Inter-Reader Agreement in ILO Readings of Radiographs for Pneumoconiosis

Marek Mikulski[1], Patrick Hartley[2], Nancy Sprince[1], George Delclos[3], Spencer Lourens[4], Laurence Fuortes[1]
[1]Department of Occupational and Environmental Health, College of Public Health, University of Iowa, Iowa City, IA
[2]Carver College of Medicine, University of Iowa, Iowa City, IA
[3]Division of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Sciences Center at Houston School of Public Health, Houston, TX
[4]Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA

Background

The International Labor Organization's International System (ILO system) for Classification of Radiographs for Pneumoconiosis has been accepted by the scientific community for use in studies of pneumoconiosis. The National Institute for Occupational Safety and Health (NIOSH) recommends using multiple ILO trained readers to increase accuracy and precision of readings. The significance of between-reader agreement on the estimates of prevalence of ILO abnormalities in medical screening programs in general and screenings of former nuclear weapons workers in particular has not been studied extensively. We screened over 2,650 workers from two sites in Iowa with a chest x-ray, spirometry and sensitization to beryllium as part of the U.S. Department of Energy (DoE) funded Former Nuclear Weapons Worker Medical Screening Program (DoE FWP). The films were reviewed according to ILO (rev. 2000) guidelines by three occupational medicine physicians (clinician readers) and the prevalence of parenchymal abnormalities (median profusion score ≥1/0) from one site was found to be up-to twice as high as in other DoE populations with higher potential for exposure to parenchymal lung disease hazards reported to date (Dement et al., 2003, Makie et al., 2005). Other than suggestion of asbestos for isolated pleural fibrosis no occupational etiology was identified (Mikulski et al., 2011) but age was statistically significantly associated with prevalence of isolated parenchymal and pleural abnormalities. The agreement between the readers ranged from moderate to substantial for dichotomous parenchymal and pleural abnormalities and was also substantial for ordinal profusion scoring. This pattern of ILO readings could reflect an effect of age, or a true high rate of pneumoconiosis but increased sensitivity by non-B clinician readers cannot be ruled out. A sample of 500 films was reviewed by independent ILO B-reader for quality assurance purposes and results are reported.

DoE Former Worker Medical Screening Programs

The University of Iowa College of Public Health medical screening programs for former nuclear weapons workers from the state of Iowa started in 2001 and 2006.

Part of the nationwide screening program funded by DoE under Public Law 102-484 Section 3162 of the 1993 Defense Authorization Act

Goal: *Identifying, locating, and providing former Iowa Army Ammunition Plant (IAAAP) and Ames Lab DoE workers employed in research and manufacture of nuclear weapons with medical evaluation of long term health effects that might have resulted from employment*

DoE Former Worker Medical Screening Programs

Site 1 – Iowa Army Ammunition Plant Located in Middletown, Iowa - primarily manufacture of conventional munitions for Department of Defense (DoD) but nuclear weapons assembled, disassembled and repaired between 1949 and mid-1975 on Line 1 under Atomic Energy Commission (AEC, pre-DoE) contractual agreements with Silas-Mason Company.
Workforce approximately 7,000 workers

Exposures (expanded list available at http://www.sem.dol.gov/index.cfm)

- *Ionizing radiation*
- *Beryllium*
- *Asbestos*
- *Solvents*
- *High Explosives incl. Barium*
- *Isocyanates*
- *Epoxy adhesives*
- *Curing agents*

Site 2 – Ames Lab located on the Iowa State University campus in Ames, Iowa – operational since 1942, primarily involved in research and development for nuclear weapons industry but in early years commercial purification of significant amounts of uranium and thorium for use in experiments and weaponry. Workforce approximately 12,000 workers

Exposures (expanded list available at http://www.sem.dol.gov/index.cfm)

- *Ionizing radiation (uranium, plutonium, thorium)*
- *Beryllium*
- *Asbestos*
- *Silica*
- *Solvents*

Methods

Participants for the FWP screenings were recruited by mail, telephone, press releases, town-hall meetings and word of mouth. All participating in the medical screenings were offered CXR, spirometry and testing for Beryllium sensitization (BeS).

