

Inter-Rater Agreement for a Retrospective Exposure Assessment of Asbestos, Chromium, Nickel and Welding Fumes in a Study of Lung Cancer and Ionizing Radiation

E. A. SEEL^{1*}, D. D. ZAEBST¹, M. J. HEIN¹, J. LIU², S. J. NOWLIN¹ and P. CHEN¹

¹National Institute for Occupational Safety and Health, Division of Surveillance, Hazard Evaluations, and Field Studies, 4676 Columbia Parkway, Cincinnati, OH 45226, USA; ²Westat Inc., 1650 Research Boulevard, Rockville, MD 20850, USA

Received 21 November 2006; in final form 10 July 2007; published online 10 September 2007

A retrospective exposure assessment of asbestos, welding fumes, chromium and nickel (in welding fumes) was conducted at the Portsmouth Naval Shipyard for a nested case-control study of lung cancer risk from external ionizing radiation. These four contaminants were included because of their potential to confound or modify the effect of a lung cancer-radiation relationship. The exposure assessment included three experienced industrial hygienists from the shipyard who independently assessed exposures for 3519 shop/job/time period combinations. A consensus process was used to resolve estimates with large differences. Final exposure estimates were linked to employment histories of the 4388 study subjects to calculate their cumulative exposures. Inter-rater agreement analyses were performed on the original estimates to better understand the estimation process. Although concordance was good to excellent (78–99%) for intensity estimates and excellent (96–99%) for frequency estimates, overall simple kappa statistics indicated only slight agreement beyond chance ($\kappa < 0.2$). Unbalanced distributions of exposure estimates partly contributed to the weak observed overall inter-rater agreement. Pairwise weighted kappa statistics revealed better agreement between two of the three panelists ($\kappa = 0.19$ – 0.65). The final consensus estimates were similar to the estimates made by these same two panelists. Overall welding fume exposures were fairly stable across time at the shipyard while asbestos exposures were higher in the early years and fell in the mid-1970s. Mean cumulative exposure for all study subjects was 520 fiber-days cc^{-1} for asbestos and 1000 mg-days m^{-3} for welding fumes. Mean exposure was much lower for nickel (140 μg -days m^{-3}) and chromium (45 μg -days m^{-3}). Asbestos and welding fume exposure estimates were positively associated with lung cancer in the nested case-control study. The radiation-lung cancer relationship was attenuated by the inclusion of these two confounders. This exposure assessment provided exposure estimates that aided in understanding of the lung cancer-radiation relationship at the shipyard.

Keywords: expert judgment; inter-rater agreement; job-exposure matrix; retrospective exposure assessment

INTRODUCTION

A 1978 cohort mortality study at the Portsmouth Naval Shipyard (PNS) observed excess mortality from all cancers combined (Najarian and Colton, 1978). The National Institute for Occupational Safety and

Health (NIOSH) embarked on a series of cohort mortality and nested case-control studies to assess the relationship between exposure to external ionizing radiation at PNS and mortality from certain diseases. An initial cohort mortality study of white male PNS workers ($n = 24\,545$) with vital status follow-up through August 1977 did not find an excess of all cancers (Rinsky *et al.*, 1981). This study detected a relationship with lung cancer (Rinsky *et al.*, 1980).

*Author to whom correspondence should be addressed.
Tel: +513-841-4422; fax: +513-841-4470;
e-mail: lseel@cdc.gov

A nested case-control study found a significant excess of lung cancer mortality among PNS workers exposed to radiation which was reduced and no longer statistically significant after controlling for asbestos and welding fumes (Rinsky *et al.*, 1988). A mortality study of an expanded cohort of PNS workers (males and females of all races, $n = 37\ 853$) with vital status follow-up through 1996 observed elevated mortality from lung cancer among radiation-exposed workers, but did not control for possible confounding from additional occupational exposures (Silver *et al.*, 2004). A subsequent mortality study of radiation-monitored PNS workers found a positive (albeit non-significant) excess relative risk of lung cancer with external radiation dose that became negative (and remained non-significant) after controlling for socioeconomic status and asbestos and welding fume exposures (Yiin *et al.*, 2005). The retrospective exposure assessment effort for asbestos, welding fumes (iron oxide), chromium and nickel (in welding fumes) described here was part of an updated nested case-control study of external ionizing radiation exposure and lung cancer risk at PNS (Yiin *et al.*, 2007). These four contaminants were included because of their confirmed or potential contribution to lung cancer mortality and their presence at the shipyard (Selikoff and Lee, 1978; Doll *et al.*, 1990; Langard, 1990; Antonini, 2003). This exposure assessment used more rigorous methods to more accurately and thoroughly assess and measure these confounders than did the previous studies. The current exposure assessment also benefited from access to a wealth of industrial hygiene records from the shipyard that were not available at the time of the previous study.

PNS, located in Kittery, Maine was established in 1800 as the first US Navy shipyard. Nuclear submarines were built at PNS from 1958 until 1969. When submarine construction ended the shipyard continued to overhaul, repair and refuel existing submarines. Asbestos was used at the shipyard to insulate boilers and hot and cold water pipes in buildings, ships and submarines. Protective aprons, gloves, millboard, thermal curtains and blankets used for welding activities also contained asbestos. In the mid-1970s, asbestos was phased out and removed from the shipyard. Welding, brazing and cutting structural members of hull sections was a major source of welding fumes. Chromium and nickel exposures were related to the use of stainless steel in submarine construction.

