

Cancer Epidemiology, Biomarkers & Prevention



False Positives in Cancer Epidemiology

Joseph K. McLaughlin and Robert E. Tarone

Cancer Epidemiol Biomarkers Prev 2013;22:11-15. Published OnlineFirst November 1, 2012.

Updated Version Access the most recent version of this article at:
doi:[10.1158/1055-9965.EPI-12-0995](https://doi.org/10.1158/1055-9965.EPI-12-0995)

Cited Articles This article cites 38 articles, 15 of which you can access for free at:
<http://cebp.aacrjournals.org/content/22/1/11.full.html#ref-list-1>

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.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Hypothesis/Commentary

False Positives in Cancer Epidemiology

Joseph K. McLaughlin and Robert E. Tarone

Abstract

Background: A recent attempt to estimate the false-positive rate for cancer epidemiology studies is based on agents in International Agency for Research on Cancer (IARC) category 3 (agent not classifiable as to its carcinogenicity to humans) in the IARC Monographs Program.

Methods: The estimation method is critiqued regarding biases caused by its reliance on the IARC classification criteria for assessing carcinogenic potential.

Results: The privileged position given to epidemiologic studies by the IARC criteria ensures that the percentage of positive epidemiologic studies for an agent will depend strongly on the IARC category to which the agent is assigned. Because IARC category 3 is composed of agents with the lowest-assessed carcinogenic potential to which the estimation approach in question could be applied, a spuriously low estimated false-positive rate was necessarily the outcome of this approach.

Conclusions: Tendentious estimation approaches like that employed will by necessity produce spuriously low and misleading false positive rates.

Impact: The recently reported estimates of the false-positive rate in cancer epidemiology are seriously biased and contribute nothing substantive to the literature on the very real problems related to false-positive findings in epidemiology. *Cancer Epidemiol Biomarkers Prev*; 22(1); 11–15. ©2012 AACR.

A recent article (1) attempted to estimate the false-positive rate for cancer epidemiology studies based on agents in International Agency for Research on Cancer (IARC) category 3 (agent not classifiable as to its carcinogenicity to humans) in the IARC Monographs Program. The estimation method used was based on the assumption that the IARC classification of agents about their carcinogenic potential (or any similar human endeavor at reaching consensus) provides a scientifically sound basis for assessing the false-positive rate in cancer epidemiology studies. Any such consensus effort uses classification criteria, and we critique the proposed false-positive rate estimation method because of potential biases induced by its reliance on the IARC criteria for assessing carcinogenic potential. The IARC criteria give epidemiologic studies epistemic priority in working group deliberations, and accordingly, the IARC category to which an agent is assigned will depend strongly on the percentage of positive epidemiologic studies for that agent. Because IARC category 3 is composed of agents with inadequate evidence of cancer risk from epidemiologic studies, we argue that the reported false-positive rate estimates based on this IARC category seriously underestimate

the actual false-positive rate in cancer epidemiology investigations.

The false-positive rate was estimated as the percentage of positive studies among all epidemiologic studies assessing cancer risk for agents classified in IARC category 3, the IARC category corresponding to the lowest-assessed carcinogenic potential to which the proposed method could be applied (1). For an agent to be assigned to category 3, there must be inadequate evidence of carcinogenicity in both humans and animals. Because IARC gives the greatest probative weight to epidemiologic evidence, any agents, including noncarcinogens, with multiple-positive epidemiology studies would by IARC definitional criteria almost always be assigned to an IARC category indicative of a higher assessment of carcinogenic potential than that of category 3. An agent with even a few positive epidemiologic findings will likely be assigned to category 2b (possibly carcinogenic to humans), category 2a (probably carcinogenic to humans), or sometimes category 1 (carcinogenic to humans). By the IARC classification criteria, agents in category 3 will have almost no positive epidemiology studies, and thus the evaluation by the authors of the percentage of positive epidemiologic studies for category 3 agents was guaranteed in advance to provide a misleadingly low estimate of the false-positive rate.

