

## False-Positive Results in Cancer Epidemiology: A Plea for Epistemological Modesty

Paolo Boffetta, Joseph K. McLaughlin, Carlo La Vecchia, Robert E. Tarone, Loren Lipworth, William J. Blot

False-positive results are inherent in the scientific process of testing hypotheses concerning the determinants of cancer and other human illnesses. Although much of what is known about the etiology of human cancers has arisen from well-conducted epidemiological studies, epidemiology has been increasingly criticized for producing findings that are often sensationalized in the media and fail to be upheld in subsequent studies. Herein we describe examples from cancer epidemiology of likely false-positive findings and discuss conditions under which such results may occur. We suggest general guidelines or principles, including the endorsement of editorial policies requiring the prominent listing of study caveats, which may help reduce the reporting of misleading results. Increased epistemological humility regarding findings in epidemiology would go a long way to diminishing the detrimental effects of false-positive results on the allocation of limited research resources, on the advancement of knowledge of the causes and prevention of cancer, and on the scientific reputation of epidemiology and would help to prevent oversimplified interpretations of results by the media and the public.

J Natl Cancer Inst 2008;100:988–995

False-positive results are an inherent feature of biomedical research. They are a source of inconsistent and misleading evidence and have potential impact on approaches to prevent and cure diseases and on the allocation of research resources. Some have argued that the majority of positive research findings are likely to be false, especially those based on large-throughput molecular approaches that involve many associations between “determinants” and “outcomes” (1,2).

Epidemiology is particularly prone to the generation of false-positive results. Although epidemiological studies have been key to the identification and quantification of cancer risks that are associated with cigarette smoking, alcohol consumption, asbestos and other occupational exposures, radiation, infectious agents, medications, and other factors (3), the reporting of associations that are not replicated is also a common occurrence in epidemiology. The problem is induced in part by the generation in many epidemiological studies of large sets of results, which are usually in the form of associations between multiplicities of both risk and protective factors and health outcomes. The challenge of making inferences in the face of such multiple comparisons is often made more difficult by the increasing use of multiple exposure metrics and baseline reference categories in calculating risk estimates.

Results relating to the primary a priori objectives of the study are often not the only data presented. In addition, “positive” or “statistically significant” findings from secondary analyses may be selectively reported as either generating new hypotheses or confirming ex post facto derived hypotheses. Further exacerbating the problem in epidemiological studies is the search for and reporting of weak associations, among which the potential for the distorting influences of bias, confounding, and chance is further enhanced.

Avoidance of false-positive results in epidemiology is desirable for many reasons. Such erroneous scientific evidence may lead to inappropriate governmental and public health decisions, including the introduction of costly and potentially harmful measures. The

damage from erroneous findings, however, extends beyond their immediate societal and financial consequences and entails aspects such as loss of credibility of epidemiological results in the eyes of the public, waste of public resources, and misguided research objectives among epidemiologists themselves (4).

Despite the potential impact of false-positive findings, there have been few attempts to review the extent of false-positive results in nonexperimental epidemiological research and to offer some guidance for recognizing and avoiding them. Some systematic reviews have noted general inadequacies in the design, analysis, and reporting of epidemiological studies, including subgroup analyses and multiplicity of associations being explored, both of which increase the likelihood of false-positive results (5). In the area of genetic epidemiology, the problem of false-positive results has received increasing attention (1,6–8)

In this commentary, we describe examples of apparent false-positive results—some chosen from environmental or occupational settings and others from cancer epidemiology in general. Although

---

**Affiliations of authors:** Lifestyle, Environment and Cancer Group, International Agency for Research on Cancer, Lyon, France (PB); International Epidemiology Institute, Rockville, MD (JKM, RET, LL, WJB); Vanderbilt University Medical Center, Vanderbilt-Ingram Cancer Center, Nashville, TN (JKM, RET, LL, WJB); Laboratory of Epidemiology, Mario Negri Institute, Milan, Italy (CLV); Institute of Medical Statistics and Biometry, University of Milan, Milan, Italy (CLV).

**Correspondence to:** Paolo Boffetta, MD, PhD, Lifestyle, Environment and Cancer Group, International Agency for Research on Cancer, 150 cours Albert Thomas, 69008 Lyon, France (e-mail: boffetta@iarc.fr).

See “Notes” following “References.”

**DOI:** 10.1093/jnci/djn191

© 2008 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

these examples are not necessarily typical of all false-positive associations, they illustrate the problems inherent in the identification of cancer hazards and the quantification of level of risk. The failure to replicate initial positive findings in subsequent research will be illustrated using cumulative meta-analyses based on a random-effects model (9). We then suggest possible guidelines for the reduction of the false-positive problem, including a call for some degree of epistemological modesty when presenting initial observational findings and a toning down by epidemiologists in their interpretation and presentation of positive results to the scientific community as well as to the media and general public.

## Examples of False-Positive Results in Environmental and Occupational Cancer Epidemiology

Early reports of an increased risk of cancer in a particular exposed group have led to the identification of several agents as human carcinogens (3), but for many other agents the suspicion of a carcinogenic effect has not been confirmed in subsequent studies. Because cancer epidemiology has great societal impact, and avoidance of false-positive reports in this field deserves particular attention, we have selected one example from environmental epidemiology and one example of a reported occupational cancer association. Obviously, considerations addressed in these examples are applicable to other areas of epidemiological research as well.

