

## Research Article

## From Testing to Estimation: The Problem of False Positives in the Context of Carcinogen Evaluation in the IARC Monographs

Christiana A. Demetriou<sup>1,2</sup>, Kurt Straif<sup>3</sup>, and Paolo Vineis<sup>1</sup>

## Abstract

**Background:** Cancer epidemiology has been criticized for producing false-positive associations. The present analysis investigates the frequency of and factors contributing to false-positive findings in cancer epidemiology.

**Methods:** The International Agency for Research on Cancer (IARC) Monographs Group 3 agents were examined to identify potential false-positive findings. Frequency estimates for their occurrence were calculated. Comments of the Working Groups on study quality were recorded for studies with potential false-positives. These were used to determine how many of such studies were criticized for each of the study quality factors that are suspected to contribute to false-positive results.

**Results:** Of 509 agents in group 3, 37 agents were found to have potential false-positive associations in the studies reviewed in their respective IARC monograph(s). The overall frequency of potential false-positives among these agents was between 0.03 and 0.10. The individual frequencies ranged from 0.01 to 0.40. The potential false-positive findings were produced by 162 studies. The most common factors contributing to potential false-positive findings were confounding and exposure misclassification.

**Conclusions:** The frequency estimates we have obtained do not suggest that epidemiology is grossly flooded by false-positive findings. The factors for which studies with potential false-positives were most often criticized were factors that are sometimes difficult to address in cancer epidemiologic research and can bias an effect estimate toward or away from the null.

**Impact:** The low frequency of false-positives in cancer epidemiology restores faith in epidemiologic procedures, making epidemiologic findings a useful guide for public health care measures. *Cancer Epidemiol Biomarkers Prev*; 21(8); 1272–81. ©2012 AACR.

## Introduction

Public health care relies heavily on the findings of epidemiologic research. Epidemiology seeks to answer public health's relevant questions with the goal of informing policy on primary and secondary prevention, as well as evaluating outcomes.

Epidemiology and more specifically cancer epidemiology, however, has been criticized for not practicing epistemological modesty (1, 2). Epidemiologists are accused of

overemphasizing and overinterpreting newly discovered positive associations, with little consideration for the limitations of their research or for the limits of epidemiology in general. The result is allegedly overemphasizing new associations that fail to be corroborated by subsequent research. These results are considered "false positive" findings.

We believe that the issue of "false positives" is partially misleading. Epidemiologic research, like all scientific research, is a continuous process rather than being characterized by decisions based on dichotomies (positive/negative). The publication of findings stimulates both a critical assessment of the overall published evidence [typically, in International Agency for Research on Cancer (IARC) Monographs Working Groups] and the conduction of new studies to corroborate or undermine previous observations. Only limited understanding of how epidemiology works leads to emphasize the issue of false-positives.

To exemplify that the question on false-positives is misleading, we have examined the associations between chemical exposures and cancer for chemicals (or circumstances of exposure) classified as "inadequate" in the IARC

**Authors' Affiliations:** <sup>1</sup>Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom; <sup>2</sup>Department of Electron Microscopy and Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia Cyprus; and <sup>3</sup>IARC Monographs Section, International Agency for Research on Cancer, Lyon, France

**Note:** P. Vineis and K. Straif contributed equally to the work.

**Corresponding Author:** Kurt Straif, International Agency for Research on Cancer, Lyon France, 150 Cours Albert Thomas, 69372 Lyon Cedex 08, France. Phone: 33-04-72738485; Fax: 33-04-72738575; E-mail: [straif@iarc.fr](mailto:straif@iarc.fr)

**doi:** 10.1158/1055-9965.EPI-12-0276

©2012 American Association for Cancer Research.

monographs, and we have reconstructed the history of such associations by looking at the whole history of relevant publications. As a by-product of this exercise, we have also estimated the prevalence of "potential false positives" (PFP).

Factors hypothesized to contribute to "false positive" results in epidemiologic studies include study quality concerns such as multiple comparisons and subgroup analyses, information and recall bias, selection bias, confounding, small study sizes and issues with design, definitions, outcomes, and methods of analysis (1, 3–5). Ioannidis argues that most published research findings are false and states that the positive predictive value of a finding—its post-study probability of being true—depends on the above as well as other factors, not necessarily relevant to study quality. These additional factors include a finding's small probability of being true before the study, small effect sizes, large number of independent studies, financial influences and prejudices, and a hot scientific field (3). Hernberg also addresses false-positive findings in his book and states that the most common reasons for their occurrence are information bias and confounding (6). Finally, most researchers agree that publication bias, the increased likelihood of positive results to be published, may increase the occurrence of false-positives in the literature (1, 5, 7, 8). All the factors above are common and sometimes inevitable in the field of cancer epidemiology. Thus, one might expect the frequency of false-positives in this field to be high.