The fundamentally flawed nature of the estimation method used in the article (1) is best understood by considering the approach its authors indicate that they would have preferred to use to estimate the false-positive rate. They indicate that ideally they would have estimated

Authors' Affiliation: International Epidemiology Institute, Rockville, Maryland

Corresponding Author: Joseph K. McLaughlin, International Epidemiology Institute, 1455 Research Boulevard, Suite 550, Rockville, MD 20850. Phone: 301-424-1054; Fax: 301-424-1053; E-mail: jkm@iei.us

doi: 10.1158/1055-9965.EPI-12-0995

©2012 American Association for Cancer Research.

the false-positive rate as the percentage of positive epidemiologic studies for category 4 agents (agents probably not carcinogenic to humans). They could not use this approach, because of the 950 agents evaluated to date by IARC, only 1 agent resides in category 4, and there was no epidemiologic study investigating possible cancer risk associated with this agent (Appendix 1). In order for an agent to qualify for IARC category 4, there must be evidence suggesting a lack of carcinogenicity in both humans and experimental animals. Obviously, if there was even a single epidemiologic study with a positive finding for an agent and any cancer, then that agent would by definition not be eligible for category 4 (unless there were a large number of methodologically superior negative epidemiologic studies of the same agent). Thus, because of the stringent (i.e., virtually impossible to meet) IARC criteria for category 4 agents (and assuming that there were any such agents evaluated in epidemiologic studies), the estimated false-positive rate for epidemiologic studies would have been approximately 0 if the authors' preference could have been met (1). The fact that no agent actually investigated in an epidemiologic study has ever been assigned to category 4, is itself strong evidence supporting the seriousness of the problem posed by false positives in cancer epidemiologic studies. Even a noncarcinogen is likely to show some evidence of a positive association for some type of cancer in occupational or environmental epidemiology studies as such investigations are currently conducted and interpreted (2–4).

No scientifically sound evaluation of true false positives can be based on the IARC consensus approach about carcinogenic potential. To provide a reasonable and valid estimate of the false-positive rate in cancer epidemiology studies, the true state of nature would have to be known (i.e., which agents are real carcinogens and which are noncarcinogens), and the approach (1) reviewed here would have to be applied to all actual noncarcinogens that have been evaluated by IARC, regardless of the present IARC category to which they have been assigned. Because of the aforementioned priority given to epidemiologic evidence by the IARC criteria, the categories to which true noncarcinogens are assigned depend strongly on the percentage of positive epidemiologic studies that have investigated them. Hence, the percentage of positive studies would be lowest for category 3 agents (i.e., the method used by the authors), and would increase successively for agents in categories 2b, 2a, and 1. Most of the agents in category 2b probably do not cause cancer in humans in the usual definition of "probably," and several agents in category 2a, and even a few agents in category 1, probably do not cause cancer in humans (Appendix 2). True noncarcinogens classified in IARC categories 1 and 2 would be particularly informative, because they would lay bare the pitfalls of virtually all of the mechanisms that produce false positivity, particularly the bandwagon effect in research once suspicion is raised about an agent by a governmental agency, activist group, or nongovern-

mental organization (5), leading to a plethora of studies attempting to document the suspected association. Epidemiologists are not immune to the influence of information cascades, availability cascades, and reputational cascades that can lead to the perpetuation of concerns about a hypothesized risk initially generated by false-positive findings (6).

The claims (1) that the frequency of false positives in epidemiology is relatively low and that false negatives may pose a greater problem than false positives are without scientific basis. False-positive findings dominate the epidemiologic literature under the current standards of practice (4, 7), with empirical evidence indicating that there are at least 20 false-positive results published for every true-positive finding (8, 9). This is not surprising in view of the increasing number of epidemiology journals, the enormous and ballooning number of published epidemiology articles (10), and the pronounced incentives on the part of researchers and journals to publish positive results (2, 7). Thousands of potential risk factors have been investigated in the last 4 decades, yet the true known primary causes of the major cancers, save lung and cervical cancer, are extraordinarily limited (5). The potential for false positives is further increased by common practices in the analysis and interpretation of individual epidemiologic studies, including the use of multiple exposure metrics for a risk factor of interest, the fitting of multiple statistical models, the analysis of numerous subgroups, and the selective reporting of analytic results (2–4). It has been asserted, without benefit of any evidence, that false-negative findings occur about as often as false-positive findings in epidemiologic research (11). In fact, however, false positives occur much more often than false negatives in most epidemiologic study settings (9).