The first example is of an apparently strong positive result for a pollutant in the general environment that was initially suspected to give rise to cancer but for which such an association has not been confirmed in subsequent studies. In 1993, a report of a case-control study nested in a New York City-based cohort (10) showed an association between breast cancer risk and serum levels of 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene (DDE), the major metabolite of 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (DDT), an organo-

chlorine pesticide that was used extensively until the early 1970s. The study involved 58 women who had been diagnosed with breast cancer within 6 months of entry in the prospective cohort study and 171 control subjects from the same cohort, and it reported a relative risk of 3.7 (95% confidence interval [CI] = 1.0 to 13.5) for the highest vs lowest 20% of the DDE distribution in serum. This study was prompted by earlier results of small (<50 case patients) case-control studies (11–14) that had suggested a modest association between breast cancer and exposure to DDE and polychlorinated biphenyls. An editorial accompanying the report of the New York City study (15) stated that these findings may have “extraordinary implications for the prevention of breast cancer” and that this study should “serve as a wake-up call for further urgent research.”

Accordingly, a series of studies were subsequently launched, including a large case-control study of breast cancer in Long Island, NY (16), but they failed to replicate the earlier findings. The results on serum DDE level from the prospective studies, including an expanded analysis of women who had been included in the original study from New York City and a combined analysis of five prospective studies, were published between 1994 and 2001 (17–28) (Table 1). We performed a cumulative meta-analysis of the initial study and seven subsequent studies published in 1994–2000 (Figure 1). The final pooled estimate of the relative risk of breast cancer for the highest vs lowest DDE category was 0.95 (95% CI = 0.7 to 1.3).

The reason for the positive findings in the original New York City study is unknown. Chance, which is still an underappreciated problem in our field, seems likely. The short time that elapsed between blood collection and diagnosis of breast cancer may be another explanation because organochlorine compounds are stored in fat tissue, whose metabolism may be affected by the neoplastic process itself (29). Overinterpretation and an apparent lack of skepticism had major roles in the wide acceptance of the result. Although the relative risk estimate of breast cancer among women

**Table 1.** Results of prospective studies on serum 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene level and risk of breast cancer\*

Reference	Population	Period of blood collection	Follow-up, years	No. of case patients/control subjects	DDE level, reference category†	DDE level, highest category†	RR highest category (95% CI)	<i>P</i> <sub>trend</sub>
10	Volunteers, New York City	1985–1991	0.1–0.5	58/171	0.5–3.2 ppb	11.9–44.3 ppb	3.68 (1.01 to 13.5)	0.03
17	Health Maintenance Organization members, California	1964–1969	Mean 14.2	150/150	5.3–29.6 ppb	49.7–149.5 ppb	1.33 (0.68 to 2.62)	0.4
18	Volunteers, Copenhagen	1976	0.1–17.0	240/477	NA‡	NA‡	0.88 (0.56 to 1.37)	0.5
21	Volunteers, Maryland	1974	1–20	346/346	<1017 ng/g	2447–10796 ng/g	0.73 (0.40 to 13.2)	0.1
22	Volunteers, Missouri	1977–1987	Mean 9.5	105/208	31–1377 ng/g	3501–20667 ng/g	0.8 (0.4 to 1.5)	0.7
24	Volunteers, New York City	1985–1991	0.5–9.0	148/295	<664 ng/g	>1934 ng/g	1.30 (0.51 to 3.35)	1.0
25	Blood donors, Norway	1973–1991	Mean 8.8	150/150	NA§	NA§	1.2 (NA)	NA
20, 27	Nurses, United States	1989–1990	1.5–5.5	372/372	70–427 ng/g	955–1441 ng/g	0.82 (0.49 to 1.37)	0.1

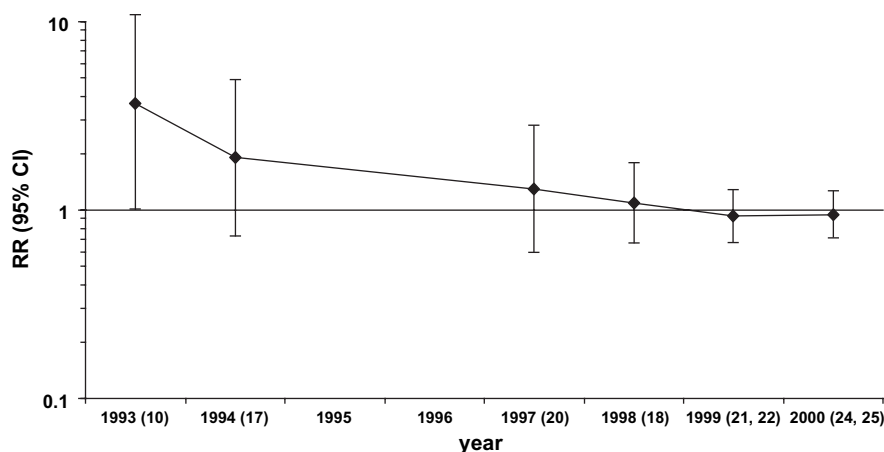
\* DDE = 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene; RR = relative risk; CI = confidence interval; HMO = health maintenance organization; NA = not available.

† DDE serum levels are either not lipid adjusted (expressed as parts per billion [ppb], equivalent to milligrams per gram serum) or lipid adjusted (expressed as nanograms per gram lipid).

‡ Mean level, control subjects: 10.2 ppb; case patients: 9.9 ppb.

§ Mean level, control subjects: 1260 ng/g; case patients: 1230 ng/g.

