposure algorithm using biological monitoring results. J Occup Environ Hyg 2:194–201.

Coble J, Hoppin JA, Engel L, Elci OC, Dosemeci M, Lynch CF, et al. 2002. Prevalence of exposure to solvents, metals, grain dust, and other hazards among farmers in the Agricultural Health Study. J Expo Anal Environ Epidemiol 12:418–426.

Cooper GS, Martin SA, Longnecker MP, Sandler DP, Germolec DR. 2004. Associations between plasma DDE levels and immunologic measures in African-American farmers in North Carolina. Environ Health Perspect 112:1080–1084.

Crawford JM, Hoppin JA, Alavanja MCR, Blair A, Sandler DP, Kamel F. 2008. Hearing loss among licensed pesticide applicators in the Agricultural Health Study. J Occup Environ Med 50:817–826.

De Roos AJ, Blair A, Rusiecki JA, Hoppin JA, Svec M, Dosemeci M, et al. 2005a. Cancer incidence among glyphosate-exposed pesticide applicators in the Agricultural Health Study. Environ Health Perspect 113:49–54.

De Roos AJ, Cooper GS, Alavanja MC, Sandler DP. 2005b. Rheumatoid arthritis among women in the Agricultural Health Study: risk associated with farming activities and exposures. Ann Epidemiol 15:762–770.

De Roos AJ, Zahm SH, Cantor KP, Weisenberg DD, Holmes FF, Burnmeister LF, et al. 2003. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. Occup Environ Med 60:E11; doi: 10.1136/oem.60.9.e11 [Online September 2003].

Dosemeci M, Alavanja MCR, Rowland AS, Mage D, Zahm SH, Rothman N, et al. 2002. A quantitative approach for estimating exposure to pesticides in the Agricultural Health Study. Ann Occup Hyg 46:245–260.

Dosemeci M, Wacholder S, Lubin JH. 1990. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 132:746–748.

Engel LS, Hill DA, Hoppin JA, Lubin JH, Lynch CF, Pierce J, et al. 2005. Pesticide use and breast cancer risk among farmers' wives in the Agricultural Health Study. Am J Epidemiol 161:121–135.

Farr SL, Cai J, Savitz DA, Sandler DP, Hoppin JA, Cooper GS. 2006. Pesticide exposure and the timing of menopause. Am J Epidemiol 163:731–742.

Farr SL, Cooper GS, Cai J, Savitz DA, Sandler DP. 2004. Pesticide use and menstrual cycle characteristics among premenopausal women in the Agricultural Health Study. Am J Epidemiol 160:1194–1204.

Flower KB, Hoppin JA, Lynch CF, Blair A, Knott C, Shore DL, et al. 2004. Cancer risk and parental pesticide application in children of Agricultural Health Study participants. Environ Health Perspect 112:631–635.

Greenburg DL, Rusiecki J, Koutros S, Dosemeci M, Patel R, Hines CJ, et al. 2008. Cancer incidence among pesticide applicators exposed to captan in the Agricultural Health Study. Cancer Causes Control 19:1401–1407.

Health Canada. 2003. Proposed Acceptability for Continuing Registration: Phase 2 of the Re-evaluation of Chlorpyrifos. Catalog No. H113-18/2003-3E-IN. Ottawa, Ontario, Canada:Pest Management Regulatory Agency.

Health Canada. 2005. Re-evaluation Note: Preliminary Risk and Value Assessments of Diazinon. Catalog No. H113-5/2005-6E. Ottawa, Ontario, Canada:Pest Management Regulatory Agency.

Health Canada. 2007. Re-evaluation Decision: Pendimethalin. Catalog No. H113-27/2007-7E. Ottawa, Ontario, Canada:Pest Management Regulatory Agency.

Health Canada. 2008. Re-evaluation Decision: *S*-ethyl dipropylthiocarbamate (EPTC). Catalog No. H113-28/2008-6E. Ottawa, Ontario, Canada:Pest Management Regulatory Agency.

Hines CJ, Deddens JA, Jaycox LB, Andrews RN, Striley CA, Alavanja MC. 2008. Captan exposure and evaluation of a pesticide exposure algorithm among orchard pesticide applicators in the Agricultural Health Study. Ann Occup Hyg 52:153–166.

Hoar SK, Blair A, Holmes FF, Boysen CD, Robel RJ, Hoover R, et al. 1986. Agricultural herbicide use and risk of lymphoma and soft-tissue sarcoma. JAMA 256:1141–1147.

Hoppin JA, Umbach DM, Kullman GJ, Henneberger PK, London SJ, Alavanja MCR, et al. 2007a. Pesticides and other agricultural factors associated with self-reported farmer's lung among farm residents in the Agricultural Health Study. Occup Environ Med 64:334–342.

Hoppin JA, Umbach DM, London SJ, Alavanja MCR, Sandler DP. 2002a. Chemical predictors of wheeze among farmer pesticide applicators in the Agricultural Health Study. Am J Respir Crit Care Med 165:683–689.

Hoppin JA, Umbach DM, London SJ, Henneberger PK, Kullman GJ, Alavanja MCR, et al. 2008. Pesticides and atopic and nonatopic asthma among farm women in the Agricultural Health Study. Am J Respir Crit Care Med 177:11–18.

Hoppin JA, Umbach DM, London SJ, Lynch CF, Alavanja MCR, Sandler DP. 2006a. Pesticides associated with wheeze among commercial pesticide applicators in the Agricultural Health Study. Am J Epidemiol 163:1129–1137.

Hoppin JA, Umbach DM, London SJ, Lynch CF, Alavanja MCR, Sandler DP. 2006b. Pesticides and adult respiratory outcomes in the Agricultural Health Study. Ann NY Acad Sci 1076:343–354.

Hoppin JA, Valcin M, Henneberger PK, Kullman GJ, Umbach DM, London SJ, et al. 2007b. Pesticide use and chronic bronchitis among farmers in the Agricultural Health Study. Am J Ind Med 50:969–979.

Hoppin JA, Yucel F, Dosemeci M, Sandler DP. 2002b. Accuracy of self-reported pesticide use duration information from licensed pesticide applicators in the Agricultural Health Study. J Expo Anal Environ Epidemiol 12:313–318.

Hou L, Lee WJ, Rusiecki J, Hoppin JA, Blair A, Bonner MR, et al. 2006. Pendimethalin exposure and cancer incidence among pesticide applicators. Epidemiology 17:302–307.

IARC (International Agency for Research on Cancer). 2010. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Available: http://www.iarc.fr/en/publications/list/monographs [accessed 1 January 2010].

Jurek AM, Greenland S, Maldonado G. 2008. How far from

non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? Int J Epidemiol 37:382–385.

Jurek AM, Greenland S, Maldonado G, Church TR. 2005. Proper interpretation of non-differential misclassification effects: expectation vs observations. Int J Epidemiol 34:680–687.

Kamel F, Boyes WK, Gladen BC, Rowland AS, Alavanja MCR, Blair A, et al. 2000. Retinal degeneration in licensed pesticide applicators. Am J Ind Med 37:618–628.

Kamel F, Engel LS, Gladen BC, Hoppin JA, Alavanja MCR, Sandler DR. 2005. Neurologic symptoms in licensed private pesticide applicators in the Agricultural Health Study. Environ Health Perspect 113:877–882.

Kamel F, Engel LS, Gladen BC, Hoppin JA, Alavanja MCR, Sandler DR. 2007a. Neurologic symptoms in licensed pesticide applicators in the Agricultural Health Study. Hum Exp Toxicol 26:243–250.

Kamel F, Tanner CM, Umbach DM, Hoppin JA, Alavanja MCR, Blair A, et al. 2007b. Pesticide exposure and self-reported Parkinson's disease in the Agricultural Health Study. Am J Epidemiol 165:364–374.

Kang D, Park SK, Beane-Freeman L, Lynch CF, Knott CE, Sandler DP, et al. 2008. Cancer incidence among pesticide applicators exposed to trifluralin in the Agricultural Health Study. Environ Res 107:271–276.

Kirrane EF, Hoppin JA, Kamel F, Umbach DM, Boyes WK, DeRoos AJ, et al. 2005. Retinal degeneration and other eye disorders in wives of farmer pesticide applicators enrolled in the Agricultural Health Study. Am J Epidemiol 161:1020–1029.

Koutros S, Lynch CF, Ma X, Lee WJ, Hoppin JA, Christensen CH, et al. 2009. Heterocyclic aromatic amine pesticide use and human cancer risk: results from the U.S. Agricultural Health Study. Int J Cancer 124:1206–1212.

Koutros S, Mahajan R, Zheng T, Hoppin JA, Ma X, Lynch CF, et al. 2008. Dichlorvos exposure and human cancer risk: results from the Agricultural Health Study. Cancer Causes Control 19:59–65.

Lash TL. 2007. Bias analysis applied to Agricultural Health Study publications to estimate non-random sources of uncertainty. J Occup Med Toxicol 2:15–23.

Lee WJ, Alavanja MCR, Hoppin JA, Rusiecki JA, Kamel F, Blair A, et al. 2007a. Mortality among pesticide applicators exposed to chlorpyrifos in the Agricultural Health Study. Environ Health Perspect 115:528–534.

Lee WJ, Blair A, Hoppin JA, Lubin JH, Rusiecki JA, Sandler DP, et al. 2004a. Cancer incidence among pesticide applicators exposed to chlorpyrifos in the Agricultural Health Study. J Natl Cancer Inst 96:1781–1791.

Lee WJ, Hoppin JA, Blair A, Lubin JH, Dosemeci M, Sandler DP, et al. 2004b. Cancer incidence among pesticide applicators exposed to alachlor in the Agricultural Health Study. Am J Epidemiol 159:373–380.

Lee WJ, Sandler DP, Blair A, Samanic C, Cross AJ, Alavanja MCR. 2007b. Pesticide use and colorectal cancer risk in the Agricultural Health Study. In J Cancer 121:339–346.

Leet T, Acquavella J, Lynch C, Anne M, Weiss NS, Vaughan T, et al. 1996. Cancer incidence among alachlor manufacturing workers. Am J Ind Med 30:300–306.

Litchfield MH. 1999. Agricultural work related injury and ill-health and the economic cost. Environ Sci Pollut Res 6:175–182.

Lynch SM, Rusiecki JA, Blair A, Dosemeci M, Lubin J, Sandler D, et al. 2006. Cancer incidence among pesticide applicators exposed to cyanazine in the Agricultural Health Study. Environ Health Perspect 114:1248–1252.

Mahajan R, Blair A, Coble J, Lynch CF, Hoppin JA, Sandler DP, et al. 2007. Carbaryl exposure and incident cancer in the Agricultural Health Study. Int J Cancer 121:1799–1805.

Mahajan R, Blair A, Lynch CF, Schroeder P, Hoppin JA, Sandler DP, et al. 2006a. Fonofos exposure and cancer incidence in the Agricultural Health Study. Environ Health Perspect 114:1838–1842.

Mahajan R, Bonner MR, Hoppin JA, Alavanja MCR. 2006b. Phorate exposure and incidence of cancer in the Agricultural Health Study. Environ Health Perspect 114:1205–1209.

Martin SA, Harlow SD, Sowers MF, Longnecker MP, Garabrant D, Shore DL, et al. 2002. DDT metabolite and androgens in African-American farmers. Epidemiology 13:454–458.

McDuffie HH, Pahwa P, McLaughlin JR, Spinelli JJ, Finchman S, Dosman JA, et al. 2001. Non-Hodgkin's lymphoma and specific pesticide exposures in men: cross-Canada study of pesticides and health. Cancer Epidemiol Biomarkers Prev 10:1155–1163.

Montgomery MP, Kamel F, Saldana TM, Alavanja MCR, Sandler DP. 2008. Incident diabetes and pesticide exposure among licensed pesticide applicators: Agricultural Health Study, 1993–2003. Am J Epidemiol 167:1235–1246.

Mozzachio AM, Rusiecki JA, Hoppin JA, Mahajan R, Patel R, Beane-Freeman L, et al. 2008. Chlorothalonil exposure and cancer incidence among pesticide applicator participants in the Agricultural Health Study. Environ Res 108:400–403.

Pearce N, Checkoway H, Kriebel D. 2007. Bias in occupational epidemiology studies. Occup Environ Med 64:562–568.

Pesatori AC, Sontag JM, Lubin JH, Consonni D, Blair A. 1994. Cohort mortality and nested case-control study of lung cancer among structural pest control workers in Florida (United States). Cancer Causes Control 5:310–318.

Popendorf W, Donham KJ. 1991. Agricultural hygiene. In: Patty's Industrial Hygiene and Toxicology, 4th ed, Vol 1, Pt A. New York:John Wiley & Sons, 721–761.

Purdue MP, Hoppin JA, Blair A, Dosemeci M, Alavanja MCR. 2006. Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. Int J Cancer 120:642–649.

Rusiecki JA, De Roos A, Lee WJ, Dosemeci M, Lubin JH, Hoppin JA, et al. 2004. Cancer incidence among pesticide applicators exposed to atrazine in the Agricultural Health Study. J Natl Cancer Inst 96:1375–1382.

Rusiecki JA, Hou L, Lee WJ, Blair A, Dosemeci M, Lubin JH, et al. 2006. Cancer incidence among pesticide applicators exposed to metolachlor in the Agricultural Health Study. Int J Cancer 118:3118–3123.

