



# Open Medicine Journal

Content list available at: [www.benthamopen.com/MEDJ/](http://www.benthamopen.com/MEDJ/)

DOI: 10.2174/1874220301603010238



## ARTICLE COMMENTARY

# Methodological Challenges in the Statistical Analysis of Epidemiology Studies: use of Average Exposure Metrics in Historical Cohort Designs

Thomas B. Starr<sup>1,\*</sup> and Gary M. Marsh<sup>2</sup>

<sup>1</sup>TBS Associates, 7500 Rainwater Road, Raleigh NC 27615-3700, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill NC, USA

<sup>2</sup>Department of Biostatistics, Department of Epidemiology, Department of Clinical & Translational Science, and Center for Occupational Biostatistics & Epidemiology, University of Pittsburgh, Pittsburgh PA, USA

Received: March 22, 2016

Revised: August 31, 2016

Accepted: September 17, 2016

**Abstract:** An important methodological challenge in the analysis of historical occupational cohort data is choosing the most appropriate metric for the average exposure of the workers under study. We describe and illustrate the many issues associated with this challenge using a recent re-analysis by Kopylev [1] of lung cancer mortality in the National Cancer Institute (NCI) acrylonitrile cohort study. Kopylev proposed the routine use of both Average Exposure and Average Intensity when analyzing epidemiological cohort data. However, due to the methodological issues that arise with these metrics, we have concerns about the validity of his finding of a significant positive association between workers' acrylonitrile exposure and increased lung cancer mortality in a subset of the NCI cohort. These include 1) the opportunity for substantial selection bias to have impacted the results; 2) the failure to account properly for latency; 3) the absence of a convincing biological rationale or other *a priori* justification for Kopylev's preferred exposure metrics; 4) the absence of meaningful differences in Average Exposure- and Average Intensity- based risk estimates; 5) the lack of a logical basis for using either of these exposure metrics and 6) the conclusion that smoking was not a significant positive confounder, which is at odds with all other such findings for this cohort.

**Keywords:** Average exposure, Average intensity, Confounding by smoking, Cumulative exposure, Lung cancer mortality, Selection bias.

## INTRODUCTION

An important methodological challenge in the analysis of historical occupational cohort data is choosing the most appropriate metric for average exposure in the evaluation of exposure-response relationships between one or more agents and a health endpoint of interest. Myriad approaches can be used to define average exposure for the workers under study, and the choice of the most appropriate approach should be guided by *a priori* hypotheses about the exposure-response relationship of interest, not by the favored outcome of the evaluation. We have used a recent re-analysis by Kopylev [1] of lung cancer mortality in the National Cancer Institute (NCI) acrylonitrile cohort [2] to describe and illustrate several of the methodological issues associated with this important problem.

## BACKGROUND

Substantial evidence that lung cancer mortality is not associated with acrylonitrile exposure has been obtained from four large acrylonitrile worker cohort studies: NCI (Blair *et al.* [2]), DuPont (Symons *et al.* [3]), Dutch (Swaen *et al.* [4]), and United Kingdom (Benn and Osborne [5]), as well as from independent studies of subcohorts of workers from

\* Address correspondence to this author at the TBS Associates, 7500 Rainwater Road, Raleigh NC 27615-3700, USA; Tel: 919-876-0203; E-mail: [tbstarr@mindspring.com](mailto:tbstarr@mindspring.com)

two of the NCI study sites in Fortier, LA/Santa Rosa, FL (Collins *et al.* [6]) and Lima, OH (Marsh and Zimmerman [7]). Nevertheless, in 2014, Kopylev [1] speculated that a significant positive association may have been missed in the Blair *et al.* [2] analysis because the NCI investigators did not employ the “correct” metric for acrylonitrile exposure. This seems unlikely given that Blair *et al.* [2] considered 19 different exposure metrics in their original report, concluding that none of these showed a strong exposure-response gradient or a statistically significant exposure-response trend in relation to lung cancer mortality.

With no biologically plausible or other *a priori* justification, Kopylev proposed two approaches to the temporal averaging of occupational exposures in the NCI cohort, both of which use traditional Cumulative Exposure as the numerator: 1) Average Intensity, *i.e.*, Cumulative Exposure divided by the Cumulative Duration of non-zero exposure time; and 2) Average Exposure, *i.e.*, Cumulative Exposure divided by Cumulative Duration of employment. Kopylev then fit Cox proportional hazards models to the data from a restricted subset of the NCI cohort (white males with at least 10 years of Time Since First Exposure) using the Cumulative Exposure, Average Intensity, and Average Exposure metrics. While Cumulative Exposure ( $p=0.58$ ) and Average Intensity ( $p=0.15$ ) were not statistically significant predictors of lung cancer mortality, either alone or in combination with birth year and plant, Kopylev found Average Exposure to be a marginally significant predictor of lung cancer mortality both alone ( $p = 0.045$ ) and in combination ( $p = 0.039$ ) with birth year and plant. On this basis, Kopylev recommended that both Average Intensity and Average Exposure be included as additional exposure metric alternatives to Cumulative Exposure when the NCI cohort study is updated, and in other epidemiologic studies as well. The methodological issues in choosing among the alternative average exposure metrics described in the Kopylev report follow.