Sorting out true causal relationships under such a high-background rate of false-positive findings is a daunting task. Even good-faith efforts are going to lead to mistakes because of a variety of cognitive biases affecting the dynamics of consensus decision making by *ad hoc* deliberative bodies (12–15), such as the working groups that evaluate the carcinogenicity of agents in the IARC Monographs Program. These problems are exacerbated when selection criteria for working group membership produce conditions conducive to group thinking (12, 14–16), leading to inherent biases in the consensus evaluation process (17). Ignoring the difficulties caused by false positives, working group composition and dynamics, and various factors affecting decision making, including the fundamental scientific conflict-of-interest posed by researchers evaluating the soundness and importance of their own work, will continue to undermine the scientific value of the IARC Monographs Program (16).

Appendix 1

Examination of the one agent currently included in IARC category 4, caprolactam, illustrates the virtual impossibility of meeting the classification criteria for

category 4 (i.e., evidence suggesting a lack of carcinogenicity is required) and the subjective application of IARC criteria in practice. There were no epidemiologic studies of caprolactam (i.e., no studies suggesting lack of carcinogenicity in humans), but it was put in category 4 on the basis of an option permitting an agent with inadequate human evidence to be assigned to category 4, provided there is evidence suggesting lack of carcinogenicity of the agent in animals. The National Cancer Institute (NCI) conducted a study in male and female mice and rats that found generally lower tumor rates in animals exposed to caprolactam (18). The only compound-related effects reported in the study, however, were dose-related decreases in body weight for both rats and mice exposed to caprolactam, and a dose-related decrease in food consumption in rats. Lower tumor rates should be expected in rodents with restricted food consumption and reduced weight gain, and thus the NCI rodent experiments are not interpretable as providing evidence against carcinogenicity of caprolactam. On the basis of the IARC evaluation criteria, it seems that caprolactam should have been assigned to IARC category 3.

Appendix 2

Coffee is classified by the IARC Monographs Program as an agent that is possibly carcinogenic to humans (i.e., category 2B). Few agents have been investigated more widely in epidemiologic studies than coffee. There is strong evidence that coffee consumption is inversely associated with endometrial cancer risk (19) and substantial evidence for an inverse association between coffee intake and liver cancer (20, 21). A recent comprehensive meta-analysis reported an inverse association between exposure to coffee and total cancer risk, and evaluation by organ site led to the conclusion that "coffee drinking was associated with a reduced risk of bladder, breast, buccal and pharyngeal, colorectal, endometrial, esophageal, hepatocellular, leukemic, pancreatic, and prostate cancers" (22). It seems that coffee may be the only agent evaluated by IARC for which the epidemiologic evidence meets the criteria for category 4, yet it currently resides in category 2B based on the IARC working group deliberations. The sheer number of cancer sites with evidence of protection from coffee drinking is surprising, and many of the reduced relative risk estimates may have been the result of residual confounding or confounding that reflects as yet unrecognized differences between coffee drinkers and people who do not drink coffee. Nevertheless, the weight of the epidemiologic evidence indicates that coffee probably does not cause cancer in humans. It should be noted that the false-positive estimation approach evaluated in this critique should be applied only to agents that neither increase nor decrease the risk of cancer, and thus coffee should not be included in such an estimation effort.