The assessment of exposure to these four substances for the nested case-control study required retrospective methods since potential for exposure began in 1945 for some study subjects and for others extended through the 1990s. A retrospective exposure assessment of solvents in the automobile industry inspired some of the methods used in this study (Macaluso *et al.*, 1993). Macaluso assembled a five-member in-

dustrial hygiene panel and developed a site-specific department/job/year exposure matrix from employment histories. Panelists estimated exposure intensity using six semiquantitative categories. The PNS exposure assessment included a three-member industrial hygiene panel very familiar with the shipyard's activities and exposures. Shop/job/year exposure matrices were also based on employment histories of study subjects. Exposure intensity categories were created for all four substances and panel members independently completed exposure matrices for them. Like Macaluso, concordance criteria were established to distinguish similar (concordant) and dissimilar (discordant) estimates. Unlike Macaluso, the PNS exposure assessment included panel consensus meetings to resolve discordant estimates. The purpose of this paper is to describe these methods, the inter-rater agreement of the three expert panelists and the exposure estimates.

METHODS

Employment history

Available employment histories for 4388 subjects in the nested case-control study were compiled into an employment history database. This database was used to identify shops, job titles and time periods of the study population to develop exposure matrices for the four contaminants. Shops are occupational groups describing major shipyard tasks such as maintenance, transportation and quality control, while job titles describe specific functions like electrician, mechanic and welder. The database contained 349 original shop names and 2288 original job titles which reduced to 57 unique shop names and 238 unique job titles after eliminating redundancies due to spelling variations, abbreviations and word order. Different job titles with similar functions were also grouped (e.g. machinist, machinist apprentice and machinist helper were grouped into a single machinist job title). Job titles were not combined if there was any uncertainty about the similarity of exposures to the four contaminants. This resulted in 1174 unique final standardized shop/job combinations. Work histories of study subjects began in 1945 and continued until 1996. When years of employment were considered the number of unique shop/job/year combinations was >15 000 cells per matrix. Longer increments of time were considered to reduce the number of estimates needed to a manageable number without losing precision. Three-year time periods were chosen, reducing the number of potential estimates for each shop/job combination from 51 to 18. The four exposure matrices each consisted of 3519 cells.

Exposure data

A second database of results of air sampling performed at PNS supported the development of the

numeric values for the asbestos intensity categories. This database was constructed from industrial hygiene records of air sampling results from the 1940s to the 1990s collected from the shipyard. The database contained 2558 asbestos, but few chromium, nickel and iron oxide in welding fume sample results. Because air sampling at the shipyard was complaint based up until the mid-1970s, and then compliance-driven, samples were not collected systematically to fully portray exposures at the shipyard. Much of the sampling occurred in the 1980s and 1990s, a job title was rarely associated with an air sample, and only four shops were sampled before 1975.

NIOSH conducted independent air sampling at PNS in 1979 and 1980 that included 89 asbestos samples. All but nine were area samples representing a few shops. Personal samples were associated with a job-coding scheme that could not be related to the job titles found in employment histories. It was concluded that these two sources of air sampling data were not distributed sufficiently among job titles, shops or time periods to directly contribute to this exposure assessment.

Over 500 original shipyard documents containing information about processes, controls, respirator use, ventilation and work practices that could impact exposures across time at the shipyard were compiled into a third database. Two timelines and a summary of this information, not initially provided to the panelists for their independent assessments, were later used to support the consensus process.

Exposure intensity

Exposure intensity was defined as the time-weighted exposure for an 8-h workday. For asbestos, four 'exposed' categories were developed using 915 PNS asbestos air sample results (sample time >6 h). Most of the results ($n = 852$) were below the limit of detection of 0.004 fibers cc^{-1} (f cc^{-1}); the remaining 63 samples ranged from 0.004 to 25.0 f cc^{-1} . Quartiles of the 63 samples (0.35, 1.0 and 3.0 f cc^{-1}) were used as boundaries for the four categories and the arithmetic means of all 915 samples within the

quartiles were used for the asbestos exposure intensities (Table 1). A fifth exposure category representing background was developed from information in the documents database suggesting that ambient air levels prior to the late 1970s were higher at the shipyard because of lacking or ineffective ventilation. The industrial hygiene panel verified that little ventilation existed in the early years and when installed, these systems exhausted into the ambient air unfiltered until the mid-1970s. The mean of the ambient asbestos air samples (0.004 f cc^{-1}) was used for the background intensity. In order to account for improved ventilation and filtration systems at the shipyard, 0.002 f cc^{-1} was considered background for 1978 and later. These values were similar to estimates in a previous study of asbestos exposures conducted at another navy shipyard in the 1980s that assessed background and bystander air levels (Mangold *et al.*, 2006).

Three exposure intensity categories (background, low and high) were developed for welding fumes, chromium and nickel due to the paucity of sampling data and program information for them at PNS. Numeric concentrations were assigned to these intensity categories using fractions of current occupational exposure limits (Coble *et al.*, 1997) and the expertise of the industrial hygiene panel. The panel examined a range of concentrations for each category based on occupational exposure limits for each contaminant. Panelists independently selected a concentration from each range to represent the 'midpoint' of each intensity category. For all categories and contaminants, at least two panelists picked the same value and the panel majority was used as the numeric intensity (Table 1). Background intensity for chromium and nickel was set to 0.0 $\mu\text{g m}^{-3}$ but 0.025 mg m^{-3} was used for welding fumes because like asbestos, welding fumes were more widespread at the shipyard. For comparison purposes, the American Conference of Governmental Industrial Hygienists (ACGIH) Threshold Limit Values (TLV) for asbestos, welding fumes and chromium are provided in Table 1 (ACGIH, 2005). The NIOSH recommended exposure limit is provided for nickel due to its considerably lower value when compared to the TLV (NIOSH, 2005).