On the other hand, there are many factors that counteract a high frequency of false-positive results. To begin with, study quality limitations, such as information bias, selection bias, and confounding, can also bias an association toward the null. The result would be attenuation in the frequency of false-positive results and even an increased frequency of false-negative results. Furthermore, in occupational epidemiology, publication bias may be reversed. For example, some suggest that there are incentives for industry-funded epidemiologists to keep positive studies unpublished (9). If the argument is true, then it is difficult to conclude whether, in occupational cancer epidemiology in particular, the presence of publication bias would increase the frequency of false-positive results. Given these forces, acting to increase the number of false-negative results, one cannot predict with certainty the actual frequency of false-positive results in cancer epidemiology.

In addition to the factors mentioned above, the frequency of false-positives in epidemiology will also depend on the technical definition of "false-positives" used (10). With the exception of the unambiguous associations, "unfortunately, there is no 100% certainty or gold standard telling us which result is a false-positive and which a true positive" (8). Therefore, the answer concerning the frequency of occurrence of false-positive results depends on the definition applied.

Despite the concerns and debate, there have been few attempts to estimate the frequency of false-positives in

nonexperimental epidemiologic research (1). Also, empirical data are lacking to support the hypotheses on which factors contribute to the generation of false-positives results (8). However, understanding the frequency and causes of false-positives is important to understand more about the caveats of and improve the epidemiologic research process.

Since 1971, the IARC monographs, the most comprehensive and respected collection of systematically evaluated agents in the field of cancer epidemiology, have evaluated (and reevaluated) more than 900 different agents (chemicals and mixtures, biologic and physical agents, occupational and environmental exposure, and circumstances and personal habits). Agents are selected on the basis of documented human exposure and some data suggesting a carcinogenic hazard. This provides a large database for the evaluation of the frequency of and the factors contributing to false-positive findings.

IARC, through its Monographs Programme, seeks to identify the causes of cancer in humans. A large number of agents to which humans are exposed and which have been suspected to be carcinogenic are reviewed by Working Groups to assess their causal relationship to cancer. For each agent, published epidemiologic studies, cancer bioassays in animals and other relevant data pertaining to the carcinogenicity are reviewed and the Working Group uses criteria such as those proposed by Hill (11) to determine whether a causal association between exposure and cancer exists. On the basis of the review of the available epidemiologic studies, the evidence on cancer in humans is finally categorized as "sufficient," "limited," "inadequate," or "evidence suggesting lack of carcinogenicity" (12).

Because of the strength of the evaluation procedures followed by IARC, the monographs have a very strong positive reputation and the consensus reached by Working Groups is regarded as authoritative. The present analysis was designed to investigate the frequency of and the factors contributing to PFP findings in cancer epidemiology, using the evaluations in the IARC monographs.

## Materials and Methods

The IARC monographs evaluate the strength of evidence on cancer in humans about chemical, physical, and biologic exposures, mixtures, personal habits, and other exposure scenarios, collectively referred to as agents. The analysis consisted of a quantitative and a qualitative component. The quantitative component identified PFP findings in the IARC Monographs of Group 3 agents and estimated their frequency of occurrence. The qualitative component examined the factors contributing to the generation of PFP findings and investigated whether these are consistent with the factors suggested in the literature.

To identify PFPs, agents that have been evaluated as "probably not carcinogenic to humans (group 4)" had to

be investigated. However, as classification in this group requires evidence suggesting lack of carcinogenicity in humans and in animals, only one agent is included in group 4 and there was no relevant data on human cancer. Therefore, group 3 agents were reviewed instead.

Group 3 agents are agents that are not classifiable with regard to their carcinogenicity and for which typically data on cancer in experimental animals were less than "sufficient" and data on cancer in humans were not available to the Working Group or were evaluated as "inadequate", that is, the reviewed studies lacked consistency or statistical power to permit a conclusion about their carcinogenicity (12). Because the consensus of the Working Group on the totality of the evidence did not support an association, statistically significant positive findings within studies reviewed in the IARC Monographs of Group 3 agents will be regarded as PFP in this analysis, with the caveats reported below.

#### **Quantitative component: operational definition of "false positive"**

The most up-to-date list of agents classified as group 3 by IARC up to and including monograph volume 103 was used. Volume 100 was excluded as this was focused on group 1 carcinogens.

For agents evaluated in more than one volume of the monographs, the most recent summary of evaluations and the sum of the studies included in all monographs were considered.

With the limitations described above about the concept of "false positive" (that does not consider evidence as a cumulative process), we have established an operational definition. For the purposes of this analysis, we have used the *P* value to dichotomize "positive" from "negative" in the first choice of the exposures to be analyzed. The level of statistical significance required to classify a finding as positive was therefore the conventional *P* value of less than 0.05 or a 95% confidence interval that excluded the null (13). When *P* values or confidence intervals were not reported, the study was not regarded as positive with statistical significance, unless the IARC Working Group had stated that the effect measure of the association was statistically significant. Therefore, in this screening, a maximum sensitivity criterion was applied to identify positive findings.

Once agents for which there has been at least one statistically significant positive study in the corresponding monograph were identified, all the relevant studies in that monograph were considered in detail to identify and quantify PFP findings.