Rusiecki JA, Patel R, Koutros S, Beane-Freeman L, Landgren O, et al. 2009. Cancer incidence among pesticide applicators exposed to permethrin in the Agricultural Health Study. Environ Health Perspect 117:581–586.

Saldana TM, Basso O, Hoppin JA, Baird DD, Knott C, Blair A, et al. 2007. Pesticide exposure and self-reported gestational diabetes mellitus in the Agricultural Health Study. Diabetes Care 30:529–534.

Samanic C, Rusiecki J, Dosemeci M, Hou L, Hoppin JA, Sandler DP, et al. 2006. Cancer incidence among pesticide applicators exposed to dicamba in the Agricultural Health Study. Environ Health Perspect 114:1521–1526.

Shaver CS, Tong T. 1991. Chemical hazards to agricultural workers. Occup Med State Art Review 6:391–413.

Smegal DC. 2002. Human Health Risk Assessment—Chlorpyrifos. Washington, DC:U.S. Environmental Protection Agency.

Sorahan T, Gilthorpe MS. 1994. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. Occup Environ Med 51:839–840.

Sprince NL, Park H, Zwerling C, Lynch CF, Whitten PA, Thu K, et al. 2002. Risk factors for machinery-related injury among Iowa livestock farmers: a case-control study nested in the Agricultural Health Study. Int J Occup Environ Health 8:332–338.

Sprince NL, Park H, Zwerling C, Lynch CF, Whitten PS, Thu K, et al. 2003a. Risk factors for animal-related injury among Iowa livestock farmers: a case-control study nested in the Agricultural Health Study. J Rural Health 19:165–173.

Sprince NL, Park H, Zwerling C, Whitten PA, Lynch CF, Burmeister LF, et al. 2007. Risk factors for low back injury among farmers in Iowa: a case-control study nested in the Agricultural Health Study. J Occup Environ Hyg 4:10–16.

Sprince NL, Zwerling C, Lynch CF, Whitten PA, Thu K, Gillette PP, et al. 2003b. Risk factors for falls among Iowa farmers: a case-control study nested in the Agricultural Health Study. Am J Ind Med 44:265–272.

Sprince NL, Zwerling C, Lynch CF, Whitten PS, Thu K, Logsden-Sackett N, et al. 2003c. Risk factors for agricultural injury: a case-control analysis of Iowa farmers in the Agricultural Health Study. J Agric Saf Health 9:5–18.

Tarone RE, Alavanja MCR, Zahm SH, Lubin JH, Sandler DP, McMaster SB, et al. 1997. The Agricultural Health Study: factors affecting completion and return of self-administered questionnaires in a large prospective cohort study of pesticide applicators. Am J Ind Med 31:233–242.

Thomas D. 1995. When will nondifferential misclassification of an exposure preserve the direction of trend? Am J Epidemiol 142:782–783.

Thomas KW, Dosemeci M, Coble JB, Hoppin JA, Sheldon LS, Chapa G, et al. 2009. Assessment of a pesticide exposure intensity algorithm in the Agricultural Health Study. J Expo Sci Environ Epidemiol; doi:10.1038/jes.2009.54 [Online 4 November 2009].

U.S. EPA. 1999. Reregistration Eligibility Decision: EPTC. EPA/738/R/99/006. Washington, DC:U.S. Environmental Protection Agency.

U.S. EPA (U.S. Environmental Protection Agency). 2007. List of Chemicals Evaluated for Carcinogenic Potential. Available: http://www.epa.gov/pesticides/carlist [accessed 1 March 2009].

U.S. EPA. 2009. Reregistration Eligibility Decision: Permethrin. EPA/738/R/09/306. Washington, DC:U.S. Environmental Protection Agency.

Valcin M, Henneberger PK, Kullman GJ, Umbach DM, London SJ, Alavanja MCR, et al. 2007. Chronic bronchitis among nonsmoking farm women in the Agricultural Health Study. J Occup Environ Med 49:574–583.

van Bemmel DM, Visvanathan K, Beane-Freeman LE, Coble J, Hoppin J, Alavanja MCR. 2008. *S*-ethyl-*N*,*N*-dipropylthiocarbamate exposure and cancer incidence among male pesticide applicators in the Agricultural Health Study: a prospective cohort. Environ Health Perspect 116:1541–1546.

Waddell BL, Zahm SH, Baris D, Weisenburger DD, Holmes F, Burmeister LF, et al. 2001. Agricultural use of organophosphate pesticides and the risk of non-Hodgkin's lymphoma among male farmers (United States). Cancer Causes Control 12:509–517.

White G, Cessna A. 1989. Occupational hazards of farming. Can Fam Physician 35:2331–2336.

Zahm SH, Weisenburger DD, Babbit PA, Saal RC, Vaught JB, Cantor KP, et al. 1990. A case-control study of non-Hodgkin's lymphoma and the herbicide 2,4-dichlorophenoxyacetic acid in Eastern Nebraska. Epidemiology 1:349–356.

Zheng T, Zahm SH, Cantor KP, Weisenburger DD, Zhang Y, Blair A. 2001. Agricultural exposure to carbamate pesticides and risk of non-Hodgkin lymphoma. J Occup Environ Med 43:641–649.

Differences in the carcinogenic evaluation of glyphosate between the International Agency for Research on Cancer (IARC) and the European Food Safety Authority (EFSA)

Christopher J Portier,[1] Bruce K Armstrong,[2] Bruce C Baguley,[3] Xaver Baur,[4] Igor Belyaev,[5] Robert Bellé,[6] Fiorella Belpoggi,[7] Annibale Biggeri,[8] Maarten C Bosland,[9] Paolo Bruzzi,[10] Lygia Therese Budnik,[11] Merete D Bugge,[12] Kathleen Burns,[13] Gloria M Calaf,[14] David O Carpenter,[15] Hillary M Carpenter,[16] Lizbeth López-Carrillo,[17] Richard Clapp,[18] Pierluigi Cocco,[19] Dario Consonni,[20] Pietro Comba,[21] Elena Craft,[22] Mohamed Aqiel Dalvie,[23] Devra Davis,[24] Paul A Demers,[25] Anneclaire J De Roos,[26] Jamie DeWitt,[27] Francesco Forastiere,[28] Jonathan H Freedman,[29] Lin Fritschi,[30] Caroline Gaus,[31] Julia M Gohlke,[32] Marcel Goldberg,[33] Eberhard Greiser,[34] Johnni Hansen,[35] Lennart Hardell,[36] Michael Hauptmann,[37] Wei Huang,[38] James Huff,[39] Margaret O James,[40] C W Jameson,[41] Andreas Kortenkamp,[42] Annette Kopp-Schneider,[43] Hans Kromhout,[44] Marcelo L Larramendy,[45] Philip J Landrigan,[46] Lawrence H Lash,[47] Dariusz Leszczynski,[48] Charles F Lynch,[49] Corrado Magnani,[50] Daniele Mandrioli,[51] Francis L Martin,[52] Enzo Merler,[53] Paola Michelozzi,[54] Lucia Miligi,[55] Anthony B Miller,[56] Dario Mirabelli,[57] Franklin E Mirer,[58] Saloshni Naidoo,[59] Melissa J Perry,[60] Maria Grazia Petronio,[61] Roberta Pirastu,[62] Ralph J Portier,[63] Kenneth S Ramos,[64] Larry W Robertson,[65] Theresa Rodriguez,[66] Martin Röösli,[67] Matt K Ross,[68] Deodutta Roy,[69] Ivan Rusyn,[70] Paulo Saldiva,[71] Jennifer Sass,[72] Kai Savolainen,[73] Paul T J Scheepers,[74] Consolato Sergi,[75] Ellen K Silbergeld,[76] Martyn T Smith,[77] Bernard W Stewart,[78] Patrice Sutton,[79] Fabio Tateo,[80] Benedetto Terracini,[81] Heinz W Thielmann,[82] David B Thomas,[83] Harri Vainio,[84] John E Vena,[85] Paolo Vineis,[86] Elisabete Weiderpass,[87] Dennis D Weisenburger,[88] Tracey J Woodruff,[89] Takashi Yorifuji,[90] Il Je Yu,[91] Paola Zambon,[92] Hajo Zeeb,[93] Shu-Feng Zhou[94]

The International Agency for Research on Cancer (IARC) Monographs Programme identifies chemicals, drugs, mixtures, occupational exposures, lifestyles and personal habits, and physical and biological agents that cause cancer in humans and has evaluated about 1000 agents since 1971. Monographs are written by ad hoc Working Groups (WGs) of international scientific experts over a period of about 12 months ending in an eight-day meeting. The WG evaluates all of the publicly available scientific information on each substance and, through a transparent and rigorous process,[1] decides on the degree to which the scientific evidence supports that substance's potential to cause or not cause cancer in humans.

For Monograph 112,[2] 17 expert scientists evaluated the carcinogenic hazard for four insecticides and the herbicide glyphosate.[3] The WG concluded that the data for glyphosate meet the criteria for classification as a *probable human carcinogen*.

The European Food Safety Authority (EFSA) is the primary agency of the European Union for risk assessments regarding food safety. In October 2015, EFSA reported[4] on their evaluation of the Renewal Assessment Report[5] (RAR) for glyphosate that was prepared by the Rapporteur Member State, the German Federal Institute for Risk Assessment (BfR). EFSA concluded that 'glyphosate is unlikely to pose a carcinogenic hazard to humans and the evidence does not support classification with regard to its carcinogenic potential'. Addendum 1 (the BfR Addendum) of the RAR[5] discusses the scientific rationale for differing from the IARC WG conclusion.

Serious flaws in the scientific evaluation in the RAR incorrectly characterise the potential for a carcinogenic hazard from exposure to glyphosate. Since the RAR is the basis for the European Food Safety Agency (EFSA) conclusion,[4] it is critical that these shortcomings are corrected.

THE HUMAN EVIDENCE

EFSA concluded 'that there is very limited evidence for an association between glyphosate-based formulations and non-Hodgkin lymphoma (NHL), overall inconclusive for a causal or clear associative relationship between glyphosate and cancer in human studies'. The BfR Addendum (p. ii) to the EFSA report explains that 'no consistent positive association was observed' and 'the most powerful study showed no effect'. The IARC WG concluded there is *limited evidence of carcinogenicity in humans* which means "A positive association has been observed between exposure to the agent and cancer for which a causal interpretation is considered by the Working Group to be credible, but chance, bias or confounding could not be ruled out with reasonable confidence."[1]

The finding of *limited evidence* by the IARC WG was for NHL, based on high-quality case–control studies, which are particularly valuable for determining the carcinogenicity of an agent because their design facilitates exposure assessment and reduces the potential for certain biases. The Agricultural Health Study[6] (AHS) was the only cohort study available providing information on the carcinogenicity

For numbered affiliations see end of article.

Correspondence to Dr Christopher J Portier, Environmental Health Consultant, Thun, CH-3600, Switzerland; cportier@me.com

of glyphosate. The study had a null finding for NHL (RR 1.1, 0.7–1.9) with no apparent exposure–response relationship in the results. Despite potential advantages of cohort versus case–control studies, the AHS had only 92 NHL cases in the unadjusted analysis as compared to 650 cases in a pooled case–control analysis from the USA.[7] In addition, the median follow-up time in the AHS was 6.7 years, which is unlikely to be long enough to account for cancer latency.[8]

The RAR classified all of the case–control studies as 'not reliable,' because, for example, information on glyphosate exposure, smoking status and/or previous diseases had not been assessed. In most cases, this is contrary to what is actually described in the publications. Well-designed case–control studies are recognised as strong evidence and routinely relied on for hazard evaluations.[9 10] The IARC WG carefully and thoroughly evaluated all available epidemiology data, considering the strengths and weaknesses of each study. This is key to determining that the positive associations seen in the case–control studies are a reliable indication of an association and not simply due to chance or methodological flaws. To provide a reasonable interpretation of the findings, an evaluation needs to properly weight studies according to quality rather than simply count the number of positives and negatives. The two meta-analyses cited in the IARC Monograph[11] are excellent examples of objective evaluations and show a consistent positive association between glyphosate and NHL.

The final conclusion[5] (Addendum 1, p.21) that "there was no unequivocal evidence for a clear and strong association of NHL with glyphosate" is misleading. IARC, like many other groups, uses three levels of evidence for human cancer data.[1] *Sufficient evidence* means 'that a causal relationship has been established' between glyphosate and NHL. BfR's conclusion is equivalent to deciding that there is not *sufficient evidence*. Legitimate public health concerns arise when 'causality is credible', that is, when there is *limited evidence of carcinogenicity.*

EVIDENCE FROM ANIMAL CARCINOGENICITY STUDIES

EFSA concluded 'No evidence of carcinogenicity was confirmed by the majority of the experts (with the exception of one minority view) in either rats or mice due to a lack of statistical significance in pairwise comparison tests, lack of consistency in multiple animal studies and slightly increased incidences only at dose levels at or above the limit dose/maximum tolerated dose (MTD), lack of preneoplastic lesions and/or being within historical control range'. The IARC WG review found a significant positive trend for renal tumours in male CD-1 mice,[12] a rare tumour, although no comparisons of any individual exposure group to the control group were statistically significant. The WG also identified a significant positive trend for hemangiosarcoma in male CD-1 mice,[13] again with no individual exposure group significantly different from controls. Finally, the WG also saw a significant increase in the incidence of pancreatic islet cell adenomas in two studies in male Sprague-Dawley rats.[14–16] In one of these rat studies, thyroid gland adenomas in females and liver adenomas in males were also increased. By the IARC review criteria,[1] this constitutes *sufficient evidence* in animals.