Table 1. Numeric values for asbestos, welding fumes, chromium and nickel intensity categories^a

Contaminant	Intensity category					Reference value
	Background	Very low	Low	Medium	High	
Asbestos (f cc^{-1})	0.004 ^b	0.03	0.75	1.85	8.09	0.1 ^c
Welding fumes (mg m^{-3})	0.025	—	2.5	—	10	5 ^c
Chromium ($\mu\text{g m}^{-3}$)	0 ^d	—	4.9	—	20	10 ^c
Nickel ($\mu\text{g m}^{-3}$)	0 ^d	—	9.0	—	40	15 ^e

^aNumeric values for asbestos are the arithmetic means of the air sample results that fell within the boundaries of each category.

^bThe numeric value for background intensity was 0.004 f cc^{-1} prior to 1978 and 0.002 f cc^{-1} in 1978 and later.

^cACGIH TLV.

^dThese values were 0 because chromium and nickel exposures were more isolated and contained at PNS than asbestos and welding fumes.

^eNIOSH recommended exposure limit.

Additional exposure metrics

Four additional exposure metrics were included in this exposure assessment. Exposure frequency was estimated as the number of 8-h days (from 1 to 240) per working year (excluding weekends and holidays) at the estimated intensity, with the balance of days assigned to background. An intensity range was estimated to be the lowest and highest exposure intensities (for an 8-h workday) experienced by the majority of workers in a shop/job/year combination. Respirator use and respirator type were also estimated by the panel for asbestos exposures only. For respirator use, panelists were asked if respirators were available to employees, required to be worn by employees or not available.

Industrial hygiene panel

Three industrial hygienists each with 25 years of experience with exposures and activities at PNS served as an expert panel for this study. Panel members independently assigned estimates for the exposure metrics in the four exposure matrices. Panelists were provided with detailed instructions, examples for completing cells in the matrices and lists of original shops/job titles and their corresponding standardized shops/job titles. A series of quality control steps were implemented prior to finalizing the matrices to ensure that they accurately reflected the estimates of each panelist.

Once all 12 exposure matrices were finalized, each exposure metric (intensity, frequency, exposure intensity range, respirator use and respirator type) was compared for agreement among the three panelists. Estimates that were identical for all three panelists became final. Results that differed among the panelists were divided into two categories: concordant and discordant. Concordance was defined differently for each exposure metric. Intensity estimates were concordant when they differed by no more than one exposure category. Exposure intensity range was handled the same way. Frequency estimates were concordant when they differed by <80 days. For respirator use and respirator type, any differences were discordant. When estimates were concordant, rules were established for finalizing them. For intensity and intensity range estimates, the majority estimate became the final estimate; for frequency, the average of the three estimates became the final estimate. Once the rules were applied to all concordant estimates, they were finalized and placed in the matrices, leaving only discordant estimates to reconcile.

The industrial hygiene panel met to discuss all estimates that were discordant. Information from the documents database, not provided during initial estimation, was available for reference during these consensus meetings. For all discordances, the panel arrived at final estimates through a consensus pro-

cess. These consensus estimates were added to the four exposure matrices making them complete.

Respirator protection factor

According to the industrial hygiene panel respirator use at PNS did not consistently occur until the mid-1970s. If respirators were available earlier, no confidence existed that they would have been fitted or worn properly. Respirators were used to protect workers from asbestos exposure but not welding fumes, chromium or nickel. Therefore, the panel only estimated respirator use/type for asbestos exposures. Adjustments to exposure estimates using the Occupational Safety and Health Administration's (OSHA) assigned protection factors (APFs) (OSHA, 1998) were applied only when the panel agreed that respirators were required.

Additional estimates

A small number of ineligible study subjects were replaced after the exposure matrices were finalized; consequently, some shop/job/time period combinations had not been assessed for exposure by the panel. Estimates were made for these shop/job combinations using the closest time periods for the same shop/job combination. If that shop/job combination did not exist at all in the matrices, professional judgment was used to determine the most similar shop with that job title and the closest time period for that shop/job combination was used. Additionally, employment histories for 743 study subjects included federal employment locations other than PNS. This offsite work time accounted for 9% of the overall person-years for study subjects (9.4% for controls and 8.4% for cases). An assessment of the employment history database identified 26 standardized job titles with the potential for exposure to the four contaminants. Exposures were similarly interpolated based on professional judgment of the estimates already in the exposure matrices.

Cumulative exposure

Exposures to asbestos, chromium, nickel and welding fumes were cumulated based on the employment history of each study subject and the daily exposure estimates in the exposure matrices. An annualized exposure estimate was computed for every shop/job/time period combination in each exposure matrix and was the product of intensity, frequency and the APF (the latter only for the asbestos matrix) for each cell. If the frequency of exposure was <240 days, the remaining days were assigned background so that all 240 days were accounted for. The annualized exposure estimate was divided into a daily exposure estimate which was linked to the employment history to determine cumulative exposure for a study subject. Annualized exposure estimates were calculated in units of

fiber-days cc^{-1} for asbestos, mg-days m^{-3} for welding fumes and $\mu\text{g-days m}^{-3}$ for chromium and nickel for each shop/job/year combination.