Different sites and types of cancer were considered as a separate outcome and each exposure–outcome combination was considered independently. Therefore, for each exposure, findings were analyzed by cancer site to identify any PFPs. For a finding to be considered as a PFP and be included in the numerator of the frequency estimate, the following criteria had to be fulfilled: (i) one study

(index study) must have produced a statistically significant positive finding between the agent under study and the specific cancer site and (ii) there must have been at least one other study that found a measure of association whose 95% confidence interval included 1.0, or a statistically significant negative effect estimate [relative risk (RR) lower than 1]. This study must have assessed the same exposure and the same specific cancer site as the index study. The denominator for the frequency estimate was the sum of the study results for all cancer sites, explicitly mentioned by the Working Group in the monograph of each agent. Therefore, the frequency estimate for each agent was calculated by summing up over all cancer sites the number of associations which were labeled as PFP in the studies reviewed in that agent's monograph(s), divided by the total number of associations explicitly recorded in that agent's monograph(s) by the Working Group. It is important to note that those agents without at least one statistically significant positive finding were not included in the analysis.

Where associations between exposure and cancer site were analyzed by a third variable, such as category of exposure or subset of the study population, each association was considered as a separate unit of observation. Even if only a mid-level exposure category was statistically significantly positive, it was qualified as a PFP. In addition, where both adjusted and unadjusted estimates were presented for the same association, only the adjusted estimates were used.

#### **Qualitative component**

To identify the possible causes of false-positive findings, an analysis was carried out by study. This is because under the IARC procedures, the Working Groups recorded those aspects of each reviewed study which compromised its quality, and these criticisms concerned the study's methodology in general and not necessarily individual associations in each study.

For each agent, the studies in its monograph that produced at least one PFP exposure–outcome association, as identified in the quantification stage, were selected.

For each of these studies, the criticisms expressed by the Working Group, in square brackets, on the study's quality were recorded as the possible causes for the PFP results. This information was reviewed to determine how many of these studies were criticized for each of the study quality factors that are mentioned in the literature as contributing to the generation of false-positive results (1, 3, 5). More specifically, the study quality factors investigated were multiple comparisons, subgroup analyses, information and recall bias, selection bias, confounding, small study size, and questions in design, definitions, outcomes, and analytic modes.

The frequency of each of these criticisms among studies with PFP results was calculated to determine which ones are more common and whether these criticisms are consistent with the concerns in the literature.

## Results

### Quantitative component

The list of group 3 agents, which was obtained from the official IARC website, was composed of 509 chemical, physical, and biologic agents, mixtures, and exposure circumstances. Only for 127 of these agents was data on carcinogenicity to humans available to the Working Groups. Sixty-eight of the group 3 agents had at least one statistically significant positive association reported in the corresponding monograph, satisfying our first criterion. After examination of the 68 relevant monographs, only for 37 agents was there a second study, thus satisfying our second criterion for PFP findings (Table 1).

Table 2 also lists the 37 agents (14–66) with at least one PFP finding and shows (i) the number of studies that reported PFP findings for each agent and (ii) the number of PFP associations for each agent. It is important to note that these are the associations which were labeled as PFP according to the criteria and definition used in this analysis, and they are independent of the evaluation of carcinogenicity proposed by the IARC Working Groups. The number of identified PFP associations for each agent varied from one association to 22 associations per agent.

Table 2 lists the overall and agent-specific frequency estimates for PFP findings. The overall frequency of occurrence of PFP findings among the 37 agents was calculated to be 0.10. The numerator for this frequency estimate was the total number of findings that were labeled as PFPs, based on the criteria used in this project. The conservative denominator used for Table 2 was the total number of associations explicitly reported in all the studies reviewed in the respective IARC monographs for the 37 agents with at least one PFP finding. On the basis of this estimate, about 1 of 10 associations examined by the Working Groups for these 37 monographs were PFP, according to the definition and criteria used in this analysis. The frequency of PFP findings varied for individual agents and ranged from 0.01 for spironolactone to 0.40 for leather goods manufacture. In other words, for the former, only 1 in a 100 associations reported in the literature was deemed to be a PFP, whereas for the latter, the frequency was 4 of 10.

Using an alternative definition for the denominator considerably reduced the overall PFP frequency. As stated, 127 agents had data on cancer in humans, 68 agents

had at least one statistically significant association in the corresponding monograph, and 37 of these agents had some PFP associations. Thus, 90 agents ( $127 - 37$ ) had associations reported in their monographs, but none of these associations was labeled as PFP according to the criteria used. If all the associations in all the 127 monographs with data on cancer in humans were counted, the overall denominator would approximately increase by a factor of  $127/37$ , that is, 3.4, the numerator would remain the same, and therefore the average frequency estimate of PFP findings among group 3 agents would be reduced to 0.03. If the denominator included all associations with the exception of associations in those 31 monographs, which reported at least one statistically significant association without satisfying the second criterion, the frequency estimate would be 0.04 after applying a correction factor of  $(127 - 31)/37$  to the denominator.