The IARC WG reached this conclusion using data that were publicly available in sufficient detail for independent scientific evaluation (a requirement of the IARC Preamble[1]). On the basis of the BfR Addendum, it seems there were three additional mouse studies and two additional rat studies that were unpublished and available to EFSA. Two of the additional studies were reported to have a significant trend for renal tumours, one in CD-1 mice (Sugimoto. *18-Month Oral Oncogenicity Study in Mice*. Unpublished, designated ASB2012–11493 in RAR. 1997), and one in Swiss-Webster mice (Unknown. *A chronic feeding study of glyphosate (roundup technical) in mice*. Unpublished, designated ABS2012–11491 in RAR. 2001). One of these studies (Sugimoto. Unpublished, 1997) also reported a significant trend for hemangiosarcoma. The RAR also reported two studies in CD-1 mice showing significant trends for malignant lymphoma (Sugimoto. Unpublished, 1997; Unknown. *Glyphosate Technical: Dietary Carcinogencity Study in the Mouse*. Unpublished, designated ABS2012–11492 in RAR. 2009).

The RAR dismissed the observed trends in tumour incidence because there are no individual treatment groups that are significantly different from controls and because the maximum observed response is reportedly within the range of the historical control data (Table 5.3–1, p.90). Care must be taken in using historical control data to evaluate animal carcinogenicity data. In virtually all guidelines,[1 17 18] scientific reports[19] and publications[20–23] on this issue, the recommended first choice is the use of concurrent controls and trend tests, even in the EC regulations cited in the RAR[18] (see p.375). Trend tests are more powerful than pairwise comparisons, particularly for rare tumours where data are sparse. Historical control data should be from studies in the same time frame, for the same animal strain, preferably from the same laboratory or the same supplier and preferably reviewed by the same pathologist.[17 18] While the EFSA final peer review[4] mentions the use of historical control data from the original laboratory, no specifics are provided and the only referenced historical control data[24] are in the BfR addendum.[5] One of the mouse studies[12] was clearly done before this historical control database was developed, one study (Sugimoto. Unpublished, 1997) used Crj:CD-1 mice rather than Crl:CD-1 mice, and one study[13] did not specify the substrain and was reported in 1993 (probably started prior to 1988). Hence, only a single study (Unknown. Unpublished, 2009) used the same mouse strain as the cited historical controls, but was reported more than 10 years after the historical control data set was developed.

The RAR dismissed the slightly increased tumour incidences in the studies considered because they occurred "only at dose levels at or above the limit dose/maximum tolerated dose (MTD)", and because there was a lack of preneoplastic lesions. Exceeding the MTD is demonstrated by an increase in mortality or other serious toxicological findings at the highest dose, not by a slight reduction in body weight. No serious toxicological findings were reported at the highest doses for the mouse studies in the RAR. While some would argue that these high doses could cause cellular disruption (eg, regenerative hyperplasia) leading to cancer, no evidence of this was reported in any study. Finally, a lack of preneoplastic lesions for a significant neoplastic finding is insufficient reason to discard the finding.

MECHANISTIC INFORMATION

The BfR Addendum dismisses the IARC WG finding that 'there is strong evidence that glyphosate causes genotoxicity' by suggesting that unpublished evidence not seen by the IARC WG was overwhelmingly negative and that, since the reviewed studies were not done under guideline principles, they should get less weight. To maintain transparency, IARC reviews only publicly available data. The use of confidential data submitted to the BfR makes it impossible for any scientist not associated with BfR to review this conclusion. Further weakening their interpretation,

the BfR did not include evidence of chromosomal damage from exposed humans or human cells that were highlighted in Tables 4.1 and 4.2 of the IARC Monograph [3]

The BfR confirms (p.79) that the studies evaluated by the IARC WG on oxidative stress were predominantly positive but does not agree that this is strong support for an oxidative stress mechanism. They minimise the significance of these findings predominantly because of a lack of positive controls in some studies and because many of the studies used glyphosate formulations and not pure glyphosate. In contrast, the WG concluded that (p.77) 'Strong evidence exists that glyphosate, AMPA and glyphosate-based formulations can induce oxidative stress'. From a scientific perspective, these types of mechanistic studies play a key role in distinguishing between the effects of mixtures, pure substances and metabolites.

Finally, we strongly disagree that data from studies published in the peer-reviewed literature should automatically receive less weight than guideline studies. Compliance with guidelines and Good Laboratory Practice does not guarantee validity and relevance of the study design, statistical rigour and attention to sources of bias.[25 26] The majority of research after the initial marketing approval, including epidemiology studies, will be conducted in research laboratories using various models to address specific issues related to toxicity, often with no testing guidelines available. Peer-reviewed and published findings have great value in understanding mechanisms of carcinogenicity and should be given appropriate weight in an evaluation based on study quality, not just on compliance with guideline rules.

GENERAL COMMENTS

Science moves forward on careful evaluations of data and a rigorous review of findings, interpretations and conclusions. An important aspect of this process is transparency and the ability to question or debate the findings of others. This ensures the validity of the results and provides a strong basis for decisions. Many of the elements of transparency do not exist for the RAR.[5] For example, citations for almost all references, even those from the open scientific literature, have been redacted. The ability to objectively evaluate the findings of a scientific report requires a complete list of cited supporting evidence. As another example, there are no authors or contributors listed for either document, a requirement for publication in virtually all scientific journals where financial support, conflicts of interest and affiliations of authors are fully disclosed. This is in direct contrast to the IARC WG evaluation listing all authors, all publications and public disclosure of pertinent conflicts of interest prior to the WG meeting.[27]

Several guidelines have been devised for conducting careful evaluation and analysis of carcinogenicity data, most after consultation with scientists from around the world. Two of the most widely used guidelines in Europe are the OECD guidance on the conduct and design of chronic toxicity and carcinogenicity studies[17] and the European Chemicals Agency Guidance on Commission Regulation (EU) No 286/2011;[18] both are cited in the RAR. The methods used for historical controls and trend analysis are inconsistent with these guidelines.

Owing to the potential public health impact of glyphosate, which is an extensively used pesticide, it is essential that all scientific evidence relating to its possible carcinogenicity is publicly accessible and reviewed transparently in accordance with established scientific criteria.

SUMMARY

The IARC WG concluded that glyphosate is a 'probable human carcinogen', putting it into IARC category 2A due to *sufficient evidence* of carcinogenicity in animals, *limited evidence* of carcinogenicity in humans and *strong* evidence for two carcinogenic mechanisms.

- The IARC WG found an association between NHL and glyphosate based on the available human evidence.
- The IARC WG found significant carcinogenic effects in laboratory animals for rare kidney tumours and hemangiosarcoma in two mouse studies and benign tumours in two rat studies.
- The IARC WG concluded that there was strong evidence of genotoxicity and oxidative stress for glyphosate, entirely from publicly available research, including findings of DNA damage in the peripheral blood of exposed humans.

The RAR concluded[5] (Vol. 1, p.160) that 'classification and labelling for carcinogenesis is not warranted' and 'glyphosate is devoid of genotoxic potential'.

- EFSA[4] classified the human evidence as 'very limited' and then dismissed any association of glyphosate with cancer without clear explanation or justification.
- Ignoring established guidelines cited in their report, EFSA dismissed evidence of renal tumours in three mouse studies, hemangiosarcoma in two mouse studies and malignant lymphoma in two mouse studies. Thus, EFSA incorrectly discarded all findings of glyphosate-induced cancer in animals as chance occurrences.
- EFSA ignored important laboratory and human mechanistic evidence of genotoxicity.
- EFSA confirmed that glyphosate induces oxidative stress but then, having dismissed all other findings of possible carcinogenicity, dismissed this finding on the grounds that oxidative stress alone is not sufficient for carcinogen labelling.

The most appropriate and scientifically based evaluation of the cancers reported in humans and laboratory animals as well as supportive mechanistic data is that glyphosate is a *probable human carcinogen*. On the basis of this conclusion and in the absence of evidence to the contrary, it is reasonable to conclude that glyphosate formulations should also be considered likely human carcinogens. The CLP Criteria[18] (Table 3.6.1, p.371) allow for a similar classification of Category 1B when there are 'studies showing limited evidence of carcinogenicity in humans together with limited evidence of carcinogenicity in experimental animals'.

In the RAR, almost no weight is given to studies from the published literature and there is an over-reliance on non-publicly available industry-provided studies using a limited set of assays that define the minimum data necessary for the marketing of a pesticide. The IARC WG evaluation of *probably carcinogenic to humans* accurately reflects the results of published scientific literature on glyphosate and, on the face of it, unpublished studies to which EFSA refers.

Most of the authors of this commentary previously expressed their concerns to EFSA and others regarding their review of glyphosate[28] to which EFSA has published a reply.[29] This commentary responds to the EFSA reply.

The views expressed in this editorial are the opinion of the authors and do not imply an endorsement or support for these opinions by any organisations to which they are affiliated.

Author affiliations
[1]Environmental Health Consultant, Thun, Switzerland
[2]The University of Sydney, Sydney, New South Wales, Australia
[3]The University of Auckland, Auckland, New Zealand
[4]Charité University Medicine Berlin, Berlin, Germany
[5]Cancer Research Institute, Bratislava, Slovak Republic
[6]Sorbonne Universités, UPMC Univ Paris 06, UMR8227, Roscoff, France

[7]CesareMaltoni Cancer Research Center, Bentivoglio (Bologna), Italy
[8]Institute for Cancer Prevention and Research, University of Florence, Italy
[9]University of Illinois at Chicago, Chicago, Illinois, USA
[10]National Cancer Research Institute, San Martino—IST Hospital, Genoa, Italy
[11]University of Hamburg, Hamburg, Germany
[12]STAMI, National Institute of Occupational Health, Oslo, Norway
[13]Sciencecorps, Lexington, Massachusetts, USA
[14]Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile
[15]Institute for Health and the Environment, University at Albany, Rensselaer, New York, USA
[16]Toxicologist, Maplewood, Minnesota, USA
[17]National Institute of Public Health, Cuernavaca, Morelos, Mexico
[18]Boston University School of Public Health, Boston, Massachusetts, USA
[19]Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Cagliari, Italy
[20]Department of Preventive Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
[21]Department of Environment and Primary Prevention, IstitutoSuperiore di Sanità, Rome, Italy
[22]Environmental Defense Fund, Austin, Texas, USA
[23]Center for Environmental and Occupational Health, University of Cape Town, Cape Town, South Africa
[24]Environmental Health Trust, Jackson Hole, Wyoming, USA and The Hebrew University Hadassah School of Medicine, Jerusalem, Israel.
[25]Dalla Lana School of Public Health, University of Toronto, Canada
[26]Department of Environmental and Occupational Health, Drexel University, Philadelphia, Pennsylvania, USA
[27]Brody School of Medicine, East Carolina University, Greenville, North Carolina, USA
[28]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy
[29]University of Louisville School of Medicine, Louisville, Kentucky, USA
[30]School of Public Health, Curtin University, Perth, Western Australia, Australia
[31]Department of Environmental Toxicology, The University of Queensland, Brisbane, Australia
[32]Department of Population Health Sciences, Virginia Tech, Blacksburg, Virginia, USA
[33]Paris Descartes University, France
[34]Epi.Consult GmbH, Musweiler, Germany
[35]Danish Cancer Society Research Center, Copenhagen, Denmark
[36]University Hospital, Orebra, Sweden
[37]Biostatistics Branch, Netherlands Cancer Institute, Amsterdam, The Netherlands
[38]Faculty of Department of Occupational and Environmental Health, Peking Univ School of Public Health, Beijing, China
[39]National Institute for Environmental Health Sciences, Research Triangle Park, North Carolina, USA
[40]University of Florida, Gainesville, Florida, USA
[41]CWJ Consulting, LLC, Cape Coral, Florida, USA
[42]Institute of Environment, Health and Societies, Brunel University London, London, UK
[43]Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany
[44]Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
[45]National Council of Scientific and Technological Research, National University of La Plata, Argentina
[46]Arnhold Institute for Global Health, Icahn School of Medicine at Mount Sinai, New York, USA
[47]Department of Pharmacology, Wayne State University School of Medicine, Detroit, Michigan, USA
[48]Department of Biosciences, University of Helsinki, Helsinki, Finland
[49]Department of Epidemiology, University of Iowa, Iowa City, Iowa, USA
[50]Cancer Epidemiology Unit, University of Eastern Piedmont, Novara, Italy
[51]Cesare Maltoni Cancer Research Center, Bentivoglio (Bologna), Italy
[52]Centre for Biophotonics, Lancaster University, UK
[53]Department of Prevention, Occupational Health Unit, National Health Service, Padua, Italy
[54]Department of Epidemiology Lazio Region, Rome, Italy
[55]Occupational and Environmental Epidemiology Unit, ISPO-Cancer Prevention and Research Institute, Florence, Italy
[56]Dalla Lana School of Public Health, University of Toronto, Canada
[57]Unit of Cancer Epidemiology, University of Turin and CPO-Piemonte, Torino, Italy
[58]Department of Environmental and Occupational Health Sciences, City University of New York School of Public Health, USA
[59]School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa
[60]Department of Environmental and Occupational Health, Milken Institute School of Public Health, The George Washington University, Washington DC, USA
[61]Health and Environment-Department of Prevention, Local Health Authority-Empoli, Florence, Italy
[62]Department of Biology and Biotechnology "Charles Darwin", Sapienza Rome University, Italy
[63]Department of Environmental Sciences, School of the Coast & Environment, Louisiana State University, Baton Rouge, Los Angeles, USA
[64]Center for Applied Genetics and Genomic Medicine, University of Arizona Health Sciences, Tucson, Arizona, USA
[65]Iowa Superfund Research Program and the Interdisciplinary Graduate Program in Human Toxicology, University of Iowa, Iowa City, Iowa, USA
[66]Center for Research in Health, Work and Environment (CISTA), National Autonomous University of Nicaragua (UNAN-León), León, Nicaragua
[67]Swiss Tropical and Public Health Institute, Associated Institute of the University of Basel, Basel, Switzerland
[68]College of Veterinary Medicine, Mississippi State University, Mississippi State, USA
[69]Department of Environmental and Occupational Health, Florida International University, Miami, Florida, USA
[70]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas, USA
[71]Faculty of Medicine, University of São Paulo, São Paulo, Brazil
[72]Natural Resources Defense Council and George Washington University, Washington DC, USA
[73]Nanosafety Research Centre, Finnish Institute of Occupational Health, Helsinki, Finland
[74]Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
[75]Department of Pathology, University of Alberta, Edmonton, Alberta, Canada
[76]Department of Environmental Health Sciences, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
[77]School of Public Health, University of California, Berkeley, California, USA
[78]Faculty of Medicine, University of New South Wales, Randwick, New South Wales Australia
[79]Program on Reproductive Health and the Environment, University of California, San Francisco, California, USA
[80]Istituto di Geosceinze e Georisorse (CNR), Padova, Italy
[81]University of Torino, Torino, Italy
[82]German Cancer Research Center, Heidelberg and Faculty of Pharmacy, Heidelberg University, Germany
[83]Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington, USA
[84]Faculty of Public Health, Kuwait University, Kuwait City, Kuwait
[85]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA
[86]Department of Environmental Epidemiology, Imperial College London, London, UK
[87]Department of Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway; Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway; Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; and Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland.
[88]Department of Pathology, City of Hope Medical Center, Duarte, California, USA
[89]Program on Reproductive Health and the Environment, University of California, San Francisco, USA
[90]Okayama University, Okayama, Japan
[91]Institute of Nanoproduct Safety Research, Hoseo University, Asan, The Republic of Korea
[92]University of Padua, Padova, Italy
[93]Department of Prevention and Evaluation, Leibniz-Institute for Prevention Research and Epidemiology, Bremen, Germany
[94]College of Pharmacy, University of South Florida, Tampa, Florida, USA