### **ISSUE 1 - SELECTION OF AN APPROPRIATE POPULATION OF EXPOSED WORKERS**

All of the mortality experience of 1547 exposed white male workers with less than 10 years of Time Since First Exposure appears to have been excluded from Kopylev’s analyses, and this experience includes 18 lung cancer deaths out of the 163 lung cancer deaths that occurred among all white male workers (see Blair *et al.* [2], Table 3). However, the mortality experience of these workers *prior* to their first acrylonitrile exposures should have been included as unexposed experience, along with that of never-exposed workers and the pre-exposure experience of exposed workers with at least 10 years of Time Since First Exposure. Failure to include this additional unexposed experience allows selection bias to impact the modeling results.

### **ISSUE 2 - ACCOUNTING FOR THE LATENT PERIOD BETWEEN FIRST EXPOSURE AND THE HEALTH OUTCOME**

While Kopylev opined that “For lung cancer, a minimum of 10 years is generally considered reasonable for latency with respect to agents that act through tumor initiation and progression”, his analyses do not appear to have taken latency into account. The standard accounting for latency involves lagging the exposure variable by a set number of years, so that, for example, exposures during the most recent 10 years of workers’ experience are not counted in the exposure metric. For example, Blair *et al.* presented [2] results for lung cancer in relation to Cumulative Exposure that were lagged by 5, 10, or 20 years. In contrast, Kopylev’s analyses appear to have included all of the most recent exposure information for the white male workers with a Time Since First Exposure of at least 10 years. Furthermore, in a standard 10 year lag analysis, workers exposed only during their most recent 10 years of work experience are treated as unexposed, while in Kopylev’s analyses, all the work experience of such individuals appears to have been excluded. Restricting attention solely to exposed workers with at least 10 years of Time Since First Exposure and not considering exposure lagging makes Kopylev’s modeling results susceptible to bias and very difficult to interpret.

### **ISSUE 3 - BIOLOGICAL PLAUSIBILITY OF THE METRIC AND REPRODUCIBILITY**

Kopylev provided no convincing biological rationale for his proposed alternative dose metrics, and it is not at all clear how either Average Intensity or Average Exposure could be employed in quantitative cancer risk assessments of environmental exposures. He also provided insufficient detail, both in the description of his methods and his results, to permit validation of his findings by independent replication. Only results from univariate analyses using Cumulative Exposure or Average Intensity were shown for the white male subcohort with a Time Since First Exposure of at least 10 years, and no results at all were provided for the white male subcohort with a Time Since First Exposure of at least 15 years [1]. In addition, Kopylev appears not to have replicated relevant findings from the original analysis of the NCI cohort by Blair *et al.* [2], or those from our subsequent analyses of the entire white male subcohort (Starr *et al.* [8]). Such replication could have established whether or not Kopylev’s data processing and computer programming were

error free. Kopylev also failed to cite the report by Marsh *et al.* [9], who found that internal rate ratios for Cumulative Exposure were positively biased due to an unexplained but statistically significant deficit of lung cancer deaths in the baseline Cumulative Exposure category.

#### **ISSUE 4 - OBSERVING MEANINGFUL DIFFERENCES ACROSS EXPOSURE METRICS**

Examination of Table 1 in Kopylev [1] shows that the descriptive statistics for Cumulative Exposure, Average Intensity and Average Exposure are in extremely close agreement for both the full cohort and the white male subcohort with a Time Since First Exposure of at least 10 years. For example, in the latter subcohort, mean Average Exposure was 0.61 parts per million (ppm), while mean Average Intensity was 0.63 ppm. The corresponding medians are both 0.06 ppm and the interquartile ranges are also virtually identical. Thus, for practical purposes, the differences between Average Exposure and Average Intensity appear trivial: Average Exposure and Average Intensity reflect essentially the same exposure. While the author's limited presentation of modeling results precludes a full comparison, it is not surprising, given the similarity of the Average Exposure and Average Intensity data, that the slope estimates shown in Kopylev's Table 2 for Average Intensity (univariate only) and Average Exposure (univariate and multivariate) are in very close agreement (0.054 ppm<sup>-1</sup>, 0.078 ppm<sup>-1</sup>, and 0.085 ppm<sup>-1</sup>, respectively). Despite the marginally significant slopes for the Average Exposure metric, these slopes are very close to the reported slopes for Average Intensity, and they are also very close to zero. They provide no compelling indication of a positive exposure-response association.