Statistical analyses

Raw agreement of the original estimates was assessed by calculating percent concordance. Raw agreement analyses, however, did not take into account agreement that may have occurred due to chance, therefore, additional inter-rater agreement analyses included overall and pairwise kappa statistics. Overall kappa statistics were calculated to assess agreement among all three panelists. Pairwise kappa statistics were calculated to assess agreement between two panelists at a time and used quadratic weights based on the intensity scores (Fleiss and Cohen, 1973). These weights treat small and large differences in estimates differently. The intraclass correlation coefficient (ICC) was used to assess agreement between the annualized estimates. ICCs were calculated using methods presented in Shrout and Fleiss (1979). ICC (2,1), which assumes that the panelists represent a random sample of panelists available from a larger population, was used to assess agreement of the annualized exposure estimates among the panelists. The strength of the inter-rater agreement was qualified using terms defined by Landis and Koch: results 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1.00 almost perfect (Landis and Koch, 1977). All statistical analyses were performed using SAS 9 Software (SAS Institute Inc., Cary, NC).

RESULTS

Asbestos exposure intensity estimates of the three panelists were highly concentrated in the lowest two

exposure categories (background and very low) and panelists 2 and 3 assigned no shop/job/time period combinations to the highest exposure intensity category (Table 2). For asbestos, 82% of the final intensity estimates (based on consensus) were rated background. For welding fumes, panelists 2 and 3 assigned most of their estimates to background and 90% of the final intensity estimates were background. Nickel and chromium intensity estimates among the three panelists were nearly identical. Most intensity estimates fell into the background category for chromium and nickel and panelists 1 and 3 assigned no shop/job/time period combinations to the highest exposure intensity category.

The majority of panel estimates of exposure intensity met the concordance criteria (differences of no more than one category); however, the overall simple kappa statistics for the intensity estimates indicated only slight agreement for all four contaminants (Table 3). For all contaminants, pairwise weighted kappas for panelists 1 and 2 and panelists 1 and 3 indicated only slight agreement; agreement was fair to substantial for panelists 2 and 3 (Table 3). For panelists 2 and 3, discordant asbestos intensity estimates rarely differed by more than one category; hence the pairwise weighted kappa was higher (0.42, Table 3) than the unweighted pairwise kappa (0.22, not shown).

The frequency metric indicated the number of days of exposure per year based on a 240-day work-year. Panelists were free to choose any number of days out of 240, but assigned frequency estimates using just 14 values (5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 90, 100, 120 and 240). The value of 240 days was only associated with background intensity. To assess agreement of frequency of exposure above background, zero was substituted for 240 and the 14 values were collapsed into four categories (0, 5–30,

Table 2. Distribution of original and consensus intensity estimates for 3519 shop/job/time period combinations

Contaminant	Intensity category	Original estimates			Consensus estimates
		Panelist 1	Panelist 2	Panelist 3	
Asbestos	Background	2246	2703	3243	2901
	Very low	894	797	48	491
	Low	267	9	205	78
	Medium	46	10	23	24
	High	66	0	0	25
Welding fumes	Background	1192	3103	3382	3184
	Low	1479	360	99	284
	High	848	56	38	51
Chromium	Background	3328	3444	3481	3492
	Low	191	43	38	27
	High	0	32	0	0
Nickel	Background	3330	3444	3481	3479
	Low	189	43	38	40
	High	0	32	0	0

Table 3. Percent concordance and inter-rater reliability for 3519 shop/job/time period combinations among three panelists

Metric contaminant	Percent concordance (%) ^a	Inter-rater reliability ^b			
		1 versus 2	1 versus 3	2 versus 3	Overall
Exposure intensity					
Asbestos	90.6	0.09	0.18	0.42	0.17
Welding fumes	77.9	0.11	0.06	0.65	0.05
Chromium	99.3	0.09	0.07	0.29	0.09
Nickel	99.3	0.09	0.07	0.27	0.09
Exposure frequency					
Asbestos	96	0.34	0.27	0.41	0.19
Welding fumes	98	0.12	0.02	0.19	0.08
Chromium	>99	0.09	0.11	0.51	0.11
Nickel	>99	0.09	0.11	0.51	0.11
Respirator usage					
Asbestos	90.9	0.16	0.18	0.29	0.21
Annualized exposure estimate					
Asbestos	—	0.32	0.44	0.49	0.40
Welding fumes	—	0.08	0.05	0.54	0.09
Chromium	—	0.07	0.08	0.56	0.16
Nickel	—	0.07	0.08	0.56	0.16

^aEstimated exposure intensities were concordant if they differed by no more than one exposure level; estimated exposure frequencies were concordant if they differed by <80 days; estimates of respirator usage were concordant if they did not differ.

^bFor estimated exposure intensity and frequency, pairwise estimates of inter-rater reliability (1 versus 2, 1 versus 3 and 2 versus 3) are quadratic weighted (Fleiss–Cohen weights) kappa statistics and the overall estimate is a simple (unweighted) kappa statistic; for respirator usage, pairwise and overall estimates of inter-rater reliability are simple (unweighted) kappa statistics and for the annualized exposure estimate, pairwise and overall estimates of inter-rater reliability are ICCs.

40–75 and 90–120). Overall simple kappa statistics for the frequency estimates indicated only slight agreement for all four contaminants (Table 3). Pairwise weighted kappas for asbestos indicated slight to moderate agreement for all pairs of panelists. Agreement between panelists 2 and 3 was moderate for both chromium and nickel.

Respirator use concordance was high while the overall simple kappa indicated barely fair agreement (Table 3). Pairwise kappa statistics were similar and slight between panelists 1 and the other two panelists. Agreement was again higher and fair for panelists 2 and 3.

The overall ICC for 3519 annualized asbestos exposure estimates indicated fair agreement; for welding fumes, chromium and nickel, agreement was slight (Table 3). Pairwise ICCs for asbestos ranged from fair to moderate; for welding fumes, chromium and nickel, it was slight between panelist 1 and the other two panelists, and moderate between panelists 2 and 3.