Contributors All authors to this commentary have participated in its development and approve the content. MCB, FF, LF, CWJ, HK, TR, MKR, IR and CS were all participants in the IARC WG. CJP was an Invited Specialist in the IARC WG. Many of the remaining authors have also served on IARC Working Groups that did not pertain to glyphosate.

Competing interests CJP, MTS and DDW are providing advice to a US law firm involved in glyphosate litigation. CJP also works part-time for the Environmental Defense Fund on issues not related to pesticides.

Provenance and peer review Commissioned; externally peer reviewed.

OPEN ACCESS

To cite Portier CJ, Armstrong BK, Baguley BC, *et al*. *J Epidemiol Community Health* Published Online First: [*please include* Day Month Year] doi:10.1136/jech-2015-207005

J Epidemiol Community Health 2016;**0**:1–5. doi:10.1136/jech-2015-207005

REFERENCES

1 IARC. *Preamble to the IARC Monographs*. 2006. http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf

2 Guyton KZ, Loomis D, Grosse Y, *et al*. Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. *Lancet Oncol* 2015;16:490–1.
3 IARC Working Group. Glyphosate. In: *Some organophosphate insecticides and herbicides: diazinon, glyphosate, malathion, parathion, and tetrachlorvinphos*. Vol 112. IARC Monogr Prog, 2015:1–92.
4 European Food Safety Authority. Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate. *EFSA J* 2015;13:4302.
5 European Food Safety Authority. *Final Addendum to the Renewal Assessment Report*. 2015. http://registerofquestions.efsa.europa.eu/roqFrontend/outputLoader?output=ON-4302
6 De Roos AJ, Blair A, Rusiecki JA, *et al*. Cancer incidence among glyphosate-exposed pesticide applicators in the Agricultural Health Study. *Environ Health Perspect* 2005;113:49–54.
7 De Roos AJ, Zahm SH, Cantor KP, *et al*. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. *Occup Environ Med* 2003;60:E11.
8 Weisenburger DD. Pathological classification of non-Hodgkin's lymphoma for epidemiological studies. *Cancer Res* 1992;52(19 Suppl):5456s–62s; discussion 5462s-5464s.
9 Checkoway H, Pearce N, Kriebel D. Research methods in occupational epidemiology. 2nd edn. *Monographs in epidemiology and biostatistics*. New York: Oxford University Press, 2004:xiv, 372 p.
10 Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd edn. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008:x, 758 p.
11 Schinasi L, Leon ME. Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. *Int J Environ Res Public Health* 2014;11:4449–527.
12 EPA. *Glyphosate; EPA Reg. # 524–308; mouse oncogenicity study*. 1985. V. Accession No. 251007–014. Tox review 004370. p.
13 JMPR. FAO plant production and protection paper, 178, 2004—pesticide residues in food—2004. *Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and WHO the Core Assessment Group*; Vol 178. Geneva, Switzerland: World Health Organization/Food and Agriculture Organization, 2004:95–169.
14 EPA. *Second peer review of Glyphosate*. 19911–19.
15 EPA. *Glyphosate—EPA Registration No. 524–308–2-Year Chronic Feeding/Oncogenicity Study in Rats with Technical Glyphosate*, I. William Dykstra. Toxicology Branch, Editor. 1991. V. MRID 416438–01 Tox review 008897. p.
16 EPA. *Glyphosate; 2-Year Combined Chronic Toxicity/Carcinogenicity Study in Sprague-Dawley Rats—List A Pesticide for Reregistration*, B. William Dykstra. Toxicology, Editor. 1991. V. MRID 416438–01. Tox review 008390. 1–29.
17 OECD. *Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies*, H.a.S.P. Environment, Editor. Paris: OECD, 2012.
18 European Chemicals Agency. *Guidance on the Application of the CLP Criteria: Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures*. Helsinki, Finland: European Chemicals Agency, 2015.
19 NRC Committee to Review the Styrene Assessment in The National Toxicology Program 12th Report on Carcinogens. *Review of the Styrene Assessment in the National Toxicology Program 12th Report on Carcinogens: Workshop Summary*; Washington DC: National Academies Press, 2014.
20 Keenan C, Elmore S, Francke-Carroll S, *et al*. Best practices for use of historical control data of proliferative rodent lesions. *Toxicol Pathol* 2009;37:679–93.
21 Haseman JK, Boorman GA, Huff J. Value of historical control data and other issues related to the evaluation of long-term rodent carcinogenicity studies. *Toxicol Pathol* 1997;25:524–7.
22 Greim H, Gelbke HP, Reuter U, *et al*. Evaluation of historical control data in carcinogenicity studies. *Hum Exp Toxicol* 2003;22:541–9.
23 Haseman JK, Huff J, Boorman GA. Use of historical control data in carcinogenicity studies in rodents. *Toxicol Pathol* 1984;12:126–35.
24 Giknis M, Clifford C. *Spontaneous Neoplastic Lesions in the Crl:CD-1(ICR)BR Mouse*. Charles River Laboratories; 2000.
25 Myers JP, vom Saal FS, Akingbemi BT, *et al*. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. *Environ Health Perspect* 2009;117:309–15.
26 Buonsante VA, Muilerman H, Santos T, *et al*. Risk assessment's insensitive toxicity testing may cause it to fail. *Environ Res* 2014;135:139–47.
27 IARC Monograph 112. *List of Working Group Participants*. IARC Monogr Eval Carcinog Risks Hum 2015 26 March, 2015 [cited 2015 24 November]. http://monographs.iarc.fr/ENG/Meetings/vol112-participants.pdf
28 Portier CJ, *et al*. *Open Letter: Review of the Carcinogenicity of Glyphosate by EFSA and BfR*. 2015 [cited 2016 1/18/2016]. http://www.efsa.europa.eu/sites/default/files/Prof_Portier_letter.pdf
29 Url B. Response to Open Letter: Review of the Carcinogenicity of Glyphosate by EFSA and BfR. 2016. 1/13/2016. http://www.efsa.europa.eu/sites/default/files/EFSA_response_Prof_Portier.pdf

Differences in the carcinogenic evaluation of glyphosate between the International Agency for Research on Cancer (IARC) and the European Food Safety Authority (EFSA)

Christopher J Portier, Bruce K Armstrong, Bruce C Baguley, Xaver Baur, Igor Belyaev, Robert Bellé, Fiorella Belpoggi, Annibale Biggeri, Maarten C Bosland, Paolo Bruzzi, Lygia Therese Budnik, Merete D Bugge, Kathleen Burns, Gloria M Calaf, David O Carpenter, Hillary M Carpenter, Lizbeth López-Carrillo, Richard Clapp, Pierluigi Cocco, Dario Consonni, Pietro Comba, Elena Craft, Mohamed Aqiel Dalvie, Devra Davis, Paul A Demers, Anneclaire J De Roos, Jamie DeWitt, Francesco Forastiere, Jonathan H Freedman, Lin Fritschi, Caroline Gaus, Julia M Gohlke, Marcel Goldberg, Eberhard Greiser, Johnni Hansen, Lennart Hardell, Michael Hauptmann, Wei Huang, James Huff, Margaret O James, C W Jameson, Andreas Kortenkamp, Annette Kopp-Schneider, Hans Kromhout, Marcelo L Larramendy, Philip J Landrigan, Lawrence H Lash, Dariusz Leszczynski, Charles F Lynch, Corrado Magnani, Daniele Mandrioli, Francis L Martin, Enzo Merler, Paola Michelozzi, Lucia Miligi, Anthony B Miller, Dario Mirabelli, Franklin E Mirer, Saloshni Naidoo, Melissa J Perry, Maria Grazia Petronio, Roberta Pirastu, Ralph J Portier, Kenneth S Ramos, Larry W Robertson, Theresa Rodriguez, Martin Röösli, Matt K Ross, Deodutta Roy, Ivan Rusyn, Paulo Saldiva, Jennifer Sass, Kai Savolainen, Paul T J Scheepers, Consolato Sergi, Ellen K Silbergeld, Martyn T Smith, Bernard W Stewart, Patrice Sutton, Fabio Tateo, Benedetto Terracini, Heinz W Thielmann, David B Thomas, Harri Vainio, John E Vena, Paolo Vineis, Elisabete Weiderpass, Dennis D Weisenburger, Tracey J Woodruff, Takashi Yorifuji, Il Je Yu, Paola Zambon, Hajo Zeeb and Shu-Feng Zhou

J Epidemiol Community Health published online March 3, 2016

Topic Collections

Articles on similar topics can be found in the following collections

Open access (304)

Notes

A Section 508–conformant HTML version of this article is available at http://dx.doi.org/10.1289/ehp.1409149.

IARC Monographs: 40 Years of Evaluating Carcinogenic Hazards to Humans

Neil Pearce,[1] Aaron Blair,[2] Paolo Vineis,[3] Wolfgang Ahrens,[4] Aage Andersen,[5] Josep M. Anto,[6] Bruce K. Armstrong,[7] Andrea A. Baccarelli,[8] Frederick A. Beland,[9] Amy Berrington,[2] Pier Alberto Bertazzi,[10] Linda S. Birnbaum,[11] Ross C. Brownson,[12] John R. Bucher,[13] Kenneth P. Cantor,[14] Elisabeth Cardis,[15] John W. Cherrie,[16] David C. Christiani,[8] Pierluigi Cocco,[17] David Coggon,[18] Pietro Comba,[19] Paul A. Demers,[20] John M. Dement,[21] Jeroen Douwes,[22] Ellen A. Eisen,[23] Lawrence S. Engel,[24] Richard A. Fenske,[25] Lora E. Fleming,[26] Tony Fletcher,[27] Elizabeth Fontham,[28] Francesco Forastiere,[29] Rainer Frentzel-Beyme,[30] Lin Fritschi,[31] Michel Gerin,[32] Marcel Goldberg,[33] Philippe Grandjean,[34] Tom K. Grimsrud,[5] Per Gustavsson,[35] Andy Haines,[27] Patricia Hartge,[2] Johnni Hansen,[36] Michael Hauptmann,[37] Dick Heederik,[38] Kari Hemminki,[39] Denis Hemon,[40] Irva Hertz-Picciotto,[41] Jane A. Hoppin,[42] James Huff,[43] Bengt Jarvholm,[44] Daehee Kang,[45] Margaret R. Karagas,[46] Kristina Kjaerheim,[5] Helge Kjuus,[47] Manolis Kogevinas,[48] David Kriebel,[49] Petter Kristensen,[47] Hans Kromhout,[38] Francine Laden,[8] Pierre Lebailly,[50] Grace LeMasters,[51] Jay H. Lubin,[2] Charles F. Lynch,[52] Elsebeth Lynge,[53] Andrea 't Mannetje,[22] Anthony J. McMichael,[54] John R. McLaughlin,[55] Loraine Marrett,[56] Marco Martuzzi,[57] James A. Merchant,[52] Enzo Merler,[58] Franco Merletti,[59] Anthony Miller,[60] Franklin E. Mirer,[61] Richard Monson,[8] Karl-Cristian Nordby,[47] Andrew F. Olshan,[24] Marie-Elise Parent,[62] Frederica P. Perera,[63] Melissa J. Perry,[64] Angela Cecilia Pesatori,[10] Roberta Pirastu,[19] Miquel Porta,[65] Eero Pukkala,[66] Carol Rice,[67] David B. Richardson,[24] Leonard Ritter,[68] Beate Ritz,[69] Cecile M. Ronckers,[70] Lesley Rushton,[71] Jennifer A. Rusiecki,[72] Ivan Rusyn,[73] Jonathan M. Samet,[74] Dale P. Sandler,[75] Silvia de Sanjose,[76] Eva Schernhammer,[8] Adele Seniori Costantini,[77] Noah Seixas,[25] Carl Shy,[24] Jack Siemiatycki,[78] Debra T. Silverman,[2] Lorenzo Simonato,[79] Allan H. Smith,[80] Martyn T. Smith,[81] John J. Spinelli,[82] Margaret R. Spitz,[83] Lorann Stallones,[84] Leslie T. Stayner,[85] Kyle Steenland,[86] Mark Stenzel,[87] Bernard W. Stewart,[88] Patricia A. Stewart,[89] Elaine Symanski,[90] Benedetto Terracini,[91] Paige E. Tolbert,[86] Harri Vainio,[92] John Vena,[93] Roel Vermeulen,[38] Cesar G. Victora,[94] Elizabeth M. Ward,[95] Clarice R. Weinberg,[96] Dennis Weisenburger,[97] Catharina Wesseling,[98] Elisabete Weiderpass,[99] and Shelia Hoar Zahm[100]*