### Qualitative component

The identified PFP findings were produced by 162 studies, and 102 of these studies received formal criticisms by the Working Groups.

Confounding and exposure misclassification were the 2 most common criticisms about studies with PFP results, with 47 and 23 studies criticized for each, respectively (Table 3). Selection bias was mentioned as a criticism in 11 studies and questions in design, definitions, outcomes, and analytic modes were mentioned as main criticisms in 13 studies. Multiple comparisons and small study size criticisms were stated for 10 studies each. Finally, outcome misclassification, loss to follow-up, and recall bias were rarely mentioned by the Working Group. No comment on interviewer bias and subgroup analyses was made (Table 3).

### Discussion

It is our view that the debate on "false positives" does not take into account the way science works, that is, by trial-and-error and a continuous revision and reassessment of the evidence. The IARC monographs are a powerful tool to observe "science in action," because they show the assessment of carcinogenicity in a historical way, including reassessments according to well-established rules that are consistent and coherent over time. Therefore, we have used the IARC monographs to establish whether epidemiology stands out for an abnormal

**Table 1.** Breakdown of group 3 agents reviewed for the quantitative component

- |       |  |
|-------|--|
| • 509 | • Chemical agents, exposure mixtures, and circumstances in group 3   |
| • 127 | • Agents for which data on human cancer were available to the Working Group                                |
| • 68  | • Agents with at least one statistically significant positive association in their monograph (criterion 1) |
| • 37  | • Agents with at least one PFP association in their monograph (criteria 1 and 2)                           |

NOTE: Total number of agents reviewed, number of agents with data on human cancer, and number of agents with at least one PFP association (according to our criteria) are presented.

**Table 2.** Agents and studies with PFP associations and PFP frequency estimates

<b>Agent—sorted by PFP frequency</b>	<b>No. of studies w/PFP associations</b>	<b>No. of PFP associations</b>	<b>Total no. of associations mentioned in monograph (s)</b>	<b>PFP frequency estimate</b>
Talc, non-asbestiform (inhaled; refs. 14–16)	1	1	100	0.01
Spironolactone (17)	1	1	72	0.01
Surgical implants and other foreign bodies (18)				
Orthopedic implants of complex composition	4	4	230	0.02
Silicone implants	1	1	35	0.03
Fluorides (in drinking water; refs. 19, 20)	1	1	36	0.03
Fluorescent lights (21)	1	1	25	0.04
Paracetamol (22)	3	3	70	0.04
Dimethylformamide (23, 24)	1	1	22	0.05
Saccharin (25)	5	5	102	0.05
Xylenes (26, 27)	1	1	16	0.06
Tea (28)	10	11	148	0.07
Diazepam (29)	2	3	38	0.08
Doxylamine succinate (30)	1	1	13	0.08
Vitamin K substances (31)	2	3	37	0.08
Rock (stone) wool (32, 33)	7	7	72	0.10
Phenol (34, 35)	2	3	27	0.11
Toluene (36, 37)	4	4	38	0.11
Hair coloring products (personal use; refs. 38, 39)	22	32	272	0.12
Electric fields (static and Extremely Low Frequency; ref. 40)	9	11	87	0.13
Hydrochloric acid (41)	1	2	15	0.13
Paint manufacture (occupational exposure; ref. 42)	1	2	16	0.13
Insulation glass wool (32, 33)	4	5	36	0.14
Mercury and inorganic mercury compounds (43)	7	8	53	0.15
Reserpine (44, 45)	4	4	26	0.15
Coal dust (46)	9	21	135	0.16
Sulfur dioxide (47)	6	7	44	0.16
Lumber and sawmill industries (58, 49)	6	8	48	0.17
Madder root (50)	2	3	18	0.17
Methyl methacrylate (51)	2	3	18	0.17
Polychlorinated dibenzofurans (52)	1	1	6	0.17
Paper and pulp manufacture (53, 54)	3	3	18	0.17
Mineral oils (highly refined; refs. 55, 56)	12	14	67	0.21
Leather tanning and processing (57, 58)	7	8	37	0.22
Cholesterol (blood; refs. 59, 60)	11	12	50	0.24
Isonicotinic acid hydrazine (61, 62)	1	3	11	0.27
Crude oil (63)	2	3	10	0.30
Chlorinated drinking water (ref. 64; excluding studies on chlorination by-products)	3	10	28	0.36
Leather goods manufacture (65, 66)	2	2	5	0.40
<b>Total</b>	<b>162</b>	<b>213</b>	<b>2,081</b>	<b>0.10</b>

NOTE: PFP frequency estimates were calculated by dividing the number of PFP associations in the monograph(s) by the total number of associations explicitly mentioned in the respective monograph(s).

production of false-positive results. The frequency estimate of PFPs among studies reviewed by IARC Working Groups for agents with at least one PFP finding in group 3, was between 0.03 and 0.10, depending on different assumptions for the denominator. With the most conservative denominator, frequencies for each agent varied

between 0.01 and 0.40. The frequency is related to the total amount of research carried out for each exposure, with few (positive) publications often giving rise to high PFP frequencies, with 3 of the 4 agents with high PFP frequencies also being among 3 of the 4 agents with the lowest number of total associations.