Mean annual asbestos and welding fume exposure estimates of the three panelists and the final exposure estimates were graphed by time period for all shop/job combinations (Figs 1 and 2, respectively). While these graphs are a broad sweeping view of the entirety of the estimates, they display the initial exposure estimates and the impact of the consensus process on the final estimates. Final welding estimates were similar to the original estimates of panel-

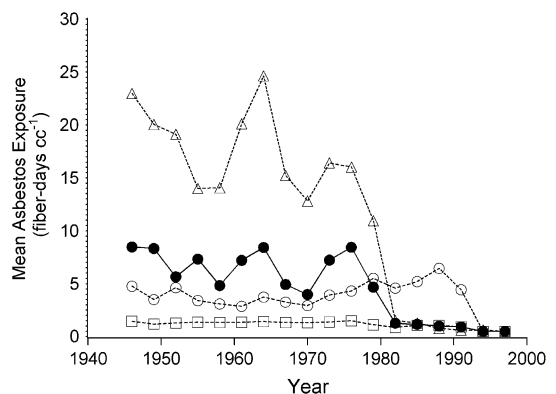


Fig. 1. Mean yearly initial panel (open triangle, panelist 1; open square, panelist 2; open circle, panelist 3) and final (closed circle) asbestos exposure estimates (fiber-days cc^{-1}) based on concordance rules and consensus meetings for 3519 shop/job/year combinations.

ists 2 and 3 while final asbestos estimates were slightly higher than estimates of panelists 2 and 3.

An assessment of the 10 highest asbestos exposed shop/jobs revealed that 19% ($n = 844$) of study subjects held at least one of these highly exposed shop/jobs, impacting 10% of the total person-years of employment in the study. Pipefitters in the pipe shop and welders in the welding shop had the greatest percentage of highly exposed workers (11% and 6%, respectively) and person-years (5% and 4%, respectively).

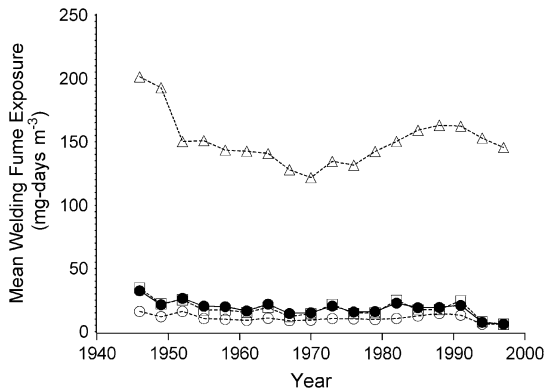


Fig. 2. Mean yearly initial panel (open triangle, panelist 1; open square, panelist 2; open circle, panelist 3) and final (closed circle) welding fume exposure estimates (mg-days m^{-3}) based on concordance rules and consensus meetings for 3519 shop/job/year combinations.

For welding fumes, 8% of study subjects were in the top 10 exposed shop/jobs (5% of total person-years). Welders in the welding shop represented the greatest percentage of study subjects (7%) and total person-years (4%) in the highest exposed shop/jobs.

Figures 3 and 4 display the exposure profiles of four job titles (pipefitters, welders, electricians and sheet metal mechanics) to asbestos and welding fumes based on the final consensus estimates. These job titles were chosen to depict a variety of exposures at the shipyard over time. Welding fume exposures remained fairly constant over time while asbestos exposures declined in the 1970s due to the phasing out of asbestos at the shipyard.

Of the 4388 subjects in the case-control study, 2796 were ever exposed to asbestos above the background intensity level during their work history. Similarly 2318 subjects were ever exposed to welding fumes above background. The mean cumulative exposure of the study subjects to asbestos and welding fumes was 520 fiber-days cc^{-1} and 1000 mg-days m^{-3} , respectively. A much smaller number of subjects were ever exposed to nickel and chromium (371 and 266, respectively). The mean cumulative exposures were 140 $\mu\text{g-days m}^{-3}$ and 45 $\mu\text{g-days m}^{-3}$ for nickel and chromium, respectively.

DISCUSSION AND CONCLUSIONS

Occupational epidemiologists have emphasized the importance of exposure assessment to the exposure-response characterization in epidemiology analyses (Stewart and Stewart, 1994; Nieuwenhuijsen, 2003; Nieuwenhuijsen *et al.*, 2006). The use of an exposure matrix has been considered the basis of retrospective exposure assessment methodology (Gomez *et al.*, 1994; Stewart, 1999; Ahrens and Stewart, 2003; Teschke, 2003). Site-specific exposure matri-

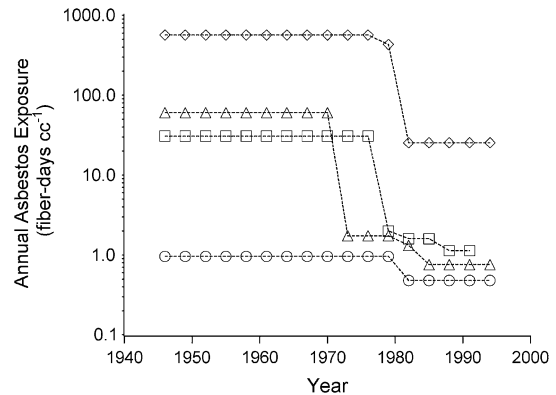


Fig. 3. Annualized asbestos exposure estimates (fiber-days cc^{-1}) for selected job titles at the Portsmouth Naval Shipyard (open triangle, welding; open square, electrician; open circle, sheet metal mechanic; open diamond, pipefitting).

