

IARC Evaluations of Cancer Hazards: Comment on the Process with Specific Examples from Volume 105 on Diesel Engine Exhaust

John F Gamble*

Retired, Consultant, USA

IARC is an international agency of WHO with a monograph program to develop “critical reviews and evaluations of evidence on the carcinogenicity of a wide range of human exposures” [1,2]. Interdisciplinary working groups (WGs) of expert scientists meet for 8 days and evaluate the weight of evidence of an agent and classify it into one of 5 categories of carcinogenicity. The Monograph program began in 1965 and has evaluated >900 agents. About 10% have been classified as carcinogenic (Group 1) and >33% as probably carcinogenic or possibly carcinogenic (Groups 2A, 2B). This approximately one-week meeting approach was chosen when the amount of data for any specific agent was much less than today (some of the monographs in the early volumes are fewer than 10 pages long). It was never reformed despite an enormous increase in the available literature and the broadening to include complex mixtures, occupational exposures, physical and biological agents and lifestyle factors.

The WGs review considers four areas of evidence: 1) Sources of exposure, which is especially important for a changing technology such as diesel engine exhaust (DEE); 2) Animal Evidence; 3) Epidemiology Evidence, which is most important in Monograph 105; and 4) Mechanistic Evidence that plays an increasingly important role in evaluating biological plausibility in the weight of evidence.

Invited expert scientists are divided into working groups to evaluate all published literature and make preliminary conclusions that are finalized in plenary sessions of all participants. Non-voting participants include a) invited specialists who have critical knowledge but may have a conflict of interest, b) representatives of health agencies, c) observers or other interested parties with relevant scientific credentials, and d) members of the IARC Secretariat who have relevant experience; only this last subgroup may participate in discussions, draft text, and prepare tables and analyses. All potential participants are assessed to determine financial, employment, and research support for potential conflicts of interest before they are invited to participate. Monographs provide an evaluation of cancer hazards and an assessment of the “strength of the available evidence that an agent could alter the incidence of cancer in humans”. This evaluation is considered an authoritative and expert-based classification with international significance. Increased interest in environmental epidemiology has resulted in publication in the scientific literature of an ever increasing number of cohort and case-control studies seeking to establish an association between exposure to an agent and an adverse effect. These studies are commonly not in complete agreement. Oftentimes inconsistencies can be traced to biases or confounding in the experimental design or errors in the analysis of the data. As stated in the Preamble the purpose of the WG is to identify these problems (IARC 2006): “When an important aspect of a study that directly impinges on its interpretation should be brought to the attention of the reader, a Working Group comment is given in square brackets”. Unless adequate biostatistical and epidemiological expertise is brought to bear on a specific topic, mistakes and misinterpretations can result. Sometimes IARC conclusions that are based on epidemiology have been controversial, if not mistaken or misinterpreted. Formaldehyde [3] and silica [4] are two examples. The recent IARC conclusion on diesel engine exhaust [1,2] is likely to be controversial as it is based

on potentially incorrect interpretations of major epidemiology studies because of limitations in the scientific assessment.

There have been wide-ranging discussions in the literature concerning potential conflicts-of-interest in participants, the make-up of the WGs, and whether the IARC selection process may be contributing to an increase in false-positive determinations in epidemiology [5-13]. In essence, the IARC review procedure for reviewing evidence has become a closed process that does not encourage open scientific debate or alternative viewpoints. This editorial relates to that discussion and examines serious flaws in the IARC process that have their roots in IARC’s policy of discouraging differing scientific viewpoints by non-voting participants. A review of the source of these flaws with regard to epidemiology is the first purpose of this editorial. The second purpose is to suggest modifications in the IARC review process to make it more scientifically robust and to improve their assessment of the weight of the evidence. My experiences as an observer in several epidemiology WGs confirm the need for increased scrutiny of individual studies and assessment of the weight of evidence used in the monograph process.