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom; [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, Maryland, USA; [3]Imperial College, London, United Kingdom; [4]Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany; [5]Department of Research, Cancer Registry of Norway, Oslo, Norway; [6]Centre for Research in Environmental Epidemiology (CREAL), IMIM (Hospital del Mar Medical Research Institute), Universitat Pompeu Fabra (UPF), and CIBER Epidemiologia y Salud Publica (CIBERESP), Barcelona, Spain; [7]School of Public Health, The University of Sydney and Sax Institute, Sydney, New South Wales, Australia; [8]Department of Environmental Health, and Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; [9]Division of Biochemical Toxicology, National Center for Toxicological Research, Jefferson, Arkansas, USA; [10]Department of Clinical Sciences and Community Health, University of Milan and IRCCS Foundation Ca'Granda Ospedale Maggiore Policlinico, Milan, Italy; [11]NCI, NIH, DHHS, Research Triangle Park, North Carolina, USA; [12]Division of Public Health Sciences, and Alvin J. Siteman Cancer Center, Washington University, School of Medicine, St. Louis, Missouri, USA; [13]Division of the National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS), NIH, DHHS, Research Triangle Park, North Carolina, USA; [14]KP Cantor Environmental LLC, Silver Spring, Maryland, USA; [15]CREAL, IMIM, and CIBERESP, Barcelona, Spain; [16]Institute of Occupational Medicine, Edinburgh, United Kingdom; [17]Department of Public Health, Clinical and Molecular Medicine, University of Cagliari-Monserrato, Cagliari, Italy; [18]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom; [19]Department of Biology and Biotechnology, "Charles Darwin" Sapienza Rome University, Rome, Italy; [20]Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Ontario, Canada; [21]Division of Occupational and Environmental Medicine, Duke University Medical Center, Durham, North Carolina, USA; [22]Centre for Public Health Research, Massey University, Wellington, New Zealand; [23]Department of Environmental Health Science, and Department of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California, USA; [24]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; [25]Department of Environmental and Occupational Health Sciences, University of Washington School of Public Health, Seattle, Washington, USA; [26]European Centre for Environment and Human Health, University of Exeter Medical School, Truro, Cornwall, United Kingdom; [27]Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, London, United Kingdom; [28]Louisiana State University School of Public Health, New Orleans, Louisiana, USA; [29]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy; [30]Center for Environmental Research and Sustainable Technology (UFT), Universität Bremen, Bremen, Germany; [31]School of Public Health, Curtin University, Perth, Western Australia, Australia; [32]Department of Environmental and Occupational Health, Ecole de Santé Publique, Universite de Montreal, Montreal, Quebec, Canada; [33]Population-based Epidemiological Cohorts Unit, Inserm UMS 011, Villejuif, France; [34]Department of Environmental Medicine, University of Southern Denmark, Odense, Denmark; [35]Institute of Environmental Medicine (IMM), Karolinska Institutet, Stockholm, Sweden; [36]Danish Cancer Society Research Center, Copenhagen, Denmark; [37]Department of Epidemiology and Biostatistics, Netherlands Cancer Institute, Amsterdam, the Netherlands; [38]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands; [39]German Cancer Research Center (DKFZ), Heidelberg, Germany; [40]Epidemiology and Biostatistics Sorbonne Paris Cite Center (CRESS), INSERM, UMR 1153, Epidemiology of Childhood and Adolescent Cancers Research Group (EPICEA), Paris Descartes University, F-75015, Paris, France; [41]Department of Public Health Sciences, University of California, Davis, Davis, California, USA; [42]Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA; [43]NIEHS, NIH, DHHS, Research Triangle Park, North Carolina, USA; [44]Department of Public Health and Clinical Medicine, Umea University, Umea, Sweden; [45]College of Medicine, Seoul National University, Seoul, Korea; [46]Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA; [47]National Institute of Occupational Health, Oslo, Norway; [48]CREAL, IMIM, and CIBERESP, Barcelona, Spain; National School of Public Health, Athens Greece; [49]Department of Work Environment, University of Massachusetts Lowell, Lowell, Massachusetts, USA; [50]Centre Francois Baclesse, Universite de Caen, Caen, France; [51]Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; [52]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa, USA; [53]Department of Public Health, University of Copenhagen, Copenhagen, Denmark; [54]National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australian Capital Territory, Australia; [55]Public Health Ontario, Toronto, Ontario, Canada; [56]Prevention and Cancer Control, Cancer Care Ontario, and Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [57]World Health Organization Regional Office for Europe, European Centre for Environment and Health, Bonn, Germany; [58]Regional Mesothelioma Register, National Health Service, Local Health Authority, Padova, Italy; [59]Unit of Cancer Epidemiology, Department of Medical Sciences, University of Turin, Turin, Italy; [60]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [61]City University of New York (CUNY) School of Public Health, New York, New York, USA; [62]INRS-Institut Armand-Frappier, Universite du Quebec, Laval, Quebec, Canada; [63]Department of Environmental Health Sciences, and Columbia Center for Children's Environmental Health, Mailman School of Public Health, Columbia University, New York, New York, USA; [64]Department of Environmental and Occupational Health, George Washington University Milken Institute School of Public Health, Washington, DC, USA; [65]IMIM, CIBERESP, and School of Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain; [66]Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland, and School of Health Sciences, University of Tampere, Tampere, Finland; [67]Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; [68]School of Environmental Sciences, Ontario Agricultural College, University of Guelph, Guelph, Ontario, Canada; [69]Department of Epidemiology, Fielding School of Public Health, University of California at Los Angeles, Los Angeles, California, USA; [70]Department of Pediatric Oncology, Emma Children's Hospital/Academic Medical Center, Amsterdam, the Netherlands; [71]MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; [72]Department of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA; [73]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas, USA; [74]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA; [75]Epidemiology Branch, NIEHS, NIH, DHHS, Research Triangle Park, North Carolina, USA; [76]Cancer Epidemiology Research Programme, Unit of Infections and Cancer, Bellvitge Biomedical Research Institute (IDIBELL), CIBERESP Catalan Institute of Oncology, Barcelona, Spain; [77]Cancer Prevention and Research Institute (ISPO), Florence, Italy; [78]Department of Social and Preventive Medicine, Ecole de Santé Publique, Universite de Montreal, Montreal, Quebec, Canada; [79]Laboratory of Public Health and Population Studies, Department of Molecular Medicine, University of Padova, Padova, Italy; [80]The School of Public Health, University of California, Berkeley, Berkeley, California, USA; [81]Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, California, USA; [82]Cancer Control Research, BC Cancer Agency and School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada; [83]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, USA; [84]Department of Psychology, Colorado State University, Fort Collins, Colorado, USA; [85]Division of Epidemiology and Biostatistics, University of Illinois at Chicago, School of Public Health, Chicago, Illinois, USA; [86]Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA; [87]Exposure Assessment Applications LLC, Arlington, Virginia, USA; [88]Cancer Control Program, South East Sydney Public Health Unit, Randwick, New South Wales, Australia; [89]Stewart Exposure Assessments LLC, Arlington, Virginia, USA; [90]Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas School of Public Health, Houston, Texas, USA; [91]University of Torino and Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica, CPO Piemonte, Torino, Italy; [92]Finnish Institute of Occupational Health, Helsinki, Finland; [93]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA; [94]Universdade Federal de Pelotas, Rio Grande do Sul, Brazil; [95]American Cancer Society Inc., Atlanta, Georgia, USA; [96]Biostatistics and Computational Biology Branch, NIEHS, NIH, DHHS, Research Triangle Park, North Carolina, USA; [97]Department of Pathology, City of Hope National Medical Center, Duarte, California, USA; [98]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; [99]Department of Community Medicine, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway; Cancer Registry of Norway, Oslo, Norway; Genetic Epidemiology Group, Folkhalsan Research Center, Helsinki, Finland; Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [100]Shelia Zahm Consulting, Hermon, Maine, USA; *Deceased

Background: Recently, the International Agency for Research on Cancer (IARC) Programme for the Evaluation of Carcinogenic Risks to Humans has been criticized for several of its evaluations, and also for the approach used to perform these evaluations. Some critics have claimed that failures of IARC Working Groups to recognize study weaknesses and biases of Working Group members have led to inappropriate classification of a number of agents as carcinogenic to humans.

Objectives: The authors of this Commentary are scientists from various disciplines relevant to the identification and hazard evaluation of human carcinogens. We examined criticisms of the IARC classification process to determine the validity of these concerns. Here, we present the results of that examination, review the history of IARC evaluations, and describe how the IARC evaluations are performed.

Discussion: We concluded that these recent criticisms are unconvincing. The procedures employed by IARC to assemble Working Groups of scientists from the various disciplines and the techniques followed to review the literature and perform hazard assessment of various agents provide a balanced evaluation and an appropriate indication of the weight of the evidence. Some disagreement by individual scientists to some evaluations is not evidence of process failure. The review process has been modified over time and will undoubtedly be altered in the future to improve the process. Any process can in theory be improved, and we would support continued review and improvement of the IARC processes. This does not mean, however, that the current procedures are flawed.

Conclusions: The IARC Monographs have made, and continue to make, major contributions to the scientific underpinning for societal actions to improve the public's health.

Citation: Pearce N, Blair A, Vineis P, Ahrens W, Andersen A, Anto JM, Armstrong BK, Baccarelli AA, Beland FA, Berrington A, Bertazzi PA, Birnbaum LS, Brownson RC, Bucher JR, Cantor KP, Cardis E, Cherrie JW, Christiani DC, Cocco P, Coggon D, Comba P, Demers PA, Dement JM, Douwes J, Eisen EA, Engel LS, Fenske RA, Fleming LE, Fletcher T, Fontham E, Forastiere F, Frentzel-Beyme R, Fritschi L, Gerin M, Goldberg M, Grandjean P, Grimsrud TK, Gustavsson P, Haines A, Hartge P, Hansen J, Hauptmann M, Heederik D, Hemminki K, Hemon D, Hertz-Picciotto I, Hoppin JA, Huff J, Jarvholm B, Kang D, Karagas MR, Kjaerheim K, Kjuus H, Kogevinas M, Kriebel D, Kristensen P, Kromhout H, Laden F, Lebailly P, LeMasters G, Lubin JH, Lynch CF, Lynge E, 't Mannetje A, McMichael AJ, McLaughlin JR, Marrett L, Martuzzi M, Merchant JA, Merler E, Merletti F, Miller A, Mirer FE, Monson R, Nordby KC, Olshan AF, Parent ME, Perera FP, Perry MJ, Pesatori AC, Pirastu R, Porta M, Pukkala E, Rice C, Richardson DB, Ritter L, Ritz B, Ronckers CM, Rushton L, Rusiecki JA, Rusyn I, Samet JM, Sandler DP, de Sanjose S, Schernhammer E, Seniori Costantini A, Seixas N, Shy C, Siemiatycki J, Silverman DT, Simonato L, Smith AH, Smith MT, Spinelli JJ, Spitz MR, Stallones L, Stayner LT, Steenland K, Stenzel M, Stewart BW, Stewart PA, Symanski E, Terracini B, Tolbert PE, Vainio H, Vena J, Vermeulen R, Victora CG, Ward EM, Weinberg CR, Weisenburger D, Wesseling C, Weiderpass E, Zahm SH. 2015. IARC Monographs: 40 years of evaluating carcinogenic hazards to humans. Environ Health Perspect 123:507–514; http://dx.doi.org/10.1289/ehp.1409149

Introduction

Important advances in human health have come from the recognition of health hazards and the development of policy actions to address them (Brownson et al. 2009; Espina et al. 2013; Samet 2000). Government and nongovernmental organizations use expert panels to review the scientific literature and to assess its relevance to public health policies. Scientific experts are charged with reviewing the quality and quantity of the scientific evidence and providing scientific interpretations of the evidence that underpin a range of health policy decisions.