**Figure 1.** Cumulative meta-analysis of cohort studies of breast cancer and serum level of 1,1-dichloro-2,2-bis(*p*-chlorophenyl) ethylene (highest vs lowest category in each study). Estimated relative risks (RRs) are shown with 95% confidence intervals (CIs) (error bars) by year of publication of subsequent reports. In parentheses are references of studies included in the cumulative meta-analyses (see Table 1 for details). Upper confidence limit for the initial RR was 13.5.



in the highest quintile of serum DDE compared with the lowest was almost 4.0, the width of the confidence interval and the marginal level of statistical significance should have tempered the enthusiasm for the finding. Similarly, a recently reported relative risk estimate of 5.0 (95% CI = 1.7 to 14.8) for serum levels of *p,p'*-DDT and breast cancer risk, based on subgroup analyses in a nested case-control study (30), also warrants a cautious interpretation. What is of ongoing concern is that given the history of studies of this exposure-disease issue, the results were still interpreted with few reservations and little if any caution.

A second example relates to the results of occupational cohort studies of lung cancer risk among workers who were exposed to acrylonitrile. Acrylonitrile is a chemical used in the manufacture of acrylic fibers, resins, and nitrile rubber, as an intermediate in the chemical industry and, in the past, as a fumigant. The first study of lung cancer mortality among workers who were exposed to acrylonitrile in a textile factory in the United States, which was reported in 1978 (31), showed a fourfold increased risk, based on six deaths from lung cancer. An International Agency for Research on Cancer (IARC) working group cited this study in its review of acrylonitrile in 1979 and concluded that "while confirmatory evidence in experimental animals and humans is desirable, acrylonitrile should be regarded as if it were carcinogenic to humans" (32). Following the initial report, a total of 16 partially overlapping occupational cohort studies have provided results on lung cancer risk among acrylonitrile-exposed workers, including an expanded analysis of the first study, which identified two additional observed and 2.9 additional expected lung cancer deaths (33) (Table 2).

When we performed a cumulative meta-analysis of the initial 1978 findings and of 15 subsequent studies of acrylonitrile and lung cancer that were published in 1980-1998, we found a steady decrease over time in the reported overall relative risk estimate of lung cancer in acrylonitrile workers (Figure 2). The final pooled estimate of the relative risk of lung cancer among the workers was 1.1 (95% CI = 0.9 to 1.4). This approach admittedly represents an overly simplified review because, for the sake of comparability, we restricted the meta-analysis to historical cohort studies and did not consider case-control studies [eg, Scélo et al. (48)] that may provide additional relevant information, we did not consider the results of internal (eg, dose-response) analyses

available for several cohorts, and we did not consider the evidence for or against bias and confounding in each of the studies. However, the declining trend in the summary relative risk estimate as further data accumulated provides evidence that the initial finding was a false-positive result. In an updated review of the evidence in 1999, an IARC working group modified the classification for acrylonitrile from "2A (probable carcinogen)" to "2B (possible carcinogen)" (49).

Similar examples of false-positive results are relevant to the evidence linking exposure to established occupational carcinogens with cancer of organs that are not the primary target of their carcinogenic effect. A detailed analysis of results on other suspected occupational carcinogens is beyond the scope of this commentary, but for several agents it would show the same pattern seen for lung cancer among workers who were exposed to acrylonitrile, that is,

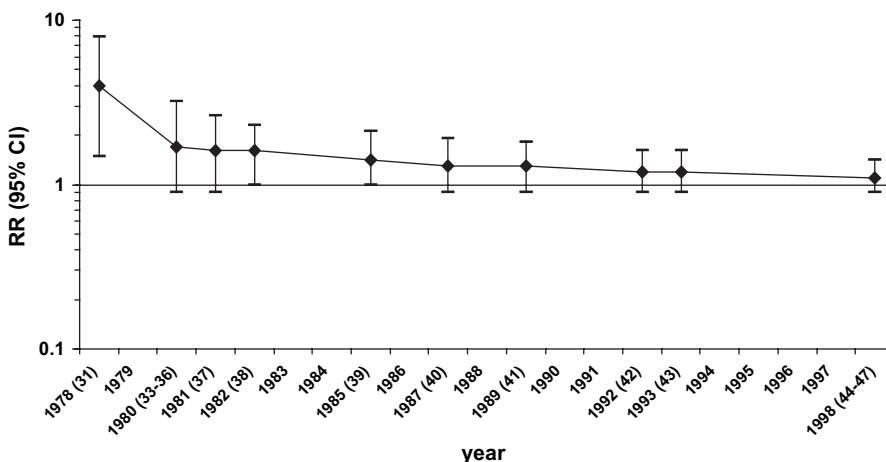
**Table 2.** Results of cohort studies on lung cancer risk among workers who were exposed to acrylonitrile\*

Reference	Overlap with previous studies,† reference	N	RR (95% CI)
31	None	6	4.0 (1.7 to 7.9)
34	None	6	0.86 (0.37 to 1.7)
33	31	2	0.69 (0.12 to 2.2)
35	None	11	2.0 (1.1 to 3.3)
36	None	1	2.0 (0.1 to 9.5)
37	None	9	1.2 (0.62 to 2.1)
38	None	9	1.5 (0.8 to 2.7)
39	31, 33	6	0.83 (0.36 to 1.6)
40	31, 33, 39	5	0.72 (0.29 to 1.5)
41	None	15	1.0 (0.62 to 1.54)
42	None	16	0.82 (0.51 to 1.2)
43	None	2	0.77 (0.14 to 2.4)
44	41	119	1.23 (1.05 to 1.43)
45	31, 33, 39, 40	27	0.69 (0.49 to 0.95)
46	42	31	1.33 (0.90 to 1.89)
47	37	44	1.0 (0.77 to 1.29)

\* N = number of observed lung cancer deaths; RR = relative risk; CI = confidence interval.