Examples of Potentially “False-Positive” Interpretations of Three Diesel Studies

Two recent DEE studies of miners and truckers were considered to be the most informative studies and were used to support the WG conclusion of sufficient evidence that DEE is carcinogenic. Important aspects of these studies provide examples where the current IARC review process may have failed to accurately interpret results. It is troubling that despite 30+ years of research into diesels and cancer, two of the most influential papers were very recent and reviewers did not have time for adequate review (including replication of results by independent groups).

The WG indicated the miner case-control study [14] provided “some of the strongest evidence of an association”, with controls for smoking, low potential confounding exposures in the mines, and “well-documented” high diesel exposures. What should also have been included in square brackets were considerations that could potentially change interpretation of this study. The smoking x location interaction terms used to adjust for smoking may be statistically incorrect because of the absence of categorical variables for each of smoking and location. Since the standard procedure of including main effects was not followed, re-analyses of the data are required to sort this out. Smoking was said to

*Corresponding author: John F Gamble, Retired, Consultant, USA, E-mail: john.f.gamble@comcast.net

Received July 24, 2012; Accepted July 25, 2012; Published July 27, 2012

Citation: Gamble JF (2012) IARC Evaluations of Cancer Hazards: Comment on the Process with Specific Examples from Volume 105 on Diesel Engine Exhaust. J Clin Toxicol 2:e106. doi:10.4172/2161-0495.1000e106

Copyright: © 2012 Gamble JF. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

have been a “negative confounder” among underground (UG) workers but smoking was not associated with DEE exposure for the combined group of both UG and surface workers and so substantive confounding from smoking is unlikely [15,16]. Smoking adjustments produced the positive association, making these study results unreliable. Further, an independent analysis was unable to completely replicate results [17]. Based on the uncertainties in the published results, a valid interpretation of this study requires replication and verification by independent groups. This reanalysis should also look for evidence of compromised follow-up that produced spurious positive exposure-response trends that were noted in the original railroad worker cohort [18] and an NCI study of formaldehyde workers [3,19-22].

A cohort of US truckers [23] showed positive exposure-response trends that “were markedly more pronounced when adjustment for duration of work was included in the models”. This type of adjustment for duration can lead to model instability because of the co-linearity associated with having duration twice in the model (duration is already included in the cumulative exposure metric). The difficulties encountered in orally discussing limitations in this study prompted a letter to the editor that was accepted for publication. This peer review suggested that the multiple adjustments for duration of employment may have distorted the Cox model because of misspecification and that the straightforward interpretations that were accepted by IARC are not feasible and are likely to be incorrect [24].

Criticisms of the first study were in the published literature but were not adequately evaluated or debated in the WG or plenary sessions. The validity of the statistical analysis in the 2nd study could not be questioned in the literature as the paper only became available at the beginning of the DEE discussions in Lyon and so did not meet the IARC posted requirement for new scientific evidence. Nevertheless, industry observers orally commented on these limitations, but they were not seriously entertained or discussed.

A third study was interpreted as significant supportive evidence for the WG conclusion.

The European-Canadian pooled case-control study [25] showed a positive association in a smoking-adjusted analysis that as noted by the WG is unlikely to [“be explained by bias or confounding”]. An observer pointed out the contrary lack of an association among non-smokers, possibly the least confounded group of cases. The lack of confounding comment in the square brackets is contradicted by evidence of positive confounding from a non-representative referent group as well as inadequate adjustment for occupational confounding and SES [26-29]. Adequacy of responses to some of these criticisms [30] is unresolved. Discussion of these issues was prematurely ended. There was significant conflict of interest as three co-authors were participating in WG discussions. In general, the practice of investigators reviewing their own work, or work of close collaborators, appears to be less closely watched by IARC than the close monitoring of non-voting representatives for bias and comments on scientific issues in individual studies being considered by the WG.

Suggested Modifications in IARC Process

These abbreviated examples suggest a need to modify IARC procedures to eliminate the deficiencies in WG critiques of individual studies. The practices suggested below are consistent with IARC current guidelines but allow for increased diversity of expertise and time for rigorous peer-review. The suggestions are based on practical, proven

principles developed and practiced by Toxicology Excellence for Risk Assessment (TERA) [31] for agencies such as IARC.

