

Approaches to Calculation of Average Exposure in Analysis of Epidemiologic Cohorts Using Large Acrylonitrile Cohort as an Example

Leonid Kopylev*

National Center for Environmental Assessment, Office of Research and Development, United States Environmental Protection Agency, Washington, DC, USA

Abstract: *Objectives:* To explore two different approaches to calculate average exposure in occupational cohorts using a large occupational cohort as an example. The data for occupational cohort exposed to acrylonitrile was collected and analyzed previously by NCI; outcome was lung cancer.

Methods: Both approaches use cumulative exposure as the numerator. As the denominator, one uses the duration of exposure, while the other uses the length of employment. The former approach is used when detailed exposure history is available, and the latter is used when exposure history is less detailed. The differences are investigated for a large occupational cohort.

Results: With restricting the cohort to only those with enough latency for lung cancer, the cumulative exposure divided by the length of employment is a significant predictor of the lung cancer mortality, while cumulative exposure divided by the duration of exposure (average intensity) is not. Analysis is shown not to be positively confounded by smoking.

Conclusion: Both approaches should be considered when calculating an average exposure metric.

Keywords: Acrylonitrile, average intensity, cumulative exposure, JEM.

INTRODUCTION

Cumulative exposure is the most commonly used metric of exposure in epidemiologic studies. Another popular metric of exposure is average exposure. A recent (April 2014) Pubmed search for “cumulative exposure” in humans found 1911 items; an analogous search for “average exposure” in humans produced 631 items. While calculation of cumulative exposure is quite straightforward, the calculation of average exposure allows some choice. The two ways of averaging exposure are dividing cumulative exposure by either the length of exposure (called average intensity in the following), or by the length of employment (called average exposure in the following). The ability to separate those two approaches depends on the level of detail in the job-exposure matrix (JEM). Epidemiologists normally prefer the first approach, but when the level of information regarding exposure on a daily (or yearly, etc.) basis in the JEM is not sufficiently detailed, the second approach is often used, with the implicit assumption that length of employment serves as a proxy for the length of exposure. When workers are exposed for the entire duration of employment, both approaches yield the same result. However, when the JEM includes information on periods during employment when workers are not exposed, both average exposure and average intensity can be calculated and yield differing results.

Blair and coworkers [1] investigated overall and cause-specific mortality (focusing on lung cancer mortality) in a large (25,460 workers) cohort of acrylonitrile workers assembled by NCI and NIOSH scientists (called the NCI acrylonitrile worker cohort below). The cohort was assembled from workers at 8 acrylonitrile production plants in the United States. The level of detail in the JEM in this study is very fine. For about 30% of workers employment history includes periods of exposure and non-exposure. The rest of the workers were either exposed or non-exposed for the entire duration of their employment. The authors evaluated more than 10 metrics of exposure including cumulative exposure and average intensity. None of the exposure metrics was a significant predictor of lung cancer mortality when using continuous exposure metrics, although in categorical analysis the highest cumulative exposure group with at least 20 years since first exposure was significantly different from the non-exposed group. Similarly, a re-analysis [2] investigated lung cancer mortality among white males only (the largest subset by sex/race), and found that cumulative exposure is not a significant predictor of lung cancer mortality. That analysis did not account for time since first exposure. It is, of course, possible that there is no relationship between acrylonitrile exposure and lung cancer mortality. However, a possible reason for lack of significant dose-response relationship could be analytical, e.g. that the appropriate metric of exposure was not found. Although the NCI acrylonitrile worker cohort has the advantage of large size, it has the significant disadvantage of young average age at the end of follow-up (mean 48.5, median 47.8 years) and comprises workers with and without sufficient latency for lung cancer to develop. Latency

*Address correspondence to this author at the Mailstop 8623P1200 Pennsylvania Ave NW Washington, DC 20460, USA; Tel: 703 347 8599; Fax: +1 703 347 8692; E-mail: kopylev.leonid@epa.gov

generally refers to the time between initial exposure to an agent of interest and the development of disease. Because lung cancer mortality (and not incidence) is the endpoint in analyses [1] and [2], the latency here refers to the time to death from lung cancer.

This article considers comparison of the statistical fit of average intensity and average exposure to lung cancer mortality in an epidemiologic study with detailed exposure information [1] as an example. We also discuss how to compensate analytically for some of the design deficiencies of the original study by an appropriate restriction of the cohort to workers with sufficient latency for dying from lung cancer. Influence of smoking on the results is also investigated.