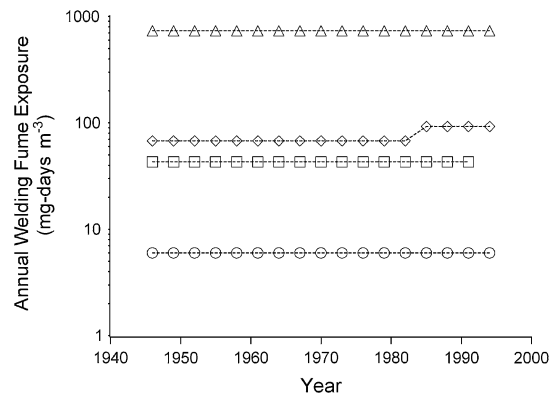


Fig. 4. Annualized welding fume exposure estimates (mg-days m^{-3}) for selected job titles at the Portsmouth Naval Shipyard (open triangle, welding; open square, electrician; open circle, sheet metal mechanic; open diamond, pipefitting).

ces, like the four in this study with shop and job title distinctions, increase precision of exposure estimates. A chronological dimension attempted to reduce misclassification of exposures over time.

Because most of the air monitoring data at PNS consisted of area samples from the 1980s forward, represented few shops and almost no job titles, it was impossible to directly assign quantitative exposure estimates to the many shop/job combinations in this study. Air monitoring data at PNS identified so few shop/job combinations that it also precluded validation of the final exposure estimates. Furthermore, even if it had been possible to link air data to shops and job titles, a direct comparison of the exposure estimates and the air sampling data would still have been questionable since the collection of air samples was complaint based until the 1970s when it became compliance driven. Air samples were never collected systematically to realistically portray

exposures at the shipyard. Misclassification would have been a real possibility if the air data were relied on for exposure estimates or even for validation of them.

Industrial hygienists are thought to have training and knowledge to estimate exposures (Stewart *et al.*, 2000; Teschke, 2003); however, for some exposure assessments, it may be very difficult for the industrial hygienists to become familiar with jobs and exposures, especially for work sites where they have spent little or no time. Providing industrial hygienists with detailed information about workplace conditions and exposure measurements overcomes this problem to some extent (Stewart *et al.*, 2000). The greatest strength of this exposure assessment is that it drew on the expertise of three very knowledgeable PNS industrial hygienists each with >25 years of experience with exposures and work practices at the shipyard from the 1970s through the 1990s. This is uncommon in most exposure assessments utilizing industrial hygienists. A growing interest in occupational exposure assessment methodology is inter-rater agreement and its evaluation (Goldberg *et al.*, 1986; Ciccone and Vineis, 1988; Macaluso *et al.*, 1993; Siemiatycki *et al.*, 1997; Rybicki *et al.*, 1998; Stewart *et al.*, 2000; Mannetje *et al.*, 2003; Correa *et al.*, 2006). Inter-rater agreement analyses using the kappa statistic and the ICC were carried out for this exposure assessment to better understand the panel's original estimates. Inter-rater agreement was low for a number of reasons. The kappa statistic, a function of both observed and expected agreement, can be influenced by unbalanced and/or asymmetric marginal distributions (Feinstein and Cicchetti, 1990). While concordance was high, overall simple kappas for frequency and intensity estimates were slight ($\kappa < 0.20$) and pairwise kappa results for frequency and intensity were wide ranging ($\kappa = 0.02$ – 0.51). When the distribution of ratings is unbalanced, as observed here with more shop/job/time period combinations assigned to the background and very low exposure intensity categories, the amount of agreement expected based on chance increases, which in turn, results in a lower kappa statistic. Consequently, the high amount of raw agreement observed here cannot be distinguished from chance agreement (Vach, 2005). Inter-rater agreement analyses of annualized exposure estimates were conducted using overall and pairwise ICC (range: 0.05–0.56). Although these statistical methods have known limitations, similar conclusions about the panel's original estimates were drawn from the different methods.

Other studies of inter-rater agreement have also had similar wide ranges of kappas and ICCs. Macaluso's study described earlier reported ICCs of <0 for rare exposures to >0.8 for frequent exposures (Macaluso *et al.*, 1993). A study of industrial hygiene

assessments of formaldehyde exposures also reported ICCs for inter-rater agreement ranging from <0 to >0.6 (Stewart *et al.*, 2000). A metal exposure assessment by industrial hygienists reported kappas from 0.15 to 0.49 (Rybicki *et al.*, 1998). In a study of chlorophenolate exposure, ICCs for industrial hygiene exposure estimates were 0.40–0.68 (Teschke *et al.*, 1989). Overall kappas and ICCs for the PNS exposure assessment tended to be lower than these other studies, but the pairwise statistics for panelists 2 and 3 were in the range of other studies' results. Pairwise statistical analyses were very informative and confirmed a difference between the estimates of panelist 1 and the other two panelists. Follow-up meetings confirmed that panelist 1 interpreted the instructions differently than the other two panelists. For example, panelist 1 assigned higher exposure estimates to all supervisor-level job titles that as non-supervisor job titles were highly exposed. As a result, estimates were inflated for job titles that were not necessarily highly exposed. In consensus meetings, panelist 1 stated that he did not want the high exposure missed for the non-supervisory job titles, failing to realize that he had also assigned high exposures to the same non-supervisor job titles.