TERA has put forth four essential key principles that are necessary for production of a scientific-based group conclusion regarding toxicity of an agent. The following discusses each principle; the limitations of the IARC process, and suggest modifications to the process.

Independence

There must be freedom from bias and conflicts of interests.

IARC screens for financial conflicts of interest that may produce industry bias. Potential issue bias and vested interests are recognized in the IARC preamble where “Care is taken to ensure that each study summary is written or reviewed by someone not associated with the study being considered”. This requirement doesn’t always work in practice and in itself does not insure an independent result. In the diesel monograph an author was the dominant voice regarding the just-published trucker study [23] and argued against and effectively prevented interpretative comments from being included in the square brackets that would have noted for the reader the destabilizing double use of the duration variable.

Suggestions

All persons have both recognized and un-recognized biases and vested interests. Potential conflicts of interest should be readily accessible for all to see, but opinions should be judged on factual accuracy and logic rather than source. Variable degrees of scientific honesty and independence are found in all affiliations and it cannot be assumed that they are only found in government, academia and NGOs.

Inclusion of Appropriate Expertise

Scientific opinion from a broad range of backgrounds and affiliations (e.g., government, academia, industry, environmental or public interest groups, consulting) are required to provide diverse scientific perspectives.

Working group participants are well-represented by government, academia and IARC staff, but industry and consultants are allowed only a limited role in the process, and then usually only at the end of a discussion after opinions have been largely formed. Expertise and preparation of working group members is, in practice, largely restricted to the chairman and the group member who wrote the initial summaries and who subsequently present their review to the WG. In the DEE epidemiology WG there was no apparent biostatistician expertise among the WG voting members adequate for accurately evaluating the questionable statistical analyses of the miner and trucker studies. My experience indicates that for specific studies, and indeed for the overall topic, the more knowledgeable and prepared individual(s) in the room include industry observers who were restricted to minimal participation in scientific discussions. These IARC restrictions have excluded appropriate expertise and have at times resulted in incorrect, if not biased, comments in the square brackets and excluded comments that are important for interpreting study results.

Suggestions

Appropriate expertise includes observers from industry and other interested parties who often have extensive work experience and knowledge of the agents under consideration. This expertise should be utilized by allowing increased participation by all knowledgeable parties in oral and written discussions of relevant scientific issues. Qualified

chairmen should be able to control oral discussion so important scientific issues are raised and discussed at the meeting. Opportunities for written comments, which become part of the session record, should also be allowed when and where appropriate.

Transparency

Activities and results should be organized and conducted so that those within and external to the process can judge for themselves the adequacy and credibility of the results.

Transparency of results is provided for each study in square brackets and in the general summary of the evidence that produced the conclusion. Square brackets are supposed to include any supplementary analyses, or more commonly, "important aspect[s] of a study that directly impinge on its interpretation" and that "should be brought to the attention of the reader". Currently, contents of the square brackets can be non-informative or patently incorrect, especially when there are time limitations on the amount of debate that can take place. Comments from observers and national representatives in the WG meetings regarding study limitations were not acted on, or were curtailed for lack of "time." Comments in square brackets were at times arbitrarily selected so as to include strengths only, without comments on study limitations. The examples discussed previously are particularly egregious instances of missing, incorrect or misleading comments in the square brackets.

Suggestions

The comments included in square brackets should include both strengths and weaknesses and a clear indication of the WG interpretation of each study based on credible input from all invited participants, whether voters or not-voters.

Robust Scientific Process

Such a process is dependent on sufficient numbers of appropriate experts with appropriate access to the subject matter and adequate time to review and critique the relevant studies.

The current IARC process begins on the first day of the monograph meeting when initial drafts are handed out. Scientific discussion is largely confined to the first few meeting days of the WGs when individual studies are reviewed. WG reviews are approved and summaries then written and approved for plenary sessions where occasional editorial changes but minimal discussion of scientific issues is the predominant activity. Inadequate time is provided for participants to review accuracy and completeness of the draft reviews and summaries on which the classification is based. A robust scientific process requires sufficient time for critical evaluations and discussion of strengths and limitations of each study for insertion in the square brackets, and for reviewing the weight of evidence.