† For studies overlapping with previous reports, only the additional observed lung cancers are reported (and the relative risk is based on these additional cancers only).

**Figure 2.** Cumulative meta-analysis of cohort studies of lung cancer and occupational exposure to acrylonitrile. Estimated relative risks (RRs) are shown with 95% confidence intervals (CIs) (error bars) by year of publication of subsequent reports. In parentheses are references of studies included in the cumulative meta-analyses (see Table 2 for details).



early positive results were not confirmed, or were shown to be less strong, in subsequent studies.

### Sources of False-Positive Results

The issue of multiplicity of exposures and outcomes, and consequent multiple comparisons and subgroup analyses leading to many “statistically significant” (yet false) findings arising by chance, is likely the major contributor to false-positive findings in epidemiology. Epidemiological studies, whether cross-sectional, case-control, or cohort in design, tend to obtain information on multiple exposures and disease outcomes, so the possibility often exists for examining hundreds or thousands of exposure–disease combinations. A similar situation arises in emerging genetic epidemiological studies in which even larger numbers of traits, up to tens or hundreds of thousands of variants, can be examined. In a quantitative approach to this problem, Ioannidis et al. (8) reviewed 36 genetic disease associations and found that for 25 of them, the first report gave a stronger estimate of genetic association than did subsequent studies; in 10 of these associations, the difference between the first and the subsequent results was statistically significant. The same authors subsequently considered the results of 55 meta-analyses (50), including 579 study comparisons of genetic associations, and found that for each association the largest study generally yielded more conservative results than the meta-analysis, that in 26% of the meta-analyses the association was statistically significantly stronger in the first study, and that in only 16% was the genetic association found in the first study replicated without evidence of heterogeneity and bias. Statistical techniques aimed at reducing the likelihood of false-positive associations, such as false discovery rate (51,52), false-positive report probability (53), and Bayesian false discovery probability (54), are now included in the analysis of databases with large numbers of genetic variants [eg, Wellcome Trust Case Control Consortium (55)]. It seems appropriate that similar approaches be applied in other areas of cancer epidemiology that are prone to reporting false-positive results, including studies of potential occupational and environmental carcinogens.

There are multiple reasons besides chance why positive results, particularly initial results, in epidemiological studies may be false (2,56). To formally address the determinants of false-positive

results in occupational cancer, Swaen et al. (57) scored the main characteristics of studies reporting a positive association with established target (true positive) and nontarget (false positive) organs of 19 occupational carcinogens. The main determinants of false-positive findings were the absence of a specific a priori hypothesis, a small magnitude of the association, the absence of a dose–effect relationship, and the lack of adjustment for tobacco smoking. Although this analysis can be criticized because of subjective aspects in the definition of target organs and in quality scoring, it highlights the important roles of confounding and bias, together with chance, in the generation of false-positive results.

Although chance is a major cause of false-positive results, it is less appreciated that false leads may also be generated by bias. Bias consists of a systematic alteration of the research findings due to factors related to study design, data acquisition, analysis, or reporting of results. There is a fundamental distinction between false-positive results that are generated by chance and those caused by bias—the former will rarely be replicated in subsequent investigations, whereas bias may operate in a similar fashion in different settings and populations and thus will provide a consistent pattern of independently generated results. Even if only a relatively low proportion of results are generated by bias, the probability of false-positive discovery may be substantial.

Information bias, or recall bias, is a likely source of false-positive findings. Early studies examining a possible association between induced abortion and breast cancer risk that were based on retrospectively ascertained information generated evidence indicating that women who had undergone an induced abortion were at increased risk of breast cancer (58). A collaborative reanalysis of data from 53 epidemiological studies comparing women with prospectively ascertained records of one or more induced abortions vs women with no such record (59) reported a relative risk of breast cancer equal to 0.93 (95% CI = 0.89 to 0.96). It is now apparent that information bias had a major role in generating the early false-positive results. Control women were less likely to report their previous induced abortions than were women with breast cancer (58,60). In general, patients have greater motivation (eg, to explain their disease) and thus are more likely to remember and/or report certain past exposures than are healthy individuals who are selected as control subjects.

Another source of false-positive findings is selection bias. A likely example of selection bias is the 1981 report of increased pancreatic cancer risk associated with coffee consumption in one of the earliest epidemiological studies of this disease (61). Many investigators aimed to replicate these findings, but the results of subsequent studies were in general null and by the end of the 1980s a causal association between coffee drinking and pancreatic cancer was considered unlikely (62,63). Yet for several years the suspicion that coffee drinking was associated with a highly lethal cancer was widespread in the medical community and the general public. This false-positive result was generated, at least in part, by exclusions from the control patient population, but not from the case patients, of individuals with a history of diseases related to cigarette smoking and alcohol consumption; because these exposures were highly correlated with coffee consumption, the exclusions likely led to a deficit of coffee consumers in the control group.