METHODS

Data

We obtained individual-level data published in [1] from the NCI. Basic demographic information on the occupational cohort members was described [1], and the exposure assessment protocol was described in detail [3]. The released data includes individual employment and exposure history, including calendar dates of start and end of periods of exposure, and yearly cumulative exposure data. Vital status follow-up was completed by the NCI through 1989 [1]; 96% of the cohort was successfully traced. Only 8.0% of the workers ($n = 2,038$) had died by December 31, 1989 and death certificates were located for 1,919 workers. Deaths were coded according to the 8th revision of the International Classification of Diseases (ICD) [4] and all 193 lung cancer deaths were coded as ICD-8 code 162.1, “malignant neoplasm of bronchus and lung”. 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 (e.g. [5, 6] and references therein). In our analysis, we considered the full workers cohort, but restricted the cohort to only those with at least 10 years of time since the first exposure (TSFE) as well as at least 15 years of TSFE, and looked at white males only, who constitute by far the largest subset (72.8% for the at least 10-year TSFE cohort, Table 1) of the cohort.

Statistical Analysis

The main statistical feature of the NCI acrylonitrile worker cohort is the availability of time-varying exposure information. The Cox regression model [7] is a common statistical model for the epidemiological analysis of survival and mortality in cohort studies, and the extended Cox model accommodates time-varying exposures. The extended Cox model was also used in a re-analysis of the NCI acrylonitrile worker cohort [2]. We used age as the time-scale for our analyses, as it typically has the strongest relationship to cancer mortality. This controls for age as a risk factor by design, rather than by parametric modeling, and effectively rules out age as a potential confounder. All analyses were conducted using R software [8]. We fit the extended Cox regression model (R software function *coxph*), which included both time-independent factors such as sex, race, plant of employment, and date of birth, and time-dependent measures of acrylonitrile exposure over the entire time course of each individuals’ lifetime, from their date of hiring

until death or loss to follow-up, as well as univariate models including only an exposure variable. The inclusion of date of birth in these analyses controls for any potential birth cohort effect, providing some control for changing smoking habits across time (e.g. [9]). We assessed proportionality by correlating the corresponding set of scaled Schoenfeld residuals with a suitable transformation of time (R software function *cox.zph*), although unlike the standard Cox regression model with time-independent covariates, the extended Cox regression model does not have a strong dependence on the assumption of proportionality [10, 11]. Exposure metrics considered were cumulative exposure, and the two ways of calculating average metric of exposure described above-average intensity and average exposure.

To investigate the impact of possible confounding on the model fit by smoking in the absence of data on smoking in the cohort, we followed the analytical method proposed in [12]. In this approach, the question whether an identified exposure relationship with lung cancer is confounded by unmeasured smoking, is evaluated by monitoring a disease (chronic obstructive pulmonary disease, COPD), which is known to be caused by smoking, but not thought to be related to the exposure (acrylonitrile).

RESULTS

There were 22,678 (89% of the original cohort) workers with at least 10 years of TSFE. Table 1 (middle column) presents demographic and exposure characteristics of these workers. These characteristics are generally similar to those of all workers, with the sub-cohort workers followed-up for slightly longer on average. In the sub-cohort of workers with TSFE of at least 10 years, 99.4% had known vital status, compared with 95.6% in the full cohort. However, it is still a fairly young cohort, with a median age at death/loss to follow-up of 49 years (mean age is 50.3 years) i.e. an age at which background lung cancer mortality is still low (before rapidly increasing in later years).

When we evaluated the proportionality assumption for the Cox proportional hazards model, it was rejected overall ($p < 0.02$), and particularly for the sex covariate. Despite the fact that the extended Cox regression model does not have a strong dependence on the assumption of proportionality [10,11], we restricted the remaining analysis to white males only. When the cohort is restricted to white males (Table 1, right column) with at least 10 years of TSFE ($n=145$ deaths), the test for lack of proportionality was not significant ($p=0.33$). The sub-cohort of white males followed for at least 10 years were generally similar to the full cohort with TSFE of at least 10 year (all races and both genders), with regards to demographic and exposure characteristics. There was a slightly higher proportion exposed (74.5% vs 67.2%) in the sub-cohort, and cumulative and average exposures were also somewhat increased (5.35 vs 4.12 ppm-years and 0.61 vs 0.48 ppm, respectively).

The average exposure metric was significant in both univariate and multivariate models, which included covariates for plant of employment and birth year (Table 2). Cumulative and average intensity exposures were not significant in either univariate (Table 2) or multivariate (results not shown) models. Generally similar results were obtained for white males with TSFE of at least 15 years (results not shown).

Table 1. Demographic and exposure characteristics of the NCI acrylonitrile worker cohort with TSFE of at least 10 years.