The addition of consensus meetings was an important part of this exposure assessment. While the three industrial hygienists had direct experience at the shipyard in the 1970s through the 1990s, it became clear that they were not as familiar with exposures from earlier years and were not accustomed to estimating exposures at the level of detail needed for an epidemiology study. The consensus process helped them resolve their differences as they gained knowledge from each other and from support materials provided. Consensus is used in many fields including medicine, psychology and sociology to make diagnoses and estimate various parameters. Consensus is growing in popularity in occupational and environmental health as well. One example is a study of hazardous waste landfill sites and risk of congenital anomalies, in which consensus estimates of an expert panel were found to be more reliable than their independent estimates (Vrijheid *et al.*, 2002). Another study recommended that consensus or average estimates be used to control for differences among experts (de Cock *et al.*, 1996). A third study of retrospective exposure assessment also mentioned the value of a panel of experts over single raters (Benke *et al.*, 1997).

An important lesson learned during this study was that panel members need to be carefully instructed on methods for completing exposure matrices. It may not be a good assumption to think that industrial hygienists know how to estimate exposures in a matrix for an epidemiology study with little or no advance training. In addition, industrial hygiene panelists should perform sample exercises in order to prepare

for the estimation process and complete a small part (5–10%) of the actual exposure matrices to address misunderstandings of instructions. As a quality control measure it would have also been prudent to have had the panelists repeat a percentage of their estimates to add another level of assurance that their ratings were what they intended them to be (i.e. an assessment of intra-rater reliability). Another improvement to this exposure assessment would have been to provide panel members with the qualitative background materials about exposures at PNS especially in the early years before they worked there. The use of standardized job titles instead of the original titles may have made it harder for panel members to assign exposure estimates as they were not as familiar with the standardized job titles as they were with the raw job titles. It is also important to finalize employment histories prior to finalizing exposure matrices. For this exposure assessment, the addition of employment history following panel estimation required interpolation of some estimates. These estimates were interpolated from nearby years for the same shop/job combination when possible or from nearby years of the most appropriate shop/job combination based on professional judgment. Interpolated estimates did not diverge from panel estimates for similar years and similar jobs.

This exposure assessment had many strengths. Exposure and employment history data were compiled and study-specific exposure matrices were created. Experienced industrial hygienists estimated exposure intensity and frequency. Initial panel estimates had moderate to high concordance, and panel consensus meetings resolved all discordant estimates. The major weakness of the estimates is the inability to validate them. Air monitoring data did not allow for this, therefore exposure misclassification remains a possibility.

In nested case–control analyses that relied on this exposure assessment to estimate welding fumes and asbestos, lung cancer risk was higher with exposure to welding fumes [odds ratio (OR) = 1.03 at 1000 mg-days m^{-3} , 95% confidence interval (CI) = 1.00–1.05] and asbestos (OR = 1.04 at 1000 fiber-days cc^{-1} , 95% CI = 1.00–1.09) than no exposure (Yiin *et al.*, 2007). In categorical analyses, lung cancer risk increased with increasing welding fume exposure but leveled off at the highest category. For asbestos, exposed categories had higher risks than baseline, but risk estimates slightly decreased with increasing exposure categories. In the univariate analysis, lung cancer risk increased with increasing occupational radiation dose; however, in the multivariate analysis, lung cancer risk decreased with increasing occupational radiation dose when welding fume and asbestos exposures were included in the model. Without the asbestos and welding fume exposure estimates from this study, lung cancer risk due to

radiation exposure would have been over estimated. This is consistent with the findings of the previous lung cancer case–control study at the shipyard (Rinsky *et al.*, 1988). The addition of the exposure estimates for asbestos and welding fumes improved the epidemiology capability to understand the risk of exposure to ionizing radiation and lung cancer at PNS.

FUNDING

Funding for this study was provided through an agreement between the US Department of Energy and the US Department of Health and Human Services.

Acknowledgements—The authors acknowledge Jeffrey Brann, James Braun, Mark Duval and Christopher Moore from the PNS; Doug Daniels, Don Fleming, Lynette Hartle, Greg Kinnes, Sharon Silver and James Yiin from NIOSH and Bin Chen, James Couch, Karin Davis, Mary Engle, Sarah Locke, Tim McAdams and Susan Viet of Westat, Inc.

Disclaimer—The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

REFERENCES

- ACGIH. (2005) Threshold limit values for chemical substances and physical agents and biologic exposure indices. Cincinnati, OH: American Conference of Governmental Industrial Hygienists.
- Ahrens W, Stewart P. (2003) Retrospective exposure assessment. In: Nieuwenhuijsen M, editor. Exposure assessment in occupational and environmental epidemiology. Oxford, UK: Oxford University Press; pp. 103–18.
- Antonini JM. (2003) Health effects of welding. *Crit Rev Toxicol*; 33: 61–103.
- Benke G, Sim M, Forbes A *et al.* (1997) Retrospective assessment of occupational exposure to chemicals in community-based studies: validity and repeatability of industrial hygiene panel ratings. *Int J Epidemiol*; 26: 635–42.
- Ciccone G, Vineis P. (1988) Inter-rater agreement in the assessment of occupational exposure to herbicides. *Med Lav*; 79: 363–7.
- Correa A, Min YI, Stewart PA *et al.* (2006) Inter-rater agreement of assessed prenatal maternal occupational exposures to lead. *Birth Defects Res A Clin Mol Teratol*; 76: 811–24.
- de Cock J, Kromhout H, Heederik D *et al.* (1996) Experts' subjective assessment of pesticide exposure in fruit growing. *Scand J Work Environ Health*; 22: 425–32.
- Doll R, Anderson A, Cooper W. (1990) Report of the international committee on nickel carcinogenesis in man. *Scand J Work Environ Health*; 16 (Suppl. 1): 1–82.
- Feinstein A, Cicchetti D. (1990) High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*; 43: 543–9.
- Fleiss J, Cohen J. (1973) The equivalence of weighted kappa and the intra-class correlation coefficient as measures of reliability. *Educ Psychol Meas*; 33: 613–9.
- Goldberg M, Siemiatycki J, Gerin M. (1986) Inter-rater agreement in assessing occupational exposure in a case-control study. *Br J Ind Med*; 43: 667–76.
- Gomez M, Cocco P, Dosemeci M *et al.* (1994) Occupational exposure to chlorinated aliphatic hydrocarbons: job exposure matrix. *Am J Ind Med*; 26: 171–83.