Another source of bias and false-positive findings in observational studies is confounding resulting either from incomplete statistical adjustment for measured variables or from the inability to adjust for unmeasured distorting variables. For example, results of early epidemiological studies showed an increased risk of cervical cancer among women who were infected with agents such as herpes simplex virus 2 and chlamydia. When sensitive and specific markers of infection with human papilloma virus became available, however, it was clear that the early positive results for the other sexually transmitted agents were due to confounding by human papilloma virus (64). Although the importance of residual confounding and unmeasured confounders as a source of bias in epidemiological studies has been downplayed by many (65), a recent statistical simulation study (66) showed that with plausible assumptions, effect sizes on the order of 1.5–2.0, which is a magnitude frequently reported in epidemiology studies, can be generated by residual and/or unmeasured confounding. The real effect of unmeasured confounders is of course unknown but may explain some of the notorious differences between well-known cohort studies and subsequent randomized trials.

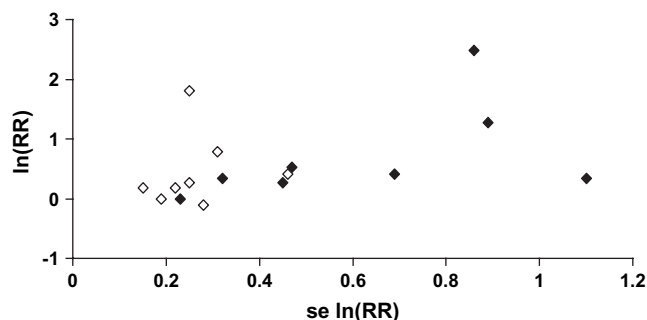
The randomized controlled trial is often presented as the gold standard of epidemiological studies, with the random assignment of treatments mitigating against the potential influences of bias and confounding. Systematic reviews have found evidence of higher relative risks in observational studies as compared with randomized controlled trials addressing the same question (67,68). However, even randomized controlled studies are not exempt from reporting of false-positive results that arise by chance. In a review of 39 highly cited (defined as studies cited 1000 or more times in the literature) randomized controlled trials that reported an original claim of an effect (69), the results of 19 were replicated by subsequent studies, whereas for nine trials subsequent results either contradicted the findings of the original report or provided evidence of a weaker effect (the remaining 11 trials did not have subsequent attempts of replication). Small size of the original study was the strongest predictor of subsequent refutation. Premature conclusions can be avoided by cautious interpretation of initial trials, especially when they are small, and by trial registration, which is a requirement of the International Committee of Medical Journal Editors (70). Moreover, there is an unfortunate tendency to highlight “positive” or “statistically significant” findings in the

abstracts of both observational studies and randomized trials, even when the results are doubtful or open to criticism (71).

## Publication Bias

Another cause of false-positive reporting in cancer epidemiology is publication or reporting bias. It originates from the tendency of authors and journal reviewers and editors to report and publish “positive” or “statistically significant” results over “null” or “non-statistically significant” results, particularly if the findings appear to confirm a previously reported association (ie, the “bandwagon effect”). As with other forms of bias, preferential publication generates a false sense of consistency among studies. Statistical tests have been proposed to assess the presence of publication bias (72–75); unfortunately, these tests tend to have low power and require a large number of independent studies to provide evidence of bias. There are, however, suspected occupational and environmental carcinogens for which a sufficiently large number of studies are available to allow testing for publication bias.

One group of suspected environmental carcinogens is the polychlorinated dibenzo-para-dioxins, which are by-products of the manufacture of chlorophenols and chlorophenoxy herbicides; they are also produced during incineration of organic material and in other thermal processes, such as metal processing and paper pulp bleaching. The most toxic is 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (2,3,7,8-TCDD), which has been found to cause tumors in rodents under certain conditions (62). Concern about the carcinogenicity of dioxins in general, and 2,3,7,8-TCDD in particular, has led to epidemiological studies that are both occupationally based (eg, workers in the chemical industry, pesticide applicators) and community based (eg, residents in contaminated areas, the general population). Non-Hodgkin lymphoma (NHL) has been among the suspected targets of dioxin carcinogenicity in humans. At the time of the most recent IARC Monograph evaluation (76), results on NHL risk were available from 16 studies of 2,3,7,8-TCDD-exposed populations—eight cohort studies and eight case-control studies (76). Although a meta-analysis suggests an increased risk of NHL in these populations (meta-relative risk = 1.6, 95% CI = 1.2 to 2.1), there is evidence of publication bias [Figure 3;  $P = .02$ , Begg test (75)]. In the absence of publication bias, the results of individual studies plotted against their precision, as in Figure 3, should be distributed in a triangular (“funnel”) pattern, with less



**Figure 3.** Funnel plot of results of studies on dioxin exposure and risk of non-Hodgkin lymphoma. **Closed diamonds**, cohort studies; **empty diamonds**, case-control studies. RR = relative risk. se = standard error. See (76) for detailed results.

precise studies (on the right side of the graph) randomly dispersed around the mean. Publication bias is suggested by the absence of results in one of the parts of the “funnel” (eg, lower right corner in Figure 3, where small null studies would be plotted). Formal testing of publication bias is helpful, but keeping the bandwagon effect concept in mind would go a long way to instilling skepticism when assessing the strength of evidence provided by a new report on a popular hypothesis.

## Caveats and Conclusions

The examples presented in this commentary suggest that false-positive results are a common problem in cancer and other types of epidemiological studies. What can be done within the practice of epidemiology to reduce the problem? One of the simplest yet potentially most effective remedies involves increasing emphasis on skepticism when assessing study results, particularly when they are new. Put another way, epidemiologists should practice some epistemological modesty when interpreting and presenting their findings. Epidemiologists by training are most often aware of the methodological limitations of observational studies, particularly those done by others, yet when it comes to practice, and especially the interpretation of their own study results, methodological vigilance is often absent. This absence of skepticism in reporting results in published papers increases the likelihood that a positive finding will receive unwarranted attention in the media and by the public.