	Full Cohort with TSFE of at Least 10 Yrs n=22,678			White Males with TSFE of at Least 10 Yrs n=16,532		
Deaths from all causes (n)	1597			1337		
Deaths from lung cancer (n)	171			145		
Deaths from COPD (n)	40			35		
Sex and ethnicity (n, %)						
Men, white	16,532 (72.8)			16,532 (100)		
Men, other ethnicity	1,850 (8.2)					
Women, white	3,577 (15.8)					
Women, other ethnicity	719 (3.2)					
Vital status (n, %)						
Alive	20965 (92.4)			15,109 (91.4)		
Deceased, confirmed	1,597 (7.0)			1,337 (8.1)		
Deceased, unconfirmed	30 (0.1)			27 (0.2)		
Unknown	86 (0.4)			59 (0.4)		
Exposure status (n, %)						
Unexposed	7,435 (32.8)			4,210 (25.5)		
Exposed	15,243 (67.2)			12,322 (74.5)		
	Mean	[25 th , 50 th , 75 th Percentile]	(Minimum, Maximum)	Mean	[25 th , 50 th , 75 th Percentile]	(Minimum, Maximum)
Year of birth	1939	[1930, 1940, 1947]	(1888, 1963)	1938	[1930, 1939, 1946]	(1888, 1963)
Year of hire	1966	[1958, 1966, 1973]	(1942, 1980)	1964	[1957, 1965, 1970]	(1942, 1980)
Age at hire (years)	27.0	[21.2, 24.7, 31.1]	(14.2, 66.9)	27.0	[21.3, 24.7, 31.0]	(14.2, 63.2)
Years of follow-up	23.3	[16.3, 23.1, 29.6]	(10.0, 47.6)	24.2	[18.8, 24.3, 30.7]	(10.0, 47.6)
Employment duration (years)	8.9	[1.3, 5.3, 15.4]	(0, 35.4)	10.0	[1.7, 6.6, 17.2]	(0, 32.6)
Cumulative exposure (ppm-years)	4.12	[0, 0.09, 1.39]	(0, 284)	5.35	[0, 0.20, 2.36]	(0, 284)
Average exposure (ppm)	0.48	[0, 0.03, 0.24]	(0, 18.78)	0.61	[0, 0.06, 0.35]	(0, 18.78)
Average intensity (ppm)	0.50	[0, 0.03, 0.25]	(0, 18.78)	0.63	[0, 0.06, 0.36]	(0, 18.78)

It is well known that smoking is a strong independent risk factor for lung cancer, but there is no evidence of a synergistic effect with acrylonitrile exposure. For analysis of confounding by smoking, following methodology by [12], we modeled the potential effects of acrylonitrile exposure on the risk of COPD mortality. Since all deaths in the NCI acrylonitrile worker cohort were coded in 8th revision of ICD [4], the definition of COPD included the following ICD-8 codes: 491 (“chronic bronchitis”), 492 (“emphysema”) and code 519.3 (“chronic obstructive lung disease without mention of asthma, bronchitis, or emphysema”); this last code was introduced in the United States and Canada during the time that the 8th revision of the ICD codes was in use [13]. There were 35 COPD deaths among white males with at least 10 years of follow-up (3 deaths with the code “491”, 12 deaths with the code “492”, and 20 deaths with the code “519.3”). Using the average exposure metric, the Cox proportional hazards model with time-dependent exposures estimated a slope for COPD of -0.133 (p-value=0.50). Because, ICD-8 code “491” may not be specific for COPD, the analysis was repeated with only the COPD deaths coded “492” and “519.3” (32 total COPD deaths) and results were similar (slope -0.109, p-value=0.57). The fact that the coefficients for exposure in the COPD Cox models were

negative is strong evidence against potential for positive confounding of the relationship between acrylonitrile exposure and lung cancer mortality; smoking is positively related to COPD risk and thus, if positive confounding is occurring, we would also expect the relationship between acrylonitrile exposure and COPD risk to be positive. However, because the coefficient for exposure is negative, it is possible that negative confounding is occurring, in which case the risk of lung cancer associated with acrylonitrile exposure would be understated.

DISCUSSION

Using lung cancer mortality analysis of the NCI acrylonitrile worker cohort [1] as an example, we illustrated that the calculation of an average exposure with length of the employment as the denominator provides a better fit than the average intensity with length of exposure as the denominator. It is also important to note that, for this cohort, previous analyses [1, 2] with cumulative exposure showed no dose-response relationship with lung cancer mortality; nevertheless, the average exposure, calculated by dividing cumulative exposure by the length of the employment, shows a significant dose-response. We also demonstrated that these results are not confounded by smoking and not

Table 2. Results for analysis of cumulative and average metrics of exposure as predictor of lung cancer risk in the White males with TSFE of at least 10 years.