Cell phone exposure is also classified by the IARC Monographs Program as being possibly carcinogenic to

humans. Cell phone exposure has been widely investigated in epidemiologic studies, with sporadic reports of increased risk estimates, most coming from case-control studies conducted by a single research group (23). The inconsistent evidence of a possible association between cell phone use and brain or salivary gland cancer risk from case-control studies is negated by the absence of any recent increase in national population cancer rates in developed countries (23–28). No increase in population rates has been observed through 2009 in countries, such as those in Scandinavia where cell phone use began early, despite the fact that increased relative-risk estimates for cell phone use and brain tumor risk were being reported in case-control studies based on cases diagnosed as early as 1994 through 1996 (29). The absence of any temporal increase in brain cancer rates in the presence of ubiquitous cell phone use contrasts with the increase in national rates of lung cancer observed in the 1930s resulting from increased cigarette smoking after World War I, particularly in view of the fact that the prevalence of cell phone use far exceeds the maximum peak prevalence of cigarette smoking during the twentieth century (particularly among women and children). The stable trends in national brain cancer rates are even more surprising if cordless phone use also increases brain cancer risk as some have claimed (30). Continued monitoring of national cancer rates for sites plausibly affected by cell phone exposure should continue, but at present the epidemiologic evidence is consistent with the conclusion that cell phone use probably does not cause cancer in humans.

Comprehensive reviews indicate that, even for some agents classified in IARC category 1, the epidemiologic evidence of carcinogenicity is weak and inconsistent (31, 32). The interpretation of epidemiologic evidence is particularly problematic for formaldehyde (31, 33, 34). Formaldehyde was originally placed in category 1 on the basis of evidence of increased nasopharyngeal cancer (NPC) risk, primarily on the basis of a single NCI cohort study of industrial workers with occupational formaldehyde exposure (35). As has been noted previously, however, all of the evidence of increased NPC risk in this study came from only 1 of the 10 plants included in the study (31, 33, 34). The standardized mortality ratio (SMR) for formaldehyde-exposed workers in plant 1 in the NCI study is 9.1 [95% confidence interval (CI), 3.3–19.8] based on 6 cases. In the remaining 9 plants of the NCI study, the SMR for exposed workers is 0.6 (95% CI, 0.1–2.3) based on 2 cases. Including all occupational cohort studies from around the world, but excluding plant 1 of the NCI cohort, the summary SMR for formaldehyde-exposed workers is 0.6 (95% CI, 0.2–1.6) based on 4 NPC cases (31, 35–40). Follow-up of an expanded cohort of workers from plant 1 of the NCI cohort study yielded an SMR of 4.4 (95% CI, 1.8–9.1) based on 7 NPC cases (41). The elevated SMR in plant 1 of the NCI study was observed despite the fact that no uniquely dangerous characteristics of formaldehyde exposure have been documented in that facility (34). Furthermore, 3 of 7 workers with NPC had formaldehyde

exposure of duration 3 months or fewer, and a fourth had formaldehyde exposure of duration 8 months (33).

IARC (42) recently concluded that formaldehyde also causes leukemia in humans, primarily on the basis of an NCI case-control study of embalmers (43). Embalmers are reported to have the highest peak exposure to formaldehyde of any occupation (43), and peak exposure was the exposure metric that showed the strongest association with NPC risk in the NCI industrial cohort study (35). Thus, embalming would seem to be an optimal occupation in which to investigate possible NPC risk. The NCI embalmer case-control study had numerous methodologic problems that are beyond the scope of this appendix, but IARC nonetheless gave it great weight in evaluating the inconsistent epidemiologic evidence of leukemia risk, including it in the discussion of cohort studies because it was treated as a case-control study nested within a cohort (42). Inexplicably, the embalmer case-control study was ignored in the IARC discussion of NPC risk (42), perhaps because the reported OR for NPC of 0.1 (95% CI, 0.01–1.2) among embalmers (43) was hard to reconcile with the consensus IARC working group conclusion that formal-

dehyde exposure is a cause of NPC in humans. A scientifically sound summary of the epidemiologic evidence about NPC risk would conclude that something in or around plant 1 of the NCI cohort study may have led to increased risk of NPC (41), but that formaldehyde, even at high-occupational exposure levels, probably does not cause NPC in humans.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: J.K. McLaughlin, R.E. Tarone

Development of methodology: R.E. Tarone

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R.E. Tarone

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R.E. Tarone

Writing, review, and/or revision of the manuscript: J.K. McLaughlin, R.E. Tarone

Study supervision: J.K. McLaughlin

Received August 24, 2012; revised October 19, 2012; accepted October 19, 2012; published OnlineFirst November 1, 2012.