#### **ISSUE 5 - LOGICAL BASIS FOR USING ALTERNATIVE EXPOSURE METRICS**

Kopylev's rationale for using his proposed alternative exposure metrics is flawed logically. In his Introduction, he noted that the ability to calculate Average Intensity requires a certain level of detail in the job-exposure matrix. While not stated explicitly, this detail would need to include the start date, stop date, and the exposure level (in terms of 8 hour time-weighted averages) for each job held by individual workers. Absent such detail, Kopylev claims that the Average Exposure approach is often used, but he offers no examples to establish this as common practice. The logical flaw is that both Average Exposure and Average Intensity require the calculation of Cumulative Exposure, defined for each job held as the product of duration of time in each job and the associated exposure level, summed over all jobs. Thus, if Cumulative Exposure cannot be calculated, then Average Exposure and Average Intensity cannot be calculated.

#### **ISSUE 6 - DO ALTERNATIVE METRICS LEAD TO RESULTS CONSISTENT WITH OTHER STUDIES?**

To his credit, Kopylev employed Richardson's method [10] to assess the potential confounding that may arise from unmeasured smoking by modeling mortality from chronic obstructive pulmonary disease, a smoking-related disease, in relation to Average Exposure in his white male subcohort with at least 10 years of Time Since First Exposure. However, he misinterpreted the marginally negative results ( $\beta=-0.133$ ,  $p=0.50$ , or  $\beta=-0.109$ ,  $p=0.57$ ), depending on how deaths from chronic obstructive pulmonary disease were defined) as somehow providing "strong" evidence against the positive confounding of a lung cancer mortality-Average Exposure association by unmeasured smoking. In our view, strong evidence against positive confounding would be provided by a negative regression coefficient for Average Exposure whose confidence interval excluded zero, not by a central estimate that is essentially null.

Furthermore, Kopylev's conclusion regarding the absence of potential confounding by smoking in the NCI study is not supported by results from a recent Monte Carlo simulation study of workers from one NCI study site in Lima, OH (Zimmerman *et al.* [11]). These investigators found that the acrylonitrile exposure-response relationship for lung cancer that had been suggested in the original Lima cohort study (Marsh *et al.* [12]) was likely to be positively confounded by smoking. In the original NCI study (Blair *et al.* [2]), the Lima, OH (Marsh and Zimmerman [7]) study, and another study of acrylonitrile workers (Collins *et al.* [6]), the prevalence of smoking was found to increase with increasing acrylonitrile exposure, thus creating the potential for positive confounding of any association between acrylonitrile exposure and lung cancer mortality. Furthermore, in a reanalysis of the original NCI nested case-cohort study of lung cancer (Blair *et al.* [2]) that addressed missing and misclassified smoking data, Cunningham and Marsh<sup>1</sup> found still further evidence that lung cancer risk estimates reported in the original full NCI study were positively confounded by smoking.

While Kopylev acknowledged the possibility that no causal relationship may exist between acrylonitrile exposure

---

<sup>1</sup> Cunningham M, Marsh GM. Reanalysis of mortality risks from lung cancer in the acrylonitrile worker cohort study of the National Cancer Institute. M.S. Thesis, University of Pittsburgh, Department of Biostatistics, 2015.

and lung cancer mortality, he also suggested that one may have been missed by previous investigators simply due to the use of inappropriate exposure metrics. We find this suggestion to have little merit, especially given Blair *et al.*'s extensive exposure-response analyses [2] that evaluated 19 different acrylonitrile exposure metrics (including Cumulative Exposure and Average Intensity) and found no evidence for a positive exposure-response relationship. Without a biologically plausible basis or other *a priori* justification, Kopylev found a single marginally significant positive acrylonitrile-lung cancer association by placing a slight, unconventional twist on the standard cohort analysis. Given the lack of precedent and any substantive scientific basis for this finding, we believe it should be regarded as no more than a coincidental finding from an unreplicated exploratory analysis.

## CONCLUSION

The currently ongoing update of the NCI acrylonitrile cohort will provide at least 20-24 more years of mortality follow-up since the first follow-up ended in 1989. This is expected to include a notable increase in the total number of lung cancer deaths, as this relatively young cohort is just now entering the peak lung cancer incidence ages. Although Kopylev's results were derived from a limited exploratory analysis, both of his methods of calculating Average Exposure might possibly be considered for use when the updated data are analyzed. However, it seems unlikely to us that Kopylev's Average Intensity or Average Exposure metrics will produce meaningfully different results given that the NCI update will include no new subjects and no extension of the work histories for original subjects still working at the end of the first follow-up. In our view, the routine use of these exposure metrics in epidemiologic analyses of worker cohorts is currently unjustified, and Kopylev's analyses illustrate the considerable challenge that is involved in selecting the most appropriate exposure metric in occupational epidemiological studies.