The *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* of the International Agency for Research on Cancer (IARC) are a prominent example of such an expert review process. The goal of the Monograph Programme is to assess carcinogenic hazards from occupational, environmental, and lifestyle exposures and agents, thus providing an essential step in the societal decision-making process to identify and then control carcinogenic hazards. For these evaluations, IARC assembles groups of scientists with a range of relevant scientific expertise (called "Working Groups") to review and assess the quality and strength of evidence from informative publications and perform a hazard evaluation to assess the likelihood that the agents of concern pose a cancer hazard to humans (Tomatis 1976). IARC has used this approach for four decades, since the first Monograph in 1972 (IARC 1972). Although widely accepted internationally, there have been criticisms of the classification of particular agents in the past, and more recent criticisms have been directed at the general approach adopted by IARC for such evaluations (Boffetta et al. 2009; Epidemiology Monitor 2012; Ioannidis 2005; Kabat 2012; McLaughlin et al. 2010, 2011).

The Monographs are widely used and referenced by governments, organizations, and the public around the world; therefore, it is critical that Working Group conclusions be clear and transparent. In addition to the actual evaluation, a major contribution of the Monographs is the assembly of relevant literature and its dissemination to the public. We recognize that no system of evaluation is perfect. It is important to foster continuing improvement of the methods used by IARC and other bodies that review scientific evidence. The IARC process itself has been modified from time to time (e.g., addition of specific evaluation of mechanistic data and greater use of formal meta-analyses and data-pooling approaches). Indeed, as recently as April 2014, the IARC Monographs program has been a subject of a review by the Advisory Group to recommend priorities for IARC Monographs during 2015–2019 (Straif et al. 2014). The Advisory Group has made a number of recommendations on further improvements in the Monographs process specifically related to conflict of interest, transparency, and the use of the systematic review procedures in data gathering and evaluation. Thus, possible changes to the process are periodically considered by IARC

Address correspondence to N. Pearce, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom. Telephone: 44-20-7958-8151. E-mail: neil.pearce@lshtm.ac.uk

This work was conducted with no direct funding, but was supported in part by the Intramural Research Program of the NIH (NCI and NIEHS).

The views expressed are those of the authors and do not necessarily reflect the official views of the Uniformed Services University of the Health Sciences, the Department of Defense, or the U.S. Food and Drug Administration.

M. Stenzel is employed by Exposure Assessment Applications LLC (Arlington, VA, USA). P. Stewart is employed by Stewart Exposure Assessments LLC (Arlington, VA, USA). S.H. Zahm is employed by Shelia Zahm Consulting (Hermon, ME, USA), and she has served as an expert witness for the plaintiffs in U.S. litigation involving polychlorinated biphenyls and non-Hodgkin lymphoma. J. Huff has served as an expert witness for plaintiffs in U.S. litigation involving exposures to acrylamide and to styrene and carcinogenicity. B. Armstrong was formerly Deputy Director of IARC (1991–1993), and J. Huff and H. Vainio have previously served as head of the IARC Monographs Program. Most (but not all) other authors have served on IARC Monograph Working groups, and several have worked for IARC in the past. Although P. Bertazzi is Director of a Department that has research and consultancy contracts with industry (including petrochemical, plastics, iron, and steel, and other chemicals), as well as governmental, private, and nonprofit agencies, he has no affiliation with industry; he has served in Italy as an expert in medical legal cases involving asbestos exposure and asbestos-induced disease and prepared reports on causation and diagnosis of asbestos-related disorders for courts. D. Christiani, J. Dement, and A. Smith have served as expert witnesses in U.S. litigation involving asbestos exposure and disease outcomes, including cancer. E. Fontham is a Senior Research Fellow at the International Prevention Research Institute. P. Comba, P. Forastiere, E. Merler, F. Merletti, R. Pirastu, B. Terracini, and P. Vineis have acted as consultants to prosecutors and judges in a number of court case trials. F. Mirer has received compensation as a consultant to the AFL-CIO and the UAW in support of litigation. E. Weiderpass is currently a member of the IARC Scientific Council. The other authors declare they have no actual or potential competing financial interests.

Received: 14 September 2014; Accepted: 20 February 2015; Advance Publication: 24 February 2015; Final Publication: 1 June 2015.

governing groups (Scientific Council and Governing Council) and Advisory Groups.

Here, we focus on current IARC processes and practices because these have been the focus of recent criticisms. The authors of this Commentary are scientists from a wide range of disciplines who are involved in designing and conducting studies that provide data used in hazard evaluations, such as those performed by IARC. Many (but not all) of us have served on IARC Monograph Working Groups, but none are current IARC staff. We first discuss the history of IARC, and describe how the IARC evaluations are performed in order to foster evidence-based policy. We then describe why unbiased evaluations, based on the evidence and free of conflicts of interest, are necessary for public health decision making. Finally, we discuss the recent criticisms of the IARC approach.

The IARC Monographs

History of the IARC Monographs. Shortly after IARC's establishment, its parent entity, the World Health Organization (WHO), asked IARC to prepare a list of agents known to cause cancer in humans. IARC recognized the need for a systematic process to determine which agents should be listed. Such a process was launched in 1972 by Lorenzo Tomatis, then Chief of the Division of Carcinogenicity of IARC (Tomatis 1976). IARC is funded by the governments of 24 countries that have decided to become members, in addition to competitive grants from funding agencies. The IARC Monograph Programme is mainly funded by the U.S. National Cancer Institute through a renewable grant subject to peer review of the program. Other sources of external funding have included the European Commission Directorate-General of Employment, Social Affairs and Equal Opportunities; the U.S. National Institute of Environmental Health Sciences; and the U.S. Environmental Protection Agency.

The IARC process antedates current systematic review methods, but anticipated some of them, for example, with regard to transparent literature identification. In the IARC process, agents are assessed for carcinogenic hazard and assigned to one of five categories, ranging from carcinogenic to humans to probably not carcinogenic to humans (Appendix 1). The classification categories are described in the preamble to the Monographs (IARC 2006). Carcinogenic hazard identification refers to an assessment of whether an agent causes cancer. Hazard identification does not predict the magnitude of cancer risks under specific conditions; this can be determined only with appropriate exposure–response information (National Research Council 2009).

The IARC Monograph process. The process for the preparation of an IARC Monograph is clearly described in the Preamble, which is published as part of each Monograph (e.g., IARC 2014a). It starts with the nomination of candidate agents. Nominations come from national regulatory agencies, scientists, and stakeholders, including public health professionals, experts in environmental or occupational hygiene, industry representatives, and private citizens. It is important to note that anyone (including private citizens) can participate in the nomination process. The Monograph Programme convenes meetings of special Advisory Groups (composed of external scientists that possess a broad range of relevant professional skills) to review agents nominated for evaluation and to suggest IARC priorities for such reviews (Ward et al. 2010). Announcements of a review are made on the IARC website (http://monographs.iarc.fr/ENG/Meetings/). For example, in 2013 IARC sought nominations for agents to be evaluated in 2015–2019 (IARC 2014b). An Advisory Group reviewed the nominated agents and exposures, added several new ones, and discussed the priorities for each.

The IARC staff makes the final selection of agents for review by taking into account the prevalence and intensity of exposure (of both occupational groups and the general population) and availability of sufficient literature for an evaluation of carcinogenicity, as well as advice from the Advisory Groups. The large majority of evaluations concern specific compounds, but there are also monographs on various occupations or industries, for example, aluminum production, insecticide applicators, firefighters, manufacture of leather goods, leather tanning and processing, welding, painters, petroleum refining, and pulp and paper manufacturing. Some individual exposures that occur in these settings have also been evaluated.

The next step is the selection of members of the Working Group (WG). IARC staff review the literature to identify Working Group candidates and specialists in relevant areas of expertise; they also seek names of possible candidates from the scientific community and advisory groups. The list of potential members, including disclosure of relevant conflicts of interest, is posted on the IARC website (http://monographs.iarc.fr/ENG/Meetings/) before the WG is convened, and anyone can send comments. Members are typically scientists who have conducted research relevant to the agent under review, but not necessarily on the specific agent. Selection procedures are evaluated yearly by the Scientific and the Governing Councils. The IARC Section of Monographs also has an external Advisory Board, made up of independent scientists, that periodically peer reviews its activities. In addition to Working Group members, invited specialists, representatives of health agencies, stakeholder observers, and the IARC Secretariat also attend meetings.

The responsibility of the Working Group is to review the literature before the Monograph meeting, discuss the literature at the meeting, and then classify whether an agent is carcinogenic, probably carcinogenic, possibly carcinogenic, not classifiable, or probably not carcinogenic to humans (see Appendix 1). Working Group members are also responsible for writing the IARC Monograph, which must both review the literature and explain why the Working Group came to their specific conclusions.

The procedures used to evaluate the scientific evidence are described in the Preamble to the Monographs (IARC 2006). It is important to stress that only Working Group members conduct the actual evaluation (Wild and Cogliano 2011; Wild and Straif 2011). IARC staff facilitate the evaluation process and ensure that the procedures described in the Preamble are followed; however, they do not determine the outcomes.

IARC assessments of carcinogenicity are based on, and necessarily limited to, scientific evidence available at the time of the review. The evidence comes from epidemiologic studies, animal bioassays, pharmacokinetic/mechanistic experiments, and surveys of human exposure. The aim is to include all relevant papers on cancer in humans and experimental animals that have been published, or accepted for publication, in peer-reviewed scientific journals and also any publicly available government or agency documents that provide data on the circumstances and extent of human exposure. To that end, the search of the literature takes a comprehensive approach. Papers that are found not to provide useful evidence can be excluded later in the process. IARC staff first use previous IARC Monographs (if available), database searches using relevant text strings, and contact with investigators in the field to identify potentially relevant material. Thus, the initial assembly of the literature is performed by individuals who are not engaged in the actual evaluation. Working Group members are then assigned various writing tasks and are instructed to perform their own literature searches to identify any further papers that might have been missed. In addition, all of the papers assembled by IARC are made available to the full Working Group before they meet, and any member can recommend other papers not previously identified that they think should be considered. Finally, papers can be recommended by stakeholder representatives before or during the Working Group meeting.

At the meeting of the Working Group, the assembled documents are reviewed and summarized by discipline-related subgroups.

However, any member of the Working Group has access to all of the assembled literature. The summaries are distributed to all subgroups, and information from all disciplines is discussed in plenary sessions prior to assigning the agents to a specific carcinogenicity category.

Because new findings continually emerge in the literature, agents are reconsidered when IARC and IARC Advisory Groups judge that there is sufficient additional information that might alter a previous evaluation. Thus, conclusions regarding human carcinogenicity of particular substances may change as new evidence becomes available. For some agents, this reevaluation has resulted in progression toward greater certainty regarding their human carcinogenicity, whereas for others the progress has been moved toward less certainty. Such movements are expected in an open, transparent, and evidence-based process. A comprehensive update of all Group 1 carcinogens was recently accomplished in Volume 100 A through F (http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php).

Usually, several agents are evaluated in a single meeting lasting more than 1 week. After discussing the evidence fully, the Working Group members follow the published IARC procedures for combining information from epidemiologic studies and bioassays to arrive at a preliminary classification (IARC 2014a). Mechanistic data are then considered in order to determine whether they warrant a change from the preliminary classification. The Working Group then votes on the final determination. Many votes are unanimous, but on occasion some reviewers may favor a higher or lower ranking than the majority. When there is dissent, alternative interpretations and their underlying reasoning are sometimes reported in the rationale for the evaluation if the dissenters feel their point of view is not sufficiently addressed in the monograph.

Consideration of the totality of the evidence. IARC Working Groups make every effort to provide full and transparent documentation of what evidence was assembled, how it was evaluated, and which papers were most important for the hazard evaluation. Consequently, the monographs are often quite lengthy, containing many evidence tables [see, for example, the recent monograph on trichloroethylene (IARC 2014c)]. Evaluations involve consideration of all of the known relevant evidence from epidemiologic, animal, pharmacokinetic/mechanistic, and exposure studies to assess cancer hazard in humans. Information on human exposure is not formally graded as part of the overall assessment of carcinogenic hazard; however, these data make a critical contribution to the process by characterizing the timing, duration, and levels of exposure in the population, and in evaluating the quality of the exposure assessment in epidemiologic studies.

Doubts and criticisms have sometimes been expressed about the relative weights attributed to evidence from individual disciplines to the assessment of cancer hazards to humans; however, each discipline provides important evidence toward the overall evaluation of causality according to the Bradford Hill considerations (Hill 1965). Because the totality of the evidence is considered, deficiencies in one discipline are often offset by strengths in another. For example, epidemiologic studies may focus on population-relevant exposures, whereas findings from animal experiments usually involve higher exposures but are less susceptible to confounding.