References

- Demetriou C, Straif K, Vineis P. From testing to estimation: the problem of false positives in the context of carcinogen evaluation in the IARC Monographs. *Cancer Epidemiol Biomarkers Prev* 2012;21:1272–81.
- Phillips CV. Publication bias *in situ*. *BMC Med Res Methodol* 2004;4:20.
- Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640–8.
- 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.
- Kabat GC. Hying health risks: environmental hazards in daily life and the science of epidemiology. New York: Columbia University Press; 2008. p. 5, 9–11.
- Kuran T, Sunstein CR. Availability cascades and risk regulation. *Stanford Law Rev* 1999;51:683–768.
- Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
- Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361:865–72.
- Ioannidis JPA, Tarone RE, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* 2011;22:450–6.
- Ioannidis JPA, Tarone RE, McLaughlin JK. X-WAS, traditional epidemiology, and policy action. *Epidemiology* 2011;22:467–8.
- 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 Perspect* 2009;117:1809–13.
- Brown R. Group processes. 2nd ed. Malden, MA: Blackwell Publishing; 2000. p. 167–224.
- Sunstein CR. Deliberating groups versus prediction markets (or Hayek's challenge to Habermas). *Episteme* 2006;3:192–213.
- Erren TC. IARC's plea for traditional 'expert' working groups—a recipe for problems? *Int J Epidemiol* 2011;40:1727–8.
- McLaughlin JK, Boffetta P, La Vecchia C, Lipworth L, Blot WJ, Tarone RE. Problems with IARC's 'expert' working groups. *Int J Epidemiol* 2011;40:1728–9.
- Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. A further plea for adherence to the principles underlying science in general and the epidemiologic enterprise in particular. *Int J Epidemiol* 2009;38:678–9.
- Kabat G. Behind the World Health Organization's "cancerous" pronouncement on cell phones. *Forbes*; 2011 [cited 2012 Jul 31]. Available from: <http://www.forbes.com/sites/realspin/2011/08/23/world-health-organization-cancerous-cell-phones/>.
- National Cancer Institute. Carcinogen bioassay of caprolactam (CAS No. 105–60-2): F344 rats and B6C3F₁ mice (feed study). Bethesda, MD: National Toxicology Program; 1982 Mar. Technical Report No. 214.
- Je Y, Giovannucci E. Coffee consumption and risk of endometrial cancer: findings from a large up-to-date meta-analysis. *Int J Cancer* 2012;131:1700–10.
- Bravi F, Bosetti C, Tavani A, La Vecchia C. Coffee drinking and hepatocellular carcinoma: an update. *Hepatology* 2009;50:1317–8.
- Gao J, Xie L, Yang WS, Zhang W, Gao S, Wang J, et al. Risk factors for hepatocellular carcinoma—current status and perspectives. *Asian Pacific J Cancer Prev* 2012;13:743–52.
- Yu X, Bao Z, Zou J, Dong J. Coffee consumption and risk of cancers: a meta-analysis of cohort studies. *BMC Cancer* 2011;11:96.
- Swerdlow AJ, Feychting M, Green AC, Kheifets L, Savitz DA International Commission for Non-Ionizing Radiation Protection Standing Committee on Epidemiology. Mobile phones, brain tumours and the interphone study: where are we now? *Environ Health Perspect* 2011;119:1534–8.
- Deltour I, Johansen C, Auvinen A, Feychting M, Klæboe L, Schüz J. Time trends in brain tumor incidence rates in Denmark, Finland, Norway, and Sweden 1974–2003. *J Natl Cancer Inst* 2009;101:1721–4.
- de Vocht F, Burstyn I, Cherrie JW. Time trends (1998–2007) in brain cancer incidence rates in relation to mobile phone use in England. *Bioelectromagnetics* 2011;32:334–9.
- Ahlbom A, Feychting M. Mobile telephones and brain tumours. *BMJ* 2011;343:d6605.
- Little MP, Rajaram P, Curtis RE, Devesa SS, Inskip PD, Check DP, et al. Mobile phone use and glioma risk: comparison of epidemiological study results with incidence trends in the United States. *BMJ* 2012;344:e1147.
- Shu X, Ahlbom A, Feychting M. Incidence trends of malignant parotid gland tumors in Swedish and Nordic adults 1970 to 2009. *Epidemiology* 2012;23:766–7.