**Table 3.** Study quality factors suspected to contribute to the generation of PFP findings and the number of studies criticized for each factor based on criticism by IARC Working Groups

<b>Study quality factors</b>	<b>Number of studies criticized (Working Group comments)</b>
Confounding	47
Information and recall bias	32
Exposure misclassification (excluding recall bias)	23
Outcome misclassification (excluding loss to follow-up)	2
Interviewer bias	0
Loss to follow-up	4
Recall bias	3
Selection bias	11
Questions in design, definitions, outcomes, and analytic methods	13
Multiple comparisons	10
Subgroup analyses	0
Small study size	10

It is important to note that the calculated frequency estimates for PFP results are only based on the evidence available for consideration at the time of each IARC Working Group meeting. As scientific research progresses and the consensus on what is "truth" about a research question changes, the estimates may increase or decrease, and some group 3 agents may turn out to be carcinogenic in the future. Therefore, the frequency of PFPs as defined here is probably an overestimate of false-positives.

Even though false-positive findings are shown to occur in cancer epidemiologic research, the frequency estimates obtained are in line with a dynamic science in which there is a continuous process of evaluation and reassessment of the evidence. Taking the overall estimate of PFPs, if only 1 in 10 findings published is a false-positive then, given that more research is carried out in a certain field, the direction of the totality of the evidence will tend toward the "truth." In terms of public health care decisions, given that the production of evidence is historical, public health care professionals are not expected to react immediately to a single positive association. Instead, they are likely to wait for further support or enough evidence to reach a consensus, and if a hypothesis is repeatedly tested, then any initial false-positive results will be quickly undermined.

The estimates above become even more in line with the normal procedures of science when one considers the likely inflation of these estimates due to the criteria and methodology used in this analysis. First, all statis-

tically significant positive associations also satisfying our second criterion were considered. The requirement of at least a second study that was not statistically significant or was statistically significant but with a negative effect estimate (RR lower than 1.0), means that we used a prudential approach, to minimize the risk of mistaking emerging hazards with PFPs. Second, even statistically significant associations for intermediate exposure categories were qualified as PFPs, although these would probably rarely be interpreted as positive studies. Also, for the denominator of the frequency estimate, only those associations explicitly mentioned in the monographs were counted. Finally, as the Working Groups only regarded published reports and studies, the publication bias may have affected the analysis for both the quantitative and the qualitative components as nonpositive studies are less likely to be published, therefore leading to underestimation of the denominator.

Moreover, several of the PFP studies have reported on cancer among occupational groups where exposures are not easily defined. For example, at the time when lumber and sawmill industries as well as paper and pulp manufacturers were classified as group 3, furniture and cabinet making was classified as group 1 (54), and later this led to the identification of wood dust as a group 1 carcinogen (67). Therefore, exposure to wood dust could explain the positive studies in the other wood-working industries. Similarly, when leather tanning and processing as well as leather goods manufacturing were classified as group 3, boot and shoe manufacturing and repair was classified as group 1 (58); later this led to the identification of leather dust as a group 1 carcinogen (68). Therefore, exposure to leather dust could be the reason for positive studies in the other leather industries.

### Significant findings

- The frequency of PFP findings among studies reviewed by IARC Working Groups for agents with at least one PFP finding in group 3 was 10%
- Frequencies for each individual agent varied between 1% and 40%
- The identified PFP findings were produced by 162 studies
- 102 of these studies received formal criticisms by the Working Groups
- Confounding and exposure misclassification were the 2 most common criticisms about studies with PFP results
- Given the relatively low frequency of PFPs in epidemiology, further research is required to look into the frequency of false-negative results and the factors that contribute to their generation

Another group of PFP studies has reported on cancer after exposure to variable mixtures. For instance, coal dust typically contains quartz and during mining and processing of coal dusts, adjacent rock strata may increase the quartz component of the airborne dust to about 10%, and crystalline silica dust, in the form of quartz or cristobalite, has been evaluated as group 1 (69). Similarly, when fluorescent lighting was evaluated, the Working Group noted that exposures may include UVA and UVB (21), and annual doses of UVR may be as high as median annual exposure received from sun-lamps and sunbeds, which have now been classified as group 1 (70). If these variable mixtures were removed from the analysis, the frequency of PFPs would be considerably reduced.

Less inflated false-positive estimates would be obtained by looking at agents classified in group 4, *probably not carcinogenic to humans*. However, only one agent has been classified in group 4, and no cancer epidemiologic studies have been reported for this agent. Therefore, we had to use group 3 agents, *not classifiable as to their carcinogenicity to humans*, as the gold standard. More recent IARC evaluations have identified agents which may be categorized as carcinogenic to humans, but for some specific cancer sites, the evidence is suggesting lack of carcinogenicity (ESL). For example, alcohol drinking is classified as carcinogenic to humans, but the evidence is suggesting lack of carcinogenicity for kidney cancers and non-Hodgkin lymphoma. Therefore, another approach to identifying frequency and reasons for false-positive studies could be based on the monographs' database using cancer site-specific evaluations of "evidence suggesting lack of carcinogenicity" as the gold standard.