Long-term animal bioassays and mechanistic studies provide critical information on the capacity of an agent to produce cancer in mammalian systems, including humans, and to contribute to decisions that would lead to better protection of human health. Bioassays are the backbone of regulatory science because they provide the opportunity to rigorously evaluate potential hazards before there is widespread human exposure. Bioassays and mechanistic studies are sometimes criticized for employing exposure routes and doses that in most instances humans would not experience, although experimental dose categories sometimes approach exposure levels found in occupational situations. There is evidence that carcinogenicity in human and animal studies is often concordant, although data may differ as to the affected cancer site (Haseman 2000; Maronpot et al. 2004; Tomatis 2002). A major effort to evaluate the concordance between animal and human results is currently under way; two Working Groups were convened at IARC in 2012, and a systematic evaluation of the correspondence between human and animal data was undertaken (a report is not yet publicly available).

Criticisms of the IARC Process

IARC Monographs are widely used to identify potential carcinogenic hazards to humans and serve as reference documents summarizing the literature on many different agents. In recent years, however, individuals have criticized both the classification of individual agents as well as the general evaluative approach (Boffetta et al. 2009; Epidemiology Monitor 2012; Kabat 2012; McLaughlin et al. 2010, 2011). We discuss four of these criticisms below.

Criticisms of epidemiology. Some of the criticisms of the IARC process have occurred in the context of more general criticisms of epidemiology as a science (Kabat 2008); these were discussed in detail by Blair et al. (2009). Potential methodological weaknesses for observational epidemiologic studies are well recognized and can be found in any epidemiologic textbook (Checkoway et al. 2004; Rothman et al. 2008). Most studies are subject to one or more methodological limitations, but this does not necessarily invalidate their findings (Blair et al. 2009). In fact, the value of epidemiologic studies has been shown by the identification of a number of well-established human carcinogens, including tobacco, asbestos, benzene, hexavalent chromium, and some viruses, in multiple studies. Some critics also argue that small or nonexistent health risks are unjustifiably highlighted and hyped by researchers who have a vested interest in continued research funding and the need to publish to benefit their careers (Boffetta et al. 2008; Kabat 2008; McLaughlin et al. 2010, 2011; Taubes 1995). However, such overstated results are unlikely to exert much of an influence in a Monograph because IARC evaluations are based on the totality of the evidence. The problem would have to occur in multiple studies, and the Working Group would have to be unable to identify it or be unwilling to weigh such studies appropriately. Incorrect positive conclusions regarding carcinogenicity may also occur in reviews of multiple studies because of publication bias, which may selectively populate the literature only with "positive" findings. However, once a topic is recognized as scientifically important, reports on relevant studies will be published regardless of the findings, so publication bias is mainly a concern for newly arising issues. To evaluate the potential for publication bias, Working Groups consider whether stronger negative studies (both in terms of design and sample size) have emerged after publication of an initial cluster of smaller and/or weaker positive studies. Funnel plots help in the assessment of bias relating to sample size and publication bias (Borenstein et al. 2009). In contrast, there are no established statistical techniques to clearly characterize strength of design.

One of the distinctive features of epidemiology is that criticism and self-criticism are firmly embedded in the discipline. A great deal of work has been done on developing methods for critical appraisal (Elwood 2007) and for assessing the likely strength and direction of possible biases (Rothman et al. 2008). Epidemiologists and other members on Working Groups routinely use various approaches to assess possible bias in study design and analysis when weighing the strengths of different studies.

The issue of false positives. Epidemiology specifically has been criticized for a tendency to produce false-positive results (i.e., individual study associations not borne out by the weight of the evidence) or to preferentially report positive findings over negative

or inconclusive findings (i.e., publication bias) (Boffetta et al. 2008, 2009; Ioannidis 2005; Kabat 2012; McLaughlin and Tarone 2013). This criticism has been most often applied to potential false positives from individual studies, but it has been inferred that this problem may also apply to overall hazard evaluations, which use findings from multiple studies. We will consider each of these issues in turn.

False-positive findings may occur by chance, particularly when many combinations of exposures and health outcomes have been examined in a single study without strong prior expectations of association; this happens often, for example, in genome-wide association studies where thousands of gene–disease associations are evaluated. Chance, of course, operates in all disciplines and in both observational and experimental studies. However, there are well-known statistical techniques to reduce the probability of declaring chance findings as "positive" (Rothman et al. 2008). Independent replication, however, is the most convincing way of checking for "chance" findings; hazard evaluations, such as those conducted by IARC Working Groups, rely heavily on reproducibility in independent studies and also interpret data following Bradford Hill principles (Hill 1965).

False negatives are more difficult to address, and perhaps they occur more frequently than false positives because of low statistical power, nondifferential misclassification of exposure and/or outcome, and incomplete follow-up, which tends to reduce the observed difference in risk between the exposed and nonexposed populations (Ahlbom et al. 1990; Blair et al. 2009; Grandjean 2005; Rothman et al. 2008). A new positive association stimulates research, whereas studies finding no associations tend to stifle further work.

There are difficulties in conducting epidemiologic studies of agents that are relatively "weak" carcinogens, or for stronger carcinogens where exposure is very low because bias and confounding can obscure weak positive associations (MacMahon et al. 1981). In general, weak carcinogens and low levels of exposure result in a smaller "signal-to-noise" ratio making the real signal more difficult to detect. Although the identification of small relative risks to humans poses special challenges to scientific research, the refinement of study designs, improvements in methods of exposure assessment, and the use of biomarkers have helped to address the problems (e.g., newer studies on the effects of air pollution, the growth in opportunities to examine gene–environment interactions) (Gallo et al. 2011). In some situations, there is less of a problem. For example, in occupational studies, exposures and relative risks may be higher while differences in lifestyle factors between different groups of workers are smaller (Checkoway et al. 2004); thus, any confounding by nonoccupational factors is likely to be weak, even from potent causes of cancer such as cigarette smoking (Siemiatycki et al. 1988). Of course, the interpretation of such studies is enhanced when there is supporting evidence from bioassays and/or mechanistic studies.

False-positive and false-negative findings in individual studies may arise by chance or bias, including bias due to confounding (Rothman et al. 2008). However, the evaluation of multiple independent epidemiologic studies from various geographic locations, involving a variety of study designs, as well as evidence from experimental studies, reduces the possibility that false-positive findings from any individual study influences the overall evaluation process. Some studies may have greater influence than others because of methodological strengths and/or large sample size. The use of information from a variety of study designs reduces the likelihood of false-positive evaluations because it is unlikely that the same biases will occur in multiple studies based on different populations under different study designs. Moreover, apparently conflicting results from epidemiologic studies do not necessarily indicate that some are false positive or false negative. This might, for example, reflect differences in levels of exposure or susceptibility to the effects of exposure (effect modification). Finally, judgment by the Working Group is not based exclusively on epidemiologic studies but usually also on results from laboratory and mechanistic studies that provide further evidence and biological coherence. For the Monographs that evaluate carcinogenic hazards associated with specific occupations or industries, the exposures of interest usually involve a complex mixture of chemicals. For these evaluations, most information comes from epidemiologic studies, although exposures to individual agents occurring at these workplaces may have been evaluated in experimental studies.

Discontent with IARC Monograph processes. The IARC Monograph evaluation process has been criticized and it has been alleged that "a number of scientists with direct experience of IARC have felt compelled to dissociate themselves from the agency's approach to evaluating carcinogenic hazards" (Kabat 2012). This is a serious charge. However, the author of this claim provided no evidence to support the charge that a "number of scientists" have dissociated themselves from the process, nor has there been any indication of how many scientists have taken this step, or for what reason. In science, we expect sweeping statements such as this to be appropriately documented. We have not been able to identify any credible support for this contention.

There is an IARC Governing Council and a Scientific Council to provide oversight and guidance to the agency. The Governing Council represents the participating states and sets general IARC policy. It appoints the IARC Director and members of the Scientific Council. The latter are independent scientists who are selected to provide scientific expertise and not as representatives of the member states. They serve for 4 years and serve without pay. The voting members of Monograph Working Groups are not employed by IARC, and they perform this task without financial compensation. There have been 111 volumes, including six separate documents under Volume 100, and three Supplements. Over the years, as the number of publications for each agent to be evaluated increased, the size of Working Groups has increased. Early in the process they were sometimes as small as 10, but now they sometimes include as many as 30 scientists. We estimate that over the entire Monograph series, approximately 1,500 scientists have served as Working Group members, and of course many scientists have also served on the Advisory Groups, Scientific Council, and Governing Council. Thus, if even a small percentage of these scientists were disenchanted with the IARC process, it would result in a considerable number of such individuals and should be easy to document. To be taken seriously, the "dissociation" criticism needs to be supported by documented information describing the number of scientists who have taken this action.

Criticisms of specific evaluations. Some criticisms of the IARC process relate to specific agents, where it is asserted that the hazard evaluations of category 2B, 2A, or 1 are not supported by the scientific literature. In the 111 volumes of the Monographs produced over the four decades since 1971, 970 agents have been considered, 114 (12%) have been classified as carcinogenic to humans (Group 1), 69 (7%) as probably carcinogenic (Group 2A), 283 (29%) as possibly carcinogenic (Group 2B), 504 (52%) as not classifiable regarding their carcinogenicity (Group 3), and 1 (< 1%) as probably not carcinogenic to humans (Group 4). Thus, even for this highly select group of agents (i.e., those selected for evaluation because there was some concern that they might be carcinogenic), more than one-half were "not classifiable" or "probably not carcinogenic," and a further 29% were placed into the category of possibly carcinogenic to humans. This distribution, based on nearly 1,000 evaluations in which fewer than one in five agents were classified as carcinogenic or probably carcinogenic to humans, does not support a conclusion that the process is heavily biased

toward classifying agents as carcinogenic (Boffetta et al. 2009; Kabat 2012).

The monographs for formaldehyde, coffee, DDT, and radiofrequency electromagnetic radiation have been cited as examples of problematic evaluations by some (Kabat 2012) [among these, only formaldehyde was classified as known to be carcinogenic to humans (Group 1) by an IARC Working Group]. These are important agents. However, to accept the charge that IARC evaluations are fundamentally biased, one has to assume that the scientists who were members of the Working Groups were incapable of appropriately evaluating weaknesses in the data, or that they distorted the evaluative process because of personal biases. In our experience, neither of these assertions is correct. Dissent among scientists is not unusual in any area of science. It is a strength of the scientific process. The IARC process capitalizes on this by bringing scientists from different disciplines together in one room to evaluate the literature and to reach a reasoned conclusion. Differences of opinion occur among Working Group members. These differences, however, typically involve disputes related to assignment to adjacent classification categories. It is instructive that there are no instances in which a carcinogen classified at the Group 1 level by one Working Group has been reversed by another. The recent review of all Group 1 agents for Volume 100 provided ample opportunity to reverse such previous classifications, but none occurred. Every scientist could probably name a substance that has been reviewed by IARC that they might personally place in a different category from that assigned by the Working Group, but this is one opinion against the collective wisdom and process of the Working Group.

Criticisms of the composition of the working groups. The composition of the Working Groups has also been criticized (Erren 2011; McLaughlin et al. 2010, 2011); it has been argued that members of the Working Groups who have conducted research on the agents under evaluation have a vested interest in advancing their own research results in the deliberations. This criticism has been addressed directly by Wild and colleagues (Wild and Cogliano 2011; Wild and Straif 2011) from IARC, and we know of no evidence to support this contention. Even if some scientists on the Working Group have performed research on some of the agents being considered, they make up a minority of the Working Group because several agents are usually evaluated in a single meeting, so the number of Working Group members who have conducted research on any one agent is typically small. Our experience has been that having some scientists who are knowledgeable about the studies of the agent under evaluation (and can therefore answer technical queries) and others from different, but related, fields provides a knowledgeable and balanced mix of scientific backgrounds for a thoughtful evaluation of the literature.

Working Group members do not receive any fee for their work, but they are paid travel expenses, and there is some prestige associated with service on an IARC Monograph. However, most scientists asked to serve on IARC Working Groups have already achieved some measure of scientific stature, and there is no reason why this should bias their evaluation in one direction or the other. In addition,

Appendix 1: Classification Categories for the Overall Evaluation for the IARC Monographs (IARC 2006)

Group 1: The agent is *carcinogenic to humans*.
This category is used when there is sufficient evidence of carcinogenicity in humans. Exceptionally, an agent may be placed in this category when evidence of carcinogenicity in humans is less than sufficient but there is sufficient evidence of carcinogenicity in experimental animals and strong evidence in exposed humans that the agent acts through a relevant mechanism of carcinogenicity.

Group 2.
This category includes agents for which, at one extreme, the degree of evidence of carcinogenicity in humans is almost *sufficient*, as well as those for which, at the other extreme, there are no human data but for which there is evidence of carcinogenicity in experimental animals. Agents are assigned to either Group 2A (*probably carcinogenic to humans*) or Group 2B (*possibly carcinogenic to humans*) on the basis of epidemiological and experimental evidence of carcinogenicity and mechanistic and other relevant data. The terms *probably carcinogenic* and *possibly carcinogenic* have no quantitative significance and are used simply as descriptors of different levels of evidence of human carcinogenicity, with *probably carcinogenic* signifying a higher level of evidence than *possibly carcinogenic*.

Group 2A: The agent is *probably carcinogenic to humans*.
This category is used when there is *limited evidence of carcinogenicity* in humans and *sufficient evidence of carcinogenicity* in experimental animals. In some cases, an agent may be classified in this category when there is *inadequate evidence of carcinogenicity* in humans and *sufficient evidence of carcinogenicity* in experimental animals and strong evidence that the carcinogenesis is mediated by a mechanism that also operates in humans. Exceptionally, an agent may be classified in this category solely on the basis of *limited evidence of carcinogenicity* in humans. An agent may be assigned to this category if it clearly belongs, based on mechanistic considerations, to a class of agents for which one or more members have been classified in Group 1 or Group 2A.