29. Hardell L, Mild KH, Pahlson A, Hallquist A. Ionizing radiation, cellular telephones and the risk for brain tumors. *Eur J Cancer Prev* 2001;10:523-9.
30. Hardell L, Carlberg M, Mild KH. Re-analysis of risk for gliomas in relation to mobile telephone use: comparison with the results of the Interphone international case-control study. *Int J Epidemiol* 2011;40:1126-8.
31. Bosetti C, McLaughlin JK, Tarone RE, Pira E, La Vecchia C. Formaldehyde and cancer risk: a quantitative review of cohort studies through 2006. *Ann Oncol* 2008;19:29-43.
32. Boffetta P, Fryzek JP, Mandel JS. Occupational exposure to beryllium and cancer risk: a review of the epidemiologic evidence. *Crit Rev Toxicol* 2012;42:107-18.
33. Tarone RE, McLaughlin JK. Re: "Mortality from solid cancers among workers in formaldehyde industries". *Am J Epidemiol* 2005;161:1089-91.
34. Marsh GM, Youk AO. Reevaluation of mortality risks from nasopharyngeal cancer in the formaldehyde cohort study of the National Cancer Institute. *Regul Toxicol Pharmacol* 2005;42:275-83.
35. Hauptmann M, Lubin JH, Stewart PA, Hayes RB, Blair A. Mortality from solid cancers among workers in formaldehyde industries. *Am J Epidemiol* 2004;159:1117-30.
36. Edling C, Järholm B, Andersson L, Axelson O. Mortality and cancer incidence among workers in an abrasive manufacturing industry. *Br J Ind Med* 1987;44:57-9.
37. Bertazzi PA, Pesatori A, Guercilena S, Consonni D, Zocchetti C. Carcinogenic risk for resin producers exposed to formaldehyde: extension of follow-up. *Med Lav* 1989;80:111-22.
38. Andjelkovich DA, Janszen DB, Brown MH, Richardson RB, Miller FJ. Mortality of iron foundry workers: IV. Analysis of a subcohort exposed to formaldehyde. *J Occup Environ Med* 1995;37:826-37.
39. Coggon D, Harris EC, Poole J, Palmer KT. Extended follow-up of a cohort of British chemical workers exposed to formaldehyde. *J Natl Cancer Inst* 2003;95:1608-15.
40. Pinkerton LE, Hein MJ, Stayner LT. Mortality among a cohort of garment workers exposed to formaldehyde: and update. *Occup Environ Med* 2004;61:193-200.
41. Marsh GM, Youk AO, Buchanich JM, Erdal S, Esmen NA. Work in metal industry and nasopharyngeal cancer mortality among formaldehyde-exposed workers. *Regul Toxicol Pharmacol* 2007;48:308-19.
42. IARC Working Group. IARC monographs on the evaluation of carcinogenic risks to humans, vol 100F: chemical agents and related occupations. Lyon, France: IARC Press; 2012. p. 401-35.
43. Hauptmann M, Stewart PA, Lubin JH, Beane Freeman LE, Hornung RW, Herrick RF, et al. Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. *J Natl Cancer Inst* 2009;101:1696-1708.