Variable	Univariate	Multivariate		Univariate	Univariate
Exposure		Average	Average	Cumulative	Av. Intensity
Exposure		0.078	0.085	2.2e-3	0.054
Slope (SE)		(0.039)	(0.041)	(3.8e-3)	(0.038)
p-value		0.045	0.039	0.58	0.15
Birthyear			0.007		
Slope (SE)			(0.013)		
p-value			0.58		
Plant 2			-0.55		
Slope (SE)			(0.80)		
p-value			0.49		
Plant 3			0.42		
Slope (SE)			(0.54)		
p-value			0.44		
Plant 4			0.96		
Slope (SE)			(0.46)		
p-value			0.037		
Plant 5			0.33		
Slope (SE)			(0.42)		
p-value			0.43		
Plant 6			0.84		
Slope (SE)			(0.44)		
p-value			0.053		
Plant 7			0.57		
Slope (SE)			(0.50)		
p-value			0.26		
Plant 8			0.61		
Slope (SE)			(0.41)		
p-value			0.13		

sensitive to inclusion of additional covariates, such as plant of employment information.

Another key consideration here is latency; only those members of the cohort with sufficient latency for lung cancer should be included in analysis of lung cancer mortality. It is desirable that this cohort be updated, since at least 20 years of follow-up would be added since 1989, and we recommend that both methods of calculating average exposure be considered when the updated data are analyzed. It is likely that there would be a notable increase in new lung cancer cases with additional years of follow-up, as this relatively young cohort is entering peak lung cancer incidence age, but how that would impact relative fit of different exposure metrics cannot be determined from the available data.

It is not clear what biological considerations, if any, may explain the results observed here. A possible explanation may be that average exposure better approximates an internal metric of exposure compared to other exposure metrics (including cumulative exposure). Another relevant consideration is lingering effect [14], and cumulative

exposure that combines concentration and duration equally may ignore effect of lingering. Results observed here using the example of the large occupational cohort, need to be confirmed in analyses of other large cohorts, but in any case, a reasonable strategy would be to calculate average exposure both ways in analyses of epidemiologic cohorts with detailed enough exposure information.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The author thanks Krista Christensen and Glinda Cooper of US EPA for valuable advice.

DISCLAIMER

The views expressed in this manuscript are those of the author and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

REFERENCES

- [1] Blair A, Stewart PA, Zaubst DD, *et al.* Mortality of industrial workers exposed to acrylonitrile. *Scand J Work Environ Health* 1998; 24(Suppl 2): 25-41.
- [2] Starr TB, Gause C, Youk AD, Stone R, Marsh GM, Collins JJ. A risk assessment for occupational acrylonitrile exposure using epidemiology data. *Risk Anal* 2004; 24: 587-601.
- [3] Stewart PA, Zaubst D, Zey JN, *et al.* Exposure assessment for a study of workers exposed to acrylonitrile. *Scand J Work Environ Health* 1998; 24: 42-53.
- [4] National Center for Health Statistics. 8th revision, International Classification of Diseases: Adapted for use in the United States. Public Health Service 1968: Publication No. 1693.
- [5] Nadler DR, Zurbenko LG. Developing a Weibull model extension to estimate cancer latency, *ISRN Epidemiology* 2013; 2013: Article ID 750857, 6 pages.
Available from: <http://www.hindawi.com/journals/isrn/2013/750857/>
- [6] Fakir H, Hofmann W, Sachs RK. Modeling progression in radiation-induced lung adenocarcinomas. *Radiat Environ Biophys* 2010; 49(2): 169-76.
- [7] Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol* 1972; 34: 187-220.
- [8] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
Available from: <http://www.R-project.org/>. 2011
- [9] Hein MJ, Stayner LT, Lehman E, Dement JM. Follow-up study of chrysotile textile workers: Cohort mortality and exposure-response. *Occup Environ Med* 2007; 64: 616-25.
- [10] Tableman M, Kim JS. Survival analysis using S: Analysis of time-to-event data. Virginia Beach, VA: Chapman and Hall/CRC Press 2004.
- [11] Kleinbaum DG, Klein M. Survival analysis: a self-learning text. New York: Springer-Verlag 1996.
- [12] Richardson DB. Occupational exposures and lung cancer: Adjustment for unmeasured confounding by smoking. *Epidemiology* 2010; 21: 181-6.
- [13] National Center for Health Statistics. Estimates of selected comparability ratios based on dual coding of 1976 death certificates by the 8th and 9th revisions of the international classification of diseases. *Public Health Service* 1980; 28 (Suppl 11).
- [14] Chen C. Lingering effect: epidemiological information useful for risk assessment. *Regul Toxicol Pharmacol* 2008; 52: 242-7.

Received: January 23, 2014

Revised: May 7, 2014

Accepted: May 9, 2014

© Leonid Kopylev; Licensee *Bentham Open*.This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.