Much of this discourse could, and should, be completed prior to the Monograph meeting in Lyon. This would allow time to focus on the most important scientific issues in older studies and for more extended discussion of very recent studies published too late for the initial review. The current process is so limited in time that only a few minutes of reading are available before oral summary, and the subsequent discussion is limited mostly to editorial changes. The current IARC process is conducted largely by the WG chairman, a ghost writer of the summary, and the IARC Secretariat before the Monograph meeting. There is inadequate time allowed for participants and observers to review and discuss scientific issues, particularly for 2 studies accepted

for publication just before or during the meeting [23,32]. Out of my 24 or so days of IARC monograph meeting time, perhaps 3 hours was actually spent on robust scientific discourse.

Suggestions

First drafts for each WG should be distributed at least one month in advance to all invited participants and observers for a written peer-review. These peer-reviews would be returned to IARC for editing and distributed to all participants at least a week before the scheduled monograph meeting in Lyon. This will allow time for robust scientific discussions of the issues and wording in the square brackets at the Lyon meeting. Robust discussions of scientific issues for each study are needed to provide an appropriate basis for evaluating the weight of evidence. This proposed process of a priori peer-review may ameliorate the frenetic pace and curtailed debate that tends to occur with the current process. Another possible change would be to allow a period for everybody to submit comments to the final draft monographs. These comments could be included in the printed monograph in an appendix.

Another possible change is to allow a period for everybody to submit comments to the final draft monographs. These comments will be included in the printed monograph (e.g., in an appendix).

There is a need for scientific summaries and recommendations/classifications of agents that may be associated with cancer in humans. IARC has taken on that responsibility. However, the IARC system is in need of a readjustment to make their reviews and conclusions more independent, transparent, unbiased and scientifically robust. To accomplish this result requires that provisions be made to provide all participants with working drafts that have already been reviewed prior to the Lyon meeting. This would allow adequate time for oral discussion of scientific issues which is particularly needed for newly published studies where the normal scientific review process is impossible because of the limited time. For these newly published studies the WG itself must of necessity conduct the peer-review so the results can be incorporated into the Monograph. Such studies require more careful scrutiny than studies published 6 months or more prior to the Lyon meeting. Recent results suggest the critical assessment by IARC of individual study limitations and the weight of evidence appear inadequate; if so, the IARC conclusions are unreliable.

References

1. Benbrahim-Tallaa L, Baan RA, Yann Grosse, Béatrice Lauby-Secretan, Fatiha El Ghissassi, et al. (2012) Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncol* 13: 663-664.
2. IARC (2012) Diesel and gasoline engine exhausts and some nitroarenes. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon, France, International Agency for Research on Cancer.
3. Marsh GM, Youk AO, Morfeld P, Collins JJ, Symons JM (2010) Incomplete follow-up in the National Cancer Institute's formaldehyde worker study and the impact on subsequent reanalyses and causal evaluations. *Regul Toxicol Pharmacol* 58: 233-236.
4. Gamble JF (2011) Crystalline silica and lung cancer: a critical review of the occupational epidemiology literature of exposure-response studies testing this hypothesis. *Crit Rev Toxicol* 41: 404-465.
5. Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, et al. (2009) A further plea for adherence to the principles underlying science in general and the epidemiologic enterprise in particular. *Int J Epidemiol* 38: 678-679.
6. Coglian V, Straif K (2010) Re: False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 102: 134-135.
7. Crosignani P (2009) Re: False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 101: 212-213.