From the qualitative component of our review, inadequate adjustment for potential confounders and information bias were the 2 most common comments by the Working Groups. Similar to the frequency estimates, the results on the factors contributing to the generation of PFP findings should also be interpreted with caution. The present analysis concentrated on studies in the monographs of group 3 agents. Therefore, there were no true positives to compare with false-positives to determine which study quality factors are associated with the generation of PFP results.

Our results are in agreement with Hernberg who argued that the most common reasons for false positive findings in occupational epidemiology are information bias and confounding (6). Swaen and colleagues (7, 8) used the IARC monographs in an attempt to investigate design characteristics associated with a false-positive study outcome in occupational cancer epidemiology studies. They reviewed 20 occupational carcinogens from IARC group 1, and the target organs in which the carcinogenic effect of these agents was observed. Studies reporting carcinogenic effects in other organs were regarded as false-positive. The authors concluded that having a specific *a priori* hypothesis and adjusting for

smoking were some of the factors that decreased the odds of producing a false-positive result. In our review, about 30% of the studies with PFP findings were criticized by the Working Group for inadequately adjusting for potential confounders, especially for smoking habits.

In the present analysis, studies reporting results on multiple exposures or between one exposure and several cancer sites as well as multiple subgroup analyses may be considered as not having a specific *a priori* hypothesis. For 10 studies, the working groups noted that multiple comparisons were carried out. In total, 35 of the 162 studies with PFP results carried out more than 5 comparisons. Subgroup analyses, however, were not raised as criticism by the Working Groups.

It is important to note that the 2 most common comments by the Working Groups on studies with PFP findings are study quality factors that can bias away from or toward the null. Exposure misclassification and particularly differential exposure misclassification can be a problem in retrospective studies and the most common example is recall bias. Differential exposure misclassification can underestimate or exaggerate an effect, and contribute to false-positive results, whereas nondifferential exposure misclassification usually biases toward the null (71). Similarly, inadequate adjustment for confounders can lead to under- or overestimation of an effect or can even change the apparent direction of an effect. Therefore, uncontrolled confounding can also bias either away or toward the null.

Given the nature of cancer epidemiologic research, the methodologic issues that can lead to false-positive results cannot always be avoided. Only rarely one single study can establish causality, generally the whole body of relevant data needs to be considered when evaluating the evidence. Therefore, researchers have an obligation to publish their results, even when the findings are inconclusive, or turn out false-negative or false-positive as science progresses. When a positive association arises, the motivation to weaken or support the initial finding leads to more studies, which can contribute to a more robust data set (72). Therefore, false-positive findings may be unavoidable, but they help to advance knowledge.

Despite the limitations already discussed, our analysis had several advantages. It relied on the IARC monographs, the most comprehensive list of evaluated agents available in cancer epidemiology, with evaluation procedures largely consistent and coherent over time. Therefore, the analysis could use all evaluations without having to regard dissimilarities in the evaluation process. Moreover, the qualitative component relied on consensus comments by the IARC Working Groups. The consensus reached by the Working Group for each agent rests on all relevant published evidence, making it an "expression of the best knowledge in the scientific community" (73).

Given the relatively low frequency of PFPs in epidemiology, further research is required to look into the frequency of false-negative results and the factors that contribute to their generation. Reasons for criticisms that are

most common in studies with false-positive findings can also underestimate an association and in terms of public health care, false-negative results may be a more important problem than false-positives.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Authors' Contributions

**Conception and design:** K. Straif, P. Vineis

**Development of methodology:** C. Demetriou, K. Straif, P. Vineis

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** K. Straif

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** C. Demetriou, K. Straif  
**Writing, review, and/or revision of the manuscript:** C. Demetriou, K. Straif, P. Vineis

#### Acknowledgments

The authors thank thoughtful comments from Neil Pearce and David Kriebel and anonymous reviewers of earlier versions of the manuscript.

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.

Received March 5, 2012; revised May 18, 2012; accepted June 11, 2012; published OnlineFirst June 19, 2012.