The tendency to emphasize and overinterpret what appear to be new findings is commonplace, perhaps in part because of a belief that the findings provide information that may ultimately improve public health. Although the results may turn out to be wrong, some authors feel it is better to err on the side of overreporting or overstating potentially false-positive results than to miss the identification of a potential new hazard or an opportunity for career advancement. This tendency to hype new findings increases the likelihood of downplaying inconsistencies within the data or any lack of concordance with other sources of evidence. Furthermore, the clear acknowledgement that the statistically significant findings may arise from multiple comparisons or subgroup analyses is often missing; results from a single study are often dispersed across multiple publications, sometimes without reference to each other, hindering the detection of the multiple comparisons or subgroup reporting.

Strict adherence to the highest epidemiological standards in the design, analysis, reporting, and interpretation of studies would help reduce the likelihood and impact of false-positive results. These standards include provisions to reduce the opportunity for bias and confounding in study design, adequate statistical power, avoidance (or at least cautious interpretation) of data-driven subgroup analyses, and accounting for multiple comparisons. The strategy for reporting study results should be specified before the results are known, and selective reporting or emphasis of statistically significant results based on ex post facto subgroup analyses should be discouraged. The interpretation of positive results, especially if they are not supported by additional evidence (eg, other epidemiological or experimental studies), should be careful and cautious. These considerations are particularly important in the summary and abstract of published papers. They are also in line with many of the recommendations in the Strengthening the Reporting of Observational

studies in Epidemiology (STROBE) statement, which is aimed at strengthening the reporting of epidemiological studies (77,78). We propose that the policies of some journals, which require the explicit acknowledgement of study limitations up front in the abstract or in a note box, become standard practice. Caution should be applied in the communication of results to the media and the general public, because “positive” findings tend to attract the media and public attention, whereas findings that do not confirm a previously reported association or do not indicate a new association often receive no attention. Consequently, false-positive results tend to survive longer and have larger long-term consequences in the general public than in the scientific community. Likewise, users of epidemiological results outside the scientific community (eg, regulatory agencies, stakeholders, media, advocacy groups, trial lawyers, the general public) should be aware of the fact that statistically significant or positive results are often false, in particular when they are not supported by related studies or other lines of evidence.

The examples we have presented in this commentary were restricted largely to the field of cancer epidemiology. However, we believe that similar problems affect other areas of epidemiological research, in particular those involving studies with large sets of “determinant” and “outcome” variables. A careful analysis of factors enhancing the likelihood of false-positive reporting should be part of the training of epidemiologists, and these issues should be addressed in the “Discussion” section of every epidemiological paper. In general, epidemiologists should recognize the reporting of false-positive results as a major challenge to the scientific credibility of their discipline and institutionalize a greater level of epistemological modesty in this regard (4,79).

## References

1. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet*. 2003;361(9360):865–872.
2. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
3. Schottenfeld D, Fraumeni JF Jr (eds). *Cancer Epidemiology and Prevention*. 3rd ed. New York, NY: Oxford University Press; 2006.
4. Taubes G. Epidemiology faces its limits. *Science*. 1995;269(5221):164–169.
5. Pocock S, Collier T, Dandreo K, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. 2004;329(7471):883.
6. Freely associating [editorial]. *Nat Genet*. 1999;22(1):1–2.
7. Dahlman I, Eaves I, Kosoy R, Morrison V, Heward J, Gough S. Parameters for reliable results in genetic association studies in common disease. *Nat Genet*. 2002;30(2):149–150.
8. Ioannidis J, Ntzani E, Trikalinos T, Contopoulos-Ioannidis D. Replication validity of genetic association studies. *Nat Genet*. 2001;29(3):306–309.
9. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–188.
10. Wolff M, Toniolo P, Lee E, Rivera M, Dubin N. Blood levels of organochlorine residues and risk of breast cancer. *J Natl Cancer Inst*. 1993;85(8):648–652.
11. Wassermann M, Nogueira DP, Tomatis L, et al. Organochlorine compounds in neoplastic and adjacent apparently normal breast tissue. *Bull Environ Contam Toxicol*. 1976;15(4):478–84.
12. Unger M, Kiaer H, Blichert-Toft M, Olsen J, Clausen J. Organochlorine compounds in human breast fat from deceased with and without breast cancer and in a biopsy material from newly diagnosed patients undergoing breast surgery. *Environ Res*. 1984;34(1):24–28.
13. Mussalo-Rauhamaa H, Hasanen E, Pyysalo H, Antervo K, Kauppila R, Pantzar P. Occurrence of  $\beta$ -hexa-chlorocyclohexane in breast cancer patients. *Cancer*. 1990;66(10):2124–2128.