8. Vineis P (2009) Viewpoint: the skeptical epidemiologist. *Int J Epidemiol* 38: 675-677.
9. Hauptmann M, Ronckers CM (2010) RE: A further plea for adherence to the principles underlying science in general and the epidemiologic enterprise in particular. *Int J Epidemiol* 39: 1677-1679.
10. McLaughlin J, Lipworth L (2010) Authors' response. *Int J Epidemiol* 39: 1679-1680
11. Erren TC (2011) IARC's plea for traditional 'expert' working groups--a recipe for problems? *Int J Epidemiol* 40: 1727-1728.
12. McLaughlin J, Boffetta P (2011) Authors's Response: Problems with IARC's 'expert' working groups. *Int J Epidemiol* 40: 1728-1739.
13. Wild CP, Coglian VJ (2011) A plea on behalf of expert evaluation and the experts involved. *Int J Epidemiol* 40: 253-261.
14. Silverman DT, Samanic CM, Lubin JH, Blair AE, Stewart PA, et al. (2012) The Diesel Exhaust in Miners Study: A nested case-control study of lung cancer and diesel exhaust. *J Natl Cancer Inst* 104: 855-868.
15. Gamble JF, Nicolich MJ, Boffetta P (2012) Lung Cancer and Diesel Exhaust: An updated critical review of the occupational epidemiology literature. *Crit Rev Toxicol* 42: 549-598.
16. Morfeld P (2012) Misspecified and misinterpreted Cox model set up may lead to erroneous conclusions.
17. Crump K, van Landingham C (2012) Evaluation of an exposure assessment used in epidemiological studies of diesel exhaust and lung cancer in underground mines. *Crit Rev Toxicol* 42: 599-612.
18. Crump KS (1999) Lung cancer mortality and diesel exhaust: reanalysis of a retrospective cohort study of U.S. railroad workers. *Inhal Toxicol* 11: 1-17.
19. Hauptmann M, Lubin JH, Stewart PA, Hayes RB, Blair A (2003) Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries. *J Natl Cancer Inst* 95: 1615-1623.
20. Marsh GM, Youk AO (2004) Reevaluation of mortality risks from leukemia in the formaldehyde cohort study of the National Cancer Institute. *Regul Toxicol Pharmacol* 40: 113-124.
21. Marsh GM, Youk AO, Morfeld P (2007) Mis-specified and non-robust mortality risk models for nasopharyngeal cancer in the National Cancer Institute formaldehyde worker cohort study. *Regul Toxicol Pharmacol* 47: 59-67.
22. Beane Freeman LE, Blair A, Lubin JH, Stewart PA, Hayes RB, et al. (2009) Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the national cancer institute cohort. *J Natl Cancer Inst* 101: 751-761.
23. Garshick E, Laden F, Hart JE, Davis ME, Eisen EA, et al. (2012) Lung cancer and elemental carbon exposure in trucking industry workers. *Environ Health Perspect*.
24. Morfeld P (2012) Diesel exhaust in miners study: how to understand the findings? *J Occup Med Toxicol* 7: 10.
25. Olsson AC, Gustavsson P, Kromhout H, Peters S, Vermeulen R, et al. (2011) Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada. *Am J Respir Crit Care Med* 183: 941-948.
26. Villeneuve PJ, Parent MÉ, Sahni V, Johnson KC; Canadian Cancer Registries Epidemiology Research Group (2011) Occupational exposure to diesel and gasoline emissions and lung cancer in Canadian men. *Environ Res* 11: 727-735.
27. Mohner M, Kersten N (2012) Zum Einfluss von Dieselmotoremissionen auf das Lungenkrebsrisiko. *Arbeitsmed Sozialmed Umweltmed* 47: 150.
28. Morfeld P, Erren TC (2012) Exposures to diesel motor exhaust and lung cancer: are findings of a large pooled analysis really consistent? *Am J Respir Crit Care Med* 185: 104-106.
29. Mohner M (2012) The impact of selection bias due to increasing response rates among population controls in occupational case-control studies. *Am J Respir Crit Care Med* 185: 104-106.
30. Olsson A, Vermeulen R (2012) Correspondence: From the authors. *Am J Respir Crit Care Med* 185: 106107.
31. Meek ME, Patterson J, Strawson JE, Liteplo RG (2007) Engaging expert peers in the development of risk assessments. *Risk Anal* 27: 1609-1621.
32. Pintos J, Parent ME, Rousseau MC, Case BW, Siemiatycki J (2008) Occupational exposure to asbestos and man-made vitreous fibers, and risk of lung cancer: evidence from two case-control studies in Montreal, Canada. *J Occup Environ Med* 50: 1273-1281.