#### References

- Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 2008;100:988–95.
- Taubes G. Epidemiology faces its limits. *Science* 1995;269:164–9.
- Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
- Bhopal R. Paradigms in epidemiology textbooks: in the footsteps of Thomas Kuhn. *Am J Public Health* 1999;89:1162–5.
- Blair A, Saracci R, Vineis P, Cocco P, Forastiere F, Grandjean P, et al. Epidemiology, public health and the rhetoric of false positives. *Environ Health Perspec* 2009;117:1809–13.
- Hernberg S. Introduction to occupational epidemiology. Michigan, MI: Lewis Publishers; 1992.
- Swaen GM, Meijers JM, Slangen JJ. Risk of gastric cancer in pneumoconiotic coal miners and the effect of respiratory impairment. *Occup Environ Med* 1995;52:606–10.
- Swaen GG, Teggeler O, van Amelsvoort LG. False positive outcomes and design characteristics in occupational cancer epidemiology studies. *Int J Epidemiol* 2001;30:948–54.
- Clapp RW, Kriebel D. Re: False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 2009;101:211–2.
- Hansen SF, Krayner von Krauss MP, Tickner JA. Categorizing mistaken false positives in regulation of human and environmental health. *Risk Anal* 2007;27:255–69.
- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
- WHO-IARC. Monographs on the evaluation of carcinogenic risks to humans: preamble. Lyon, France: WHO Press; 2006.
- Kirkwood B, Sterne J. Using P-values and confidence intervals to interpret the results of statistical analyses. In: *Essential medical statistics*. 2nd ed. Massachusetts, MA: Blackwell Science Inc.; 2003. p. 71–9.
- WHO-IARC. Silica and some silicates. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 42. Lyon, France: WHO Press; 1987. p. 185–224.
- WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 349–50.
- WHO-IARC. Carbon black, titanium dioxide and non-asbestiform talc. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 93. Lyon, France: WHO Press; 2010. p. 277–413.
- WHO-IARC. Some thyrotropic agents. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 79. Lyon, France: WHO Press; 2001. p. 319–37.
- WHO-IARC. Surgical implants and other foreign bodies. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 74. Lyon, France: WHO Press; 1999. p. 1–409.
- WHO-IARC. Some aromatic amines, anthraquinones and nitroso compounds, and inorganic fluorides used in drinking water and dental preparations. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 27. Lyon, France: WHO Press; 1982. p. 237–303.
- WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 208–2010.
- WHO-IARC. Solar and ultraviolet radiation. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 55. Lyon, France: WHO Press; 1992. p. 1–316.
- WHO-IARC. Some chemicals that cause tumours of the kidney or urinary bladder in rodents and some other substances. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 73. Lyon, France: WHO Press; 1999. p. 401–49.
- WHO-IARC. Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 71. Lyon, France: WHO Press; 1999. p. 545–74.
- WHO-IARC. Some organic solvents, resin monomers and related compounds, pigments and occupational exposures in paint manufacture and painting. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 47. Lyon, France: WHO Press; 1989. p. 171–97.
- WHO-IARC. Some chemicals that cause tumours of the kidney or urinary bladder in rodents and some other substances. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 73. Lyon, France: WHO Press; 1999. p. 517–624.
- WHO-IARC. Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 71. Lyon, France: WHO Press; 1999. p. 1189–208.
- WHO-IARC. Some organic solvents, resin monomers and related compounds, pigments and occupational exposures in paint manufacture and painting. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 47. Lyon, France: WHO Press; 1989. p. 1215–56.
- WHO-IARC. Coffee, tea, mate, methylxanthines and methylglyoxal. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 51. Lyon, France: WHO Press; 1991. p. 207–71.
- WHO-IARC. Some pharmaceutical drugs. IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 66. Lyon, France: WHO Press; 1996. p. 37–96.
- WHO-IARC. Some thyrotropic agents. IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 79. Lyon, France: WHO Press; 2001. p. 145–59.
- WHO-IARC. Some antiviral and antineoplastic drugs, and other pharmaceutical agents. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 76. Lyon, France: WHO Press; 2000. p. 417–86.
- WHO-IARC. Man-made vitreous fibres. IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 81. Lyon, France: WHO Press; 2002. p. 1–418.