14. Falck F, Ricci A, Wolff MS, Godbold J, Deckers P. Pesticides and polychlorinated biphenyl residues in human breast lipids and their relation to breast cancer. *Arch Environ Health*. 1992;47(2):143–146.
15. Hunter D, Kelsey K. Pesticide residues and breast cancer: the harvest of a silent spring? *J Natl Cancer Inst*. 1993;85(8):598–599.
16. Gammon MD, Wolff MS, Neugut AI, et al. Environmental toxins and breast cancer on Long Island II. Organochlorine compound levels in blood. *Cancer Epidemiol Biomarkers Prev*. 2002;11(8):686–697.
17. Krieger N, Wolff M, Hiatt R, Rivera M, Vogelmann J, Orentreich N. Breast cancer and serum organochlorines: a prospective study among White, Black, and Asian women. *J Natl Cancer Inst*. 1994;86(8):589–599.
18. Høyer AP, Granjean P, Jørgensen T, Brock JW, Hartvig HB. Organochlorine exposure and risk of breast cancer. *Lancet*. 1998;352(9143):1816–1820.
19. Moysich K, Ambrosone C, Vena J, et al. Environmental organochlorine exposure and postmenopausal breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 1998;7(3):181–188.
20. Hunter D, Hankinson S, Laden F, et al. Plasma organochlorine levels and the risk of breast cancer. *N Engl J Med*. 1997;337(18):1253–1258.
21. Helzlsouer K, Alberg A, Huang HY, et al. Serum concentrations of organochlorine compounds and the subsequent development of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 1999;8(6):525–532.
22. Dorgan JF, Brock JW, Rothman N, et al. Serum organochlorine pesticides and PCBs and breast cancer risk: results from a prospective analysis (USA). *Cancer Causes Control*. 1999;10(1):1–11.
23. Zheng T, Holford TR, Mayne S, et al. Risk of female breast cancer associated with serum polychlorinated biphenyls and 1,1-dichloro-2,2'-bis(p-chlorophenyl)ethylene. *Cancer Epidemiol Biomarkers Prev*. 2000;9(2):167–174.
24. Wolff M, Zeleniuch-Jacquette A, Dubin N, Toniolo P. Risk of breast cancer and organochlorine exposure. *Cancer Epidemiol Biomarkers Prev*. 2000;9(3):271–277.
25. Ward E, Schulte P, Grajewski B, et al. Serum organochlorine levels and breast cancer: A nested case-control study of Norwegian women. *Cancer Epidemiol Biomarkers Prev*. 2000;9(12):1357–1367.
26. Wolff MS, Berkowitz GS, Brower S, et al. Organochlorine exposures and breast cancer risk in New York City women. *Environ Res*. 2000;84(2):151–161.
27. Laden F, Hankinson S, Wolff M, et al. Plasma organochlorine levels and the risk of breast cancer: an extended follow-up in the Nurses' Health Study. *Int J Cancer*. 2001;91(4):568–574.
28. Laden F, Collman G, Iwamoto K, et al. 1,1-Dichloro-2,2-bis(p-chlorophenyl)ethylene and polychlorinated biphenyls and breast cancer: combined analysis of five US studies. *J Natl Cancer Inst*. 2001;93(10):768–776.
29. López-Cervantes M, Torres-Sánchez L, Tobías A, Lopez-Carillo L. Dichlorodiphenyldichloroethane burden and breast cancer risk: a meta-analysis of the epidemiological evidence. *Environ Health Perspect*. 2004;112(2):207–214.
30. Cohn BA, Wolff MS, Cirillo PM, Sholtz RI. DDT and breast cancer in young women: new data on the significance of age at exposure. *Environ Health Perspect*. 2007;115(10):1406–1414.
31. United States Department of Labor. Occupational exposure to acrylonitrile (vinyl cyanide). Final standard. *Fed Regist*. 1978;43:45762–45819.
32. International Agency for Research on Cancer. Acrylonitrile. In: *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 19. Some Monomers, Plastics and Synthetic Elastomers, and Acrolein*. Lyon, France: IARC; 1979:73–114.
33. O'Berg MT. Epidemiologic study of workers exposed to acrylonitrile. *J Occup Med*. 1980;22(4):245–252.
34. Keisselbach N, Korallus U, Lange HJ, Neills A, Zwingers T. Acrylonitrile—epidemiological study—Bayer 1977. *Zentralbl Arbeitsmed*. 1979;29(10):257–259.
35. Thiess AM, Frentzel-Beyme R, Link R, Wild H. Mortality study in chemical personnel of various industries exposed to acrylonitrile. *Zentralbl Arbeitsmed Arbeitsschutz*. 1980;30(7):259–267.
36. Ott MG, Kolesar RC, Scharnweber HC, Schneider EJ, Venable IR. A mortality survey of employees engaged in the development or manufacture of styrene-based products. *J Occup Med*. 1980;22(7):445–460.
37. Werner JB, Carter JT. Mortality of United Kingdom acrylonitrile polymerization workers. *Br J Ind Med*. 1981;38(3):247–253.
38. Delzell E, Monson RR. Mortality among rubber workers: VI. Men with potential exposure to acrylonitrile. *J Occup Med*. 1982;24(10):767–769.
39. O'Berg MT, Chen JL, Burke CA, Walrath J, Pell S. Epidemiologic study of workers exposed to acrylonitrile: an update. *J Occup Med*. 1985;27(11):835–840.
40. Chen JL, Walrath J, O'Berg MT, Burke CA, Pell S. Cancer incidence and mortality among workers exposed to acrylonitrile. *Am J Ind Med*. 1987;11(2):157–163.
41. Collins JJ, Page LC, Caporossi JC, Utidjian HM, Lucas LJ. Mortality patterns among employees exposed to acrylonitrile. *J Occup Med*. 1989;31(4):368–371.
42. Swaen GMH, Bloemen LNJ, Twisk J, Scheffers T, Slangen JJM, Sturmans F. Mortality of workers exposed to acrylonitrile. *J Occup Med*. 1992;34(8):801–809.
43. Mastrangelo G, Serena R, Marzia V. Mortality from tumours in workers in an acrylic fibre factory. *Occup Med (Lond)*. 1993;43(3):155–158.
44. Blair A, Stewart PA, Zaubst DD, et al. Mortality study of industrial workers exposed to acrylonitrile. *Scand J Work Environ Health*. 1998;24(suppl 2):25–41.
45. Wood SM, Buffer PA, Burau K, Krivanek N. Mortality and morbidity experience of workers exposed to acrylonitrile in fiber production. *Scand J Work Environ Health*. 1998;24(suppl 2):54–62.
46. Swaen GMH, Bloemen LNJ, Twisk J, et al. Mortality update of workers exposed to acrylonitrile in the Netherlands. *Scand J Work Environ Health*. 1998;24(suppl 2):10–16.
47. Benn T, Osborne K. Mortality of United Kingdom acrylonitrile workers—an extended and updated study. *Scand J Work Environ Health*. 1998;24(suppl 2):17–24.
48. Scélo G, Constantinescu V, Csiki I, et al. Occupational exposure to vinyl chloride, acrylonitrile and styrene and lung cancer risk (Europe). *Cancer Causes Control*. 2004;15(5):445–452.
49. International Agency for Research on Cancer. Acrylonitrile. In: *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 71. Re-evaluation of Some Organic Chemicals, Hydrazine and Hydrogen Peroxide (part one)*. Lyon, France: IARC; 1999:43–108.
50. Ioannidis J, Trikalinos T, Ntzani E, Contopoulos-Ioannidis D. Genetic associations in large versus small studies: an empirical assessment. *Lancet*. 2003;361(9357):567–571.
51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289–300.
52. Benjamini Y, Yekutieli D. Quantitative trait loci analysis using the false discovery rate. *Genetics*. 2005;171(2):783–790.
53. Wacholder S, Chanock S, Garcia-Closas M, El-Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96(6):434–442.
54. Wakefield J. A bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet*. 2007;81(2):208–227.
55. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
56. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol*. 2005;58(6):543–549.
57. Swaen G, Teggeler O, van Amelsvoort L. False positive outcomes and design characteristics in occupational cancer epidemiology studies. *Int J Epidemiol*. 2001;30(5):948–954.
58. Rookus MA, van Leeuwen FE. Induced abortion and risk for breast cancer: reporting (recall) bias in a Dutch case-control study. *J Natl Cancer Inst*. 1996;88(23):1759–1764.
59. Beral V, Bull D, Doll R, Peto R, Reeves G; Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and abortion: collaborative reanalysis of data from 53 epidemiological studies, including 83,000 women with breast cancer from 16 countries. *Lancet*. 2004;363(9414):1007–1016.
60. Lindefors-Harris BM, Eklund G, Adami HO, Meirik O. Response bias in a case-control study: analysis utilizing comparative data concerning legal