- Landis J, Koch G. (1977) The measurement of observer agreement for categorical data. *Biometrics*; 33: 159–74.
- Langard S. (1990) One hundred years of chromium and cancer: a review of epidemiological evidence and selected case reports. *Am J Ind Med*; 17: 189–215.
- Macaluso M, Delzell E, Rose V *et al.* (1993) Inter-rater agreement in the assessment of solvent exposure at a car assembly plant. *Am Ind Hyg Assoc J*; 54: 351–9.
- Mangold C, Clark K, Madl A *et al.* (2006) An exposure study of bystanders and workers during the installation and removal of asbestos gaskets and packing. *J Occup Environ Hyg*; 3: 87–98.
- Mannetje A, Fevotte J, Fletcher T *et al.* (2003) Assessing exposure misclassification by expert assessment in multi-center occupational studies. *Epidemiology*; 14: 585–92.
- Najarian T, Colton T. (1978) Mortality from leukaemia and cancer in shipyard nuclear workers. *Lancet*; 1: 1018–20.
- Nieuwenhuijsen M. (2003) Introduction to exposure assessment. In: Nieuwenhuijsen M, editor. *Exposure assessment in occupational and environmental epidemiology*. Oxford, UK: Oxford University Press; pp. 3–19.
- Nieuwenhuijsen M, Paustenbach D, Duarte-Davidson R. (2006) New developments in exposure assessment: the impact on the practice of health risk assessment and epidemiological studies. *Environ Int*; 32: 996–1009.
- NIOSH. (2005) NIOSH pocket guide to chemical hazards. Cincinnati, OH: National Institute for Occupational Safety and Health.
- OSHA. (1998) Preamble to the final rule of the Osha respiratory protection standard. Washington, DC: Occupational Safety and Health Administration.
- Rinsky RA, Melius JM, Hornung RW *et al.* (1988) Case-control study of lung cancer in civilian employees at the Portsmouth Naval Shipyard, Kittery, Maine. *Am J Epidemiol*; 127: 55–64.
- Rinsky RA, Waxweiler RJ, Bierbaum PJ *et al.* (1980) Epidemiologic study of civilian employees at the Portsmouth Naval Shipyard. Cincinnati, OH: National Institute for Occupational Safety and Health.
- Rinsky RA, Zumwalde RD, Waxweiler RJ *et al.* (1981) Cancer mortality at a naval nuclear shipyard. *Lancet*; 1: 231–5.
- Rybicki B, Peterson E, Johnson C *et al.* (1998) Intra- and inter-rater agreement in the assessment of occupational exposure to metals. *Int J Epidemiol*; 27: 269–73.
- Selkoff I, Lee D. (1978) *Asbestos and disease*. New York: Academic Press.
- Shrout P, Fleiss J. (1979) Intra-class correlations: uses in assessing rater reliability. *Psychol Bull*; 86: 420–8.
- Siemiatycki J, Fritschi L, Nadon L *et al.* (1997) Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community-based case-control studies. *Am J Ind Med*; 31: 280–6.
- Silver SR, Daniels RD, Taulbee TD *et al.* (2004) Differences in mortality by radiation monitoring status in an expanded cohort of Portsmouth Naval Shipyard workers. *J Occup Environ Med*; 46: 677–90.
- Stewart P. (1999) Challenges to retrospective exposure assessment. *Scand J Work Environ Health*; 25: 505–10.
- Stewart P, Carel R, Schairer C *et al.* (2000) Comparison of industrial hygienists' exposure evaluations for an epidemiologic study. *Scand J Work Environ Health*; 26: 44–51.
- Stewart P, Stewart W. (1994) Occupational case-control studies: II. Recommendations for exposure assessment. *Am J Ind Med*; 26: 313–26.
- Teschke K. (2003) Exposure surrogates: job exposure matrices, self reports and expert evaluations. In: Nieuwenhuijsen M, editor. *Exposure assessment in occupational and environmental epidemiology*. Oxford, UK: Oxford University Press; pp. 119–32.
- Teschke K, Hertzman C, Dimich-Ward H *et al.* (1989) A comparison of exposure estimates by worker raters and industrial hygienists. *Scand J Work Environ Health*; 15: 424–9.
- Vach W. (2005) The dependence of Cohen's kappa on the prevalence does not matter. *J Clin Epidemiol*; 58: 655–61.
- Vrijheid M, Dolk H, Armstrong B *et al.* (2002) Hazard potential ranking of hazardous waste landfill sites and risk of congenital anomalies. *Occup Environ Med*; 59: 768–76.
- Yiin JH, Schubauer-Berigan MK, Silver SR *et al.* (2005) Risk of lung cancer and leukemia from exposure to ionizing radiation and potential confounders among workers at the Portsmouth Naval Shipyard. *Radiat Res*; 163: 603–13.
- Yiin JH, Silver SR, Daniels RD *et al.* (2007) A nested case-control study of lung cancer risk and ionizing radiation exposure at the Portsmouth Naval Shipyard. *Radiat Res*; 168: 341–48.