## AUTHORS' NOTE

This commentary on methodological issues associated with selection of the most appropriate metric for average exposure in occupational epidemiology studies was originally submitted electronically to *The Open Epidemiology Journal* on 14 July 2015 as a Letter to the Editor. However, that journal was discontinued subsequently, and we only learned this several months after the fact. Mr. Richard E. Morrissy of Bentham Science Publishers suggested that we submit our commentary instead to *The Open Medicine Journal*.

## CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

TBS and GMM received partial financial support for this work from The Acrylonitrile Group, Washington DC.

## REFERENCES

- [1] Kopylev L. Approaches to calculation of average exposure in analysis of epidemiologic cohorts using large acrylonitrile cohort as an example. *Open Epidemiol J* 2014; 7: 1-5. [<http://dx.doi.org/10.2174/1874297120140515001>]
- [2] Blair A, Stewart PA, Zaebst DD, *et al.* Mortality of industrial workers exposed to acrylonitrile. *Scand J Work Environ Health* 1998; 24(Suppl. 2): 25-41. [PMID: 9714511]
- [3] Symons JM, Kreckmann KH, Sakr CJ, Kaplan AM, Leonard RC. Mortality among workers exposed to acrylonitrile in fiber production: an update. *J Occup Environ Med* 2008; 50(5): 550-60. [<http://dx.doi.org/10.1097/JOM.0b013e318162f640>] [PMID: 18469624]
- [4] Swaen GM, Bloemen LJ, Twisk J, *et al.* Mortality update of workers exposed to acrylonitrile in The Netherlands. *J Occup Environ Med* 2004; 46(7): 691-8. [<http://dx.doi.org/10.1097/01.jom.0000128161.17144.27>] [PMID: 15247808]
- [5] Benn T, Osborne K. Mortality of United Kingdom acrylonitrile workers an extended and updated study. *Scand J Work Environ Health* 1998; 24(Suppl. 2): 17-24. [PMID: 9714510]
- [6] Collins JJ, Page LC, Caporossi JC, Utidjian HM, Lucas LJ. Mortality patterns among employees exposed to acrylonitrile. *J Occup Med* 1989; 31(4): 368-71. [PMID: 2715845]
- [7] Marsh GM, Zimmerman SD. Mortality among chemical plant workers exposed to acrylonitrile: 2011 follow-up. *J Occup Environ Med* 2015;

- 57(2): 134-45.  
[<http://dx.doi.org/10.1097/JOM.0000000000000369>] [PMID: 25612296]
- [8] Starr T, Gause C, Youk AO, Stone RA, Marsh GM, Collins JJ. A risk assessment for occupational exposure to acrylonitrile using epidemiology data. *Risk Anal* 2004; 24: 587-601.  
[<http://dx.doi.org/10.1111/j.0272-4332.2004.00461.x>] [PMID: 15209932]
- [9] Marsh GM, Youk AO, Collins JJ. Reevaluation of lung cancer risk in the acrylonitrile cohort study of the National Cancer Institute and the National Institute for Occupational Safety and Health. *Scand J Work Environ Health* 2001; 27(1): 5-13.  
[<http://dx.doi.org/10.5271/sjweh.581>] [PMID: 11266147]
- [10] Richardson DB. Occupational exposures and lung cancer: adjustment for unmeasured confounding by smoking. *Epidemiology* 2010; 21(2): 181-6.  
[<http://dx.doi.org/10.1097/EDE.0b013e3181c6f7d9>] [PMID: 20081541]
- [11] Zimmerman SD, Marsh GM, Youk AO, Talbott EO. Adjustment for confounding by smoking in the presence of misclassified smoking data in a cohort study of workers exposed to formaldehyde. *J Occup Environ Med* 2015; 57: 146-51.  
[<http://dx.doi.org/10.1097/JOM.0000000000000386>] [PMID: 25612297]
- [12] Marsh GM, Gula MJ, Youk AO, Schall LC. Mortality among chemical plant workers exposed to acrylonitrile and other substances. *Am J Ind Med* 1999; 36(4): 423-36.  
[[http://dx.doi.org/10.1002/\(SICI\)1097-0274\(199910\)36:4<423::AID-AJIM3>3.0.CO;2-M](http://dx.doi.org/10.1002/(SICI)1097-0274(199910)36:4<423::AID-AJIM3>3.0.CO;2-M)] [PMID: 10470007]

---

© Starr and Marsh; Licensee *Bentham Open*

This is an open access article licensed under the terms of the Creative Commons Attribution-Non-Commercial 4.0 International Public License (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.