- abortions from two independent Swedish studies. *Am J Epidemiol.* 1991; 134(9):1003–1008.
61. MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med.* 1981;304(11):630–633.
  62. Boyle P, Hsieh C, Maisonneuve P, et al. Epidemiology of pancreas cancer. *Int J Pancreatol.* 1989;5(4):327–346.
  63. International Agency for Research on Cancer. Coffee. In: *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 51. Coffee, Tea, Mate, Methylxanthines and Methylglyoxal.* Lyon, France: IARC; 1991:41–206.
  64. Schiffman MH, Brinton LA. The epidemiology of cervical carcinogenesis. *Cancer.* 1995;76(Suppl 10):1888–1901.
  65. Blair A, Stewart P, Lubin JH, Forastiere F. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *Am J Ind Med.* 2007;50(3):199–207.
  66. Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol.* 2007;166(6):646–655.
  67. Benson K, Hartz A. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342(25):1878–1886.
  68. Ioannidis J, Haidich A, Pappa M, et al. Comparison of evidence of treatment effects in randomized and non randomized studies. *JAMA.* 2001; 286(7):821–830.
  69. Ioannidis J. Contradicted and initially stronger effects in highly cited clinical research. *JAMA.* 2005;294(2):218–228.
  70. De Angelis C, Drazen J, Frizelle F, et al. Is this clinical trial fully registered? A statement from the international committee of medical journal. *Lancet.* 2005;365(9474):1827–1829.
  71. Gotzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ.* 2006;333(7561):231–234.
  72. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol.* 2000;53(11):1119–1129.
  73. Henmi M, Copas JB, Eguchi S. Confidence intervals and P-values for meta-analysis with publication bias. *Biometrics.* 2007;63(2):475–482.
  74. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials.* 2007;4(3):245–253.
  75. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics.* 1994;50(4):1088–1101.
  76. International Agency for Research on Cancer. Polychlorinated dibenzo-para-Dioxins. In: *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 69. Polychlorinated Dibenzo-para-Dioxins and Polychlorinated Dibenzofurans.* Lyon, France: IARC; 1997:33–342.
  77. von Elm E, Altman DG, Egger M, et al. The Strengthening of the Reporting of Observational Studies in Epidemiology (STROBE) Statement—guidelines for reporting observational studies. *Epidemiology.* 2007;18(6):800–804.
  78. Vandenberg JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)—explanation and elaboration. *Epidemiology.* 2007;18(6):805–835.
  79. Trichopoulos D. The future of epidemiology. *BMJ.* 1996;313(7055): 436–437.

### Notes

The authors received no external funding for this Commentary.

Manuscript received March 3, 2008; revised April 24, 2008; accepted May 16, 2008.