33. WHO-IARC. Man-made mineral fibres and radon. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 43. Lyon, France: WHO Press; 1988. p. 39–171.
34. WHO-IARC. Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 71. Lyon, France: WHO Press; 1999. p. 749–68.
35. WHO-IARC. Some organic solvents, resin monomers and related compounds, pigments and occupational exposures in paint manufacture and painting. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 47. Lyon, France: WHO Press; 1989. p. 263–87.
36. WHO-IARC. Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 71. Lyon, France: WHO Press; 1999. p. 829–64.
37. WHO-IARC. Some organic solvents, resin monomers and related compounds, pigments and occupational exposures in paint manufacture and painting. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 47. Lyon, France: WHO Press; 1989. p. 79–123.
38. WHO-IARC. Occupational exposures of hairdressers and barbers and personal use of hair colourants; some hair dyes, cosmetic colourants, industrial dyestuffs and aromatic amines. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 57. Lyon, France: WHO Press; 1993. p. 43–118.
39. WHO-IARC. Some industrial and cosmetic dyes, and related exposures. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 99. Lyon, France: WHO Press; 2010. p. 499–658.
40. WHO-IARC. Non-ionizing radiation, part 1: static and extremely low-frequency (ELF) electric and magnetic fields. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 80. Lyon, France: WHO Press; 2002. p. 1–429.
41. WHO-IARC. Occupational exposures to mists and vapours from strong inorganic acids; and other industrial chemicals. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 54. Lyon, France: WHO Press; 1992. p. 189–211.
42. WHO-IARC. Some organic solvents, resin monomers and related compounds, pigments and occupational exposures in paint manufacture and painting. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 47. Lyon, France: WHO Press; 1989. p. 330–442.
43. WHO-IARC. Beryllium, cadmium, mercury, and exposures in the glass manufacturing industry. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 58. Lyon, France: WHO Press; 1988. p. 239–345.
44. WHO-IARC. Some pharmaceutical drugs. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 24. Lyon, France: WHO Press; 1988. p. 211–41.
45. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 330–2.
46. WHO-IARC. Silica, some silicates, coal dust and *para*-aramid fibrils. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 68. Lyon, France: WHO Press; 1997. p. 337–406.
47. WHO-IARC. Occupational exposures to mists and vapours from strong inorganic acids; and other industrial chemicals. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 54. Lyon, France: WHO Press; 1992. p. 131–88.
48. WHO-IARC. Wood, leather and some associated industries. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 25. Lyon, France: WHO Press; 1981. p. 49–97.
49. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon: WHO Press; 1987a. p. 383–5.
50. WHO-IARC. Some traditional herbal medicines, some mycotoxins, naphthalene and styrene. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 82. Lyon, France: WHO Press; 2002. p. 49–97.
51. WHO-IARC. Wood, some industrial chemicals. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 60. Lyon, France: WHO Press; 1994. p. 445–74.
52. WHO-IARC. Polychlorinated dibenzo-*para*-dioxins and polychlorinated dibenzofurans. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 69. Lyon, France: WHO Press; 1997. p. 345–423.
53. WHO-IARC. Wood, leather and some associated industries. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 25. Lyon, France: WHO Press; 1981. p. 157–97.
54. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 385–7.
55. WHO-IARC. Polynuclear aromatic compounds, Part 2: carbon blacks, mineral oils (Lubricant Base Oils and Derived Products) and some nitroarenes. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol.33. Lyon, France: WHO Press; 1984. p. 87–168.
56. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 252–4.
57. WHO-IARC. Wood, leather and some associated industries. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 25. Lyon, France: WHO Press; 1981. p. 201–47.
58. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 236–37.
59. WHO-IARC. Some food additives, feed additives and naturally occurring substances. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol.31. Lyon, France: WHO Press; 1983. p. 95–132.
60. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 161–5.
61. WHO-IARC. Some aromatic amines, hydrazine and related substances, *N*-nitroso compounds and miscellaneous alkylating agents. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 4. Lyon, France: WHO Press; 1974. p. 159–72.
62. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 227–8.
63. WHO-IARC. Occupational exposures in petroleum refining; crude oil and major petroleum fuels. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 45. Lyon, France: WHO Press; 1989. p. 119–58.
64. WHO-IARC. Chlorinated drinking-water; chlorination by-products; some other halogenated compounds; cobalt and cobalt compounds. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 52. Lyon, France: WHO Press; 1989. p. 45–141.
65. WHO-IARC. Wood, leather and some associated industries. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 25. Lyon, France: WHO Press; 1989. p. 279–92.
66. WHO-IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Overall evaluations of carcinogenicity: an updating of IARC monographs Volumes 1 to 42. Supplement 7. Lyon, France: WHO Press; 1987a. p. 235–6.
67. WHO-IARC. Wood dust and formaldehyde. IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 62. Lyon, France: WHO Press; 1995. p. 35–215.
68. WHO-IARC. A review of human carcinogens: arsenic, metals, fibres, and dusts. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 100C. Lyon, France: WHO Press; 2012. p. 317–53.

69. WHO-IARC. Silica, some silicates, coal dust and *para*-aramid fibrils. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 68. Lyon, France: WHO Press; 1997. p. 41–242.
70. WHO-IARC. A review of human carcinogens: radiation. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, Vol 100D. Lyon, France: WHO Press; 2012. p. 35–101.
71. Pearce N, Checkoway H, Kriebel D. Bias in occupational epidemiology studies. *Occup Environ Med* 2007;64:562–8.
72. Crosignani P. Re: false-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 2009;101:212–3.
73. Vineis P. The skeptical epidemiologist. *Int J Epidemiol* 2009;38:675–7.

# Cancer Epidemiology, Biomarkers & Prevention

## From Testing to Estimation: The Problem of False Positives in the Context of Carcinogen Evaluation in the IARC Monographs

Christiana A. Demetriou, Kurt Straif and Paolo Vineis

*Cancer Epidemiol Biomarkers Prev* 2012;21:1272-1281. Published OnlineFirst June 19, 2012.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1055-9965.EPI-12-0276](https://doi.org/10.1158/1055-9965.EPI-12-0276)

**Cited articles** This article cites 13 articles, 3 of which you can access for free at:  
<http://cebp.aacrjournals.org/content/21/8/1272.full#ref-list-1>

**Citing articles** This article has been cited by 1 HighWire-hosted articles. Access the articles at:  
<http://cebp.aacrjournals.org/content/21/8/1272.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cebp.aacrjournals.org/content/21/8/1272>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.