Group 2B: The agent is *possibly carcinogenic to humans*.
This category is used for agents for which there is *limited evidence of carcinogenicity* in humans and less than *sufficient evidence of carcinogenicity* in experimental animals. It may also be used when there is *inadequate evidence of carcinogenicity* in humans but there is *sufficient evidence of carcinogenicity* in experimental animals. In some instances, an agent for which there is *inadequate evidence of carcinogenicity* in humans and less than *sufficient evidence of carcinogenicity* in experimental animals together with supporting evidence from mechanistic and other relevant data may be placed in this group. An agent may be classified in this category solely on the basis of strong evidence from mechanistic and other relevant data.

Group 3: The agent is *not classifiable as to its carcinogenicity to humans*.
This category is used most commonly for agents for which the evidence of carcinogenicity is *inadequate* in humans and *inadequate* or *limited* in experimental animals.
Exceptionally, agents for which the evidence of carcinogenicity is *inadequate* in humans but *sufficient* in experimental animals may be placed in this category when there is strong evidence that the mechanism of carcinogenicity in experimental animals does not operate in humans.

Agents that do not fall into any other group are also placed in this category.

An evaluation in Group 3 is not a determination of noncarcinogenicity or overall safety. It often means that further research is needed, especially when exposures are widespread or the cancer data are consistent with differing interpretations.

Group 4: The agent is *probably not carcinogenic to humans*.
This category is used for agents for which there is *evidence suggesting lack of carcinogenicity* in humans and in experimental animals. In some instances, agents for which there is *inadequate evidence of carcinogenicity* in humans but *evidence suggesting lack of carcinogenicity* in experimental animals, consistently and strongly supported by a broad range of mechanistic and other relevant data, may be classified in this group.

IARC strictly requires that any conflict of interests be divulged, and does not allow those with conflicts of interest to serve on Working Groups, although nonvoting observers who may have conflicts of interest are able to attend the Working Group meetings.

Conclusions

For more than four decades the IARC Monograph Programme has provided evaluations of cancer hazards to humans from many different exposures and agents. These are often the first evaluations of new and emerging threats to public health and, consequently, are subject to intense scrutiny. Although these evaluations are widely respected and used by many organizations, institutions, companies, and government agencies to improve the public's health, IARC has recently been subject to criticism over conclusions on specific agents, the process that leads to such conclusions, and membership of the Working Groups. Debate and criticism facilitate self-correction and a check on the validity in science. We are concerned, however, that the criticisms expressed by a vocal minority regarding the evaluations of a few agents may promote the denigration of a process that has served the public and public health well for many decades for reasons that are not supported by data.

There has been very broad involvement of the scientific community in the IARC Monograph Programme through participation in the Working Groups and service on the IARC Governing and Scientific Councils and ad hoc Advisory Board for the Monograph Programme. The long list of scientists who are coauthors of this paper attests to the strong support that IARC has in the scientific community. Many exposures that IARC has evaluated have also been independently evaluated by other institutions, such as the U.S. National Toxicology Program (https://ntp.niehs.nih.gov/); U.S. Environmental Protection Agency (http://www.epa.gov/); National Academy of Sciences (http://www.nasonline.org/); the American Conference of Governmental Industrial Hygienists (ACGIH) Threshold Limit Values and Biological Exposure Indices (http://www.acgih.org/); the Nordic Expert Group for Criteria Documentation of Health Risks from Chemicals (http://www.av.se/arkiv/neg/); Institute of Occupational Medicine (http://www.iom-world.org/); World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) Expert Reports; European Chemicals Agency (https://echa.europa.eu); Swedish Criteria Group for Occupational Standards (2013); California Office of Environmental Hazard Assessment (Proposition 65; http://oehha.ca.gov/prop65/background/p65plain.html); Health Canada Bureau of Chemical Safety (http://www.hc-sc.gc.ca/ahc-asc/branch-dirgen/hpfb-dgpsa/fd-da/bcs-bsc/index-eng.php); Scientific Committee on Occupational Exposure Limits (SCOEL), European Commission, Employment, Social Affairs and Inclusion (http://ec.europa.eu/social/main.jsp?catId=148&langId=en&intPageId=684); European Food Safety Authority (EFSA 2013); and European Chemicals Agency (ECHA; http://echa.europa.eu/). Assessments from these groups typically come to conclusions similar to those from IARC. This further indicates broad agreement within the scientific community regarding evidence on carcinogenicity in the scientific literature and expands the number of scientists who do not have a "vested interest" but who have generally agreed with those conclusions.

Disagreement with the conclusions in an IARC Monograph for an individual agent is not evidence for a failed or biased approach. Some disagreement about the carcinogenic hazard of important agents seems inherent to the scientific enterprise and is unavoidable at early stages of the hazard evaluation, where IARC usually operates. Because the evaluations are not—and should not be—static, it is difficult to see how such assessments could be addressed any differently. Substances now universally recognized as human carcinogens (e.g., tobacco, asbestos) at one time went through a quite lengthy period of contentious debate (Michaels 2006, 2008). Any process can in theory be improved with fair and constructive criticism; appropriate reviews may take place from time to time, and we would support continued review and improvement of the IARC processes. However, as a group of international scientists, we have looked carefully at the recent charges of flaws and bias in the hazard evaluations by IARC Working Groups, and we have concluded that the recent criticisms are unfair and unconstructive.

References

Ahlbom A, Axelson O, Støttrup Hansen ES, Hogstedt C, Jensen J, Olsen J. 1990. Interpretation of "negative studies" in occupational epidemiology. Scand J Work Environ Health 16:153–157.

Blair A, Saracci R, Vineis P, Cocco P, Forastiere F, Grandjean P, et al. 2009. Epidemiology, public health, and the rhetoric of false positives. Environ Health Perspect 117:1809–1813; doi:10.1289/ehp.0901194.

Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. 2008. False-positive results in cancer epidemiology: a plea for epistemological modesty. J Natl Cancer Inst 100:988–995.

Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. 2009. Authors' response. A further plea for adherence to the principles underlying science in general and the epidemiologic enterprise in particular. Int J Epidemiol 38:678–679.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. 2009. Introduction to Meta-Analysis. West Sussex, England:Wiley.

Brownson RC, Chriqui JF, Stamatakis KA. 2009. Understanding evidence-based public health policy. Am J Public Health 99:1576–1583.

Checkoway H, Pearce N, Kriebel D. 2004. Research Methods in Occupational Epidemiology. 2nd ed. New York:Oxford University Press.

Elwood M. 2007. Critical Appraisal of Epidemiological Studies and Clinical Trials. 3rd ed. New York:Oxford University Press.

Epidemiology Monitor. 2012. Epidemiologists speak out about the challenge of false positives in cancer epidemiology. Epidemiology Monitor 33(11):9–10.

Erren TC. 2011. IARC's plea for traditional 'expert' working groups—a recipe for problems? [Letter]. Int J Epidemiol 40:1727–1728.

Espina C, Porta M, Schüz J, Hernández Aguado I, Percival RV, Dora C, et al. 2013. Environmental and occupational interventions for primary prevention of cancer: a cross-sectorial policy framework. Environ Health Perspect 121:420–426; doi:10.1289/ehp.1205897.

European Food Safety Authority. 2013. Scientific Opinion on the Hazard Assessment of Endocrine Disruptors: Scientific Criteria for Identification of Endocrine Disruptors and Appropriateness of Existing Test Methods for Assessing Effects Mediated by These Substances on Human Health and the Environment. Available: http://www.efsa.europa.eu/en/search/doc/3132.pdf [accessed 21 April 2015].

Gallo V, Egger M, McCormack V, Farmer PB, Ioannidis JPA, Kirsch-Volders M, et al. 2011. STrengthening the Reporting of OBservational studies in Epidemiology–Molecular Epidemiology (STROBE-ME): an extension of the STROBE Statement. PLoS Med 8:e1001117; doi:10.1371/journal.pmed.1001117.

Grandjean P. 2005. Non-precautionary aspects of toxicology. Toxicol Appl Pharmacol 207(2 suppl):652–657.

Haseman JK. 2000. Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. Drug Metab Rev 32:169–186.

Hill AB. 1965. The environment and disease: association or causation? Proc R Soc Med 58:295–300.

IARC (International Agency for Research on Cancer). 1972. Some inorganic substances, chlorinated hydrocarbons, aromatic amines, N-nitroso compounds, and natural products. IARC Monogr Eval Carcinog Risk Hum 1:17–184. Available: http://monographs.iarc.fr/ENG/Monographs/vol1-42/mono1.pdf [accessed 20 April 2015].

IARC (International Agency for Research on Cancer). 2006. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Preamble. Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed 20 April 2015].

IARC (International Agency for Research on Cancer). 2014a. Preamble. IARC Monogr Eval Carcinog Risk Hum 106:7–30. Available: http://monographs.iarc.fr/ENG/Monographs/vol106/mono106.pdf [accessed 20 April 2015].

IARC (International Agency for Research on Cancer). 2014b. Report of the Advisory Group to Recommend Priorities for IARC Monographs during 2010–2014. Available: http://monographs.IARC.fr/ENG/Publications/internrep/14-001.pdf [accessed 20 April 2015].

IARC (International Agency for Research on Cancer). 2014c. Trichloroethylene, tetrachloroethylene, and some other chlorinated agents. IARC Monogr Eval Carcinog Risk Hum 106:35–512. Available: http://monographs.iarc.fr/ENG/Monographs/vol106/mono106.pdf [accessed 20 April 2015].

Ioannidis JPA. 2005. Why most published research findings are false. Plos Med 2:e124; doi:10.137/journal.pmed.0020124.

Kabat GC. 2008. Hyping Health Risks: Environmental Hazards in Daily Life and the Science Of Epidemiology. New York:Columbia Univeristy Press.

Kabat G. 2012. How Activism Distorts the Assessment of Health Risks. Forbes, 20 November. Available: http://www.forbes.com/sites/realspin/2012/11/20/how-activism-distorts-the-assessment-of-health-risks/ [accessed 20 April 2015].

MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. 1981. Coffee and cancer of the pancreas. N Engl J Med 304:630–633.

Maronpot RR, Flake G, Huff J. 2004. Relevance of animal carcinogenesis findings to human cancer predictions and prevention. Toxicol Pathol 32(suppl 1):40–48.

McLaughlin JK, Boffetta P, La Vecchia C, Lipworth L, Blot WJ, Tarone RE. 2011. Authors' response. Problems with IARC's 'expert' working groups [Letter]. Int J Epidemiol 40:1728–1729.

McLaughlin JK, Lipworth L, Tarone RE, La Vecchia C, Blot WJ, Boffetta P. 2010. Authors' response. Re: A further plea for adherence to the principles underlying science in general and the epidemiologic enterprise in particular [Letter]. Int J Epidemiol 39:1679–1680.

McLaughlin JK, Tarone RE. 2013. False positives in cancer epidemiology. Cancer Epidemiol Biomarkers Prev 22:11–15.

Michaels D. 2006. Manufactured uncertainty: protecting public health in the age of contested science and product defense. Ann NY Acad Sci 1076:149–162.

Michaels D. 2008. Doubt is Their Product: How Industry's Assault on Science Threatens Your Health. New York:Oxford University Press.

National Research Council. 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC:National Academies Press. Available: http://www.nap.edu/openbook.php?record_id=12209 [accessed 20 April 2015].

Rothman KJ, Greenland S, Lash TL. 2008. Modern Epidemiology. 3rd ed. Philadelphia:Lippincott Williams & Wilkins.

Samet JM. 2000. Epidemiology and policy: the pump handle meets the new millennium. Epidemiol Rev 22:145–154.

Siemiatycki J, Wacholder S, Dewar R, Cardis E, Greenwood C, Richardson L. 1988. Degree of confounding bias related to smoking, ethnic group, and socioeconomic status in estimates of the associations between occupation and cancer. J Occup Med 30:617–625.

Straif K, Loomis D, Guyton K, Grosse Y, Lauby-Secretan B, El Ghissassi F, et al. 2014. Future priorities for the IARC monographs. Lancet Oncol 15:683–684.

Swedish Criteria Group for Occupational Standards. 2013. Scientific Basis for Swedish Occupational Standards XXXII. Available: https://gupea.ub.gu.se/bitstream/2077/34986/1/gupea_2077_34986_1.pdf [accessed 21 April 2015].

Taubes G. 1995. Epidemiology faces its limits. Science 269:164–169.

Tomatis L. 1976. The IARC Program on the Evaluation of Carcinogenic Risk of Chemicals to Man. Ann NY Acad Sci 271:396–409.

Tomatis L. 2002. The IARC Monographs Program: changing attitudes towards public health. Int J Occup Environ Health 8:144–152.

Ward EM, Schulte PA, Straif K, Hopf NB, Caldwell JC, Carreón T, et al. 2010. Research recommendations for selected IARC-classified agents. Environ Health Perspect 118:1355–1362; doi:10.1289/ehp.0901828.

Wild CP, Cogliano VJ. 2011. A plea on behalf of expert evaluation and the experts involved [Letter]. Int J Epidemiol 40:253.

Wild CP, Straif K. 2011. Authors' response. Expert working groups—a reliable recipe [Letter]. Int J Epidemiol 40:1730–1731.