Postero-Anterior (PA) films were reviewed by three occupational medicine clinicians using the ILO system, blinded to radiologist's reports and each other's readings but not to the industry and it's exposures. The most recent CXR with three ILO readings was used for analysis. Abnormal profusion scoring defined as ≥1/0

A sample of 500 films (read by three readers), including all those read as consistent with work-related parenchymal and/or pleural disease by ILO standards were reviewed by an external ILO B-reader blinded to exposure history, personal characteristics, radiologist's reports and other readings.

Spirometry was performed according to American Thoracic Society (ATS) guidelines using Third National Health and Nutrition Examination Survey (NHANES III) reference population and American College of Occupational and Environmental Medicine (ACOEM) recommended lower limit of normal (LLN) algorithm for interpretation.

Beryllium Sensitization (BeS) was defined as two abnormal blood Beryllium Lymphocyte Proliferation Tests (BeLPT) or one abnormal + one borderline BeLPT result from any accredited laboratory. Non-normal initial BeLPT were repeated with a split test sent to two laboratories within 6-12 months of the initial testing. Normal results offered repeat testing within three to five years.

Highest ever exposure potential to beryllium and asbestos was estimated by industrial hygienists based on job codes/job titles in subcontractor's and plant's employment records and workers' interviews. Jobs with *highest exposure potential to beryllium* (frequent, direct exposure): millwright, tool and die maker, machinist. Jobs with *occasional exposure potential to beryllium:* production operator, engineer/scientist, pipefitter, plumber, electrician, laundry operator. Jobs with *highest exposure potential to asbestos* (frequent, direct exposure): pipefitter, plumber, carpenter. Jobs with *occasional exposure potential to asbestos:* power plant operators, auto/equipment mechanics

Analysis

Multivariate logistic regression was used to validate ILO readings in models with known predictors of parenchymal and/or pleural disease. Modeling was done separately for each reader and with ILO abnormalities as a dependent variable 1) grouped together i.e. ever-abnormal (Y/N); and 2) modeled separately i.e. isolated parenchymal abnormalities (PA) in a separate model from pleural (PL) and coincident parenchymal pleural abnormalities (PA+PL). Model-fit assessed by AIC

Never smokers were defined as those with less than 20-pack smoking history during lifetime or less than one cigarette smoked per day for one year. Ex-smokers were defined as those who quit smoking >1 month before the survey. Pack-years were calculated according to pack/day*years smoked formula; cigarette conversion for other types of tobacco use - 1 small cigar = 3 cigarettes; 1 regular cigar = 5 cigarettes; 1 pipe = 4 cigarettes

Inter-reader concordance was assessed by calculating simple and weighted kappa-statistic according to Fleiss (1971) and Cohen (1960), separately for each ILO clinician reader compared to the B-reader. Generalized kappa (generalized Pi, Scott 1955) was calculated for all four readers together. Concordance was assessed separately for 1) ever-abnormal film incl. PA and PA+PL and PL (Y/N); 2) ILO type of abnormality i.e. PA vs. PA+PL vs. PL vs. NL; and 3) for ILO profusion scoring grouped according to (modified) Miller et al. 1996, Group 0: 0/-, 0/0; Group 1: 0/1; Group 2: 1/0, Group 3: 1/1; Group 4: 1/2-3/+

Results

Table 1.Characteristics of individuals included in the study by gender

Parameter	Male n= 347 (69.0%)	Female n= 153 (31.0%)	Total n=500
Age, mean (SD), range	69(12); 31-99	62(14);32-91	67(13);31-99
Age, n (%)			
<40	9 (2.6)	13 (8.5)	22 (4.4)
41-50	17 (4.9)	20 (13.1)	37 (7.4)
51-60	46 (13.2)	43 (28.1)	89 (17.8)
61-70	91 (26.2)	22 (14.4)	113 (22.6)
71-80	129 (37.2)	36 (23.5)	165 (33.0)
>80	55 (15.9)	19 (12.4)	74 (14.8)
Race, n (%)			
White	338 (97.4)	145 (94.9)	483 (96.6)
African-American	3 (0.9)	6 (3.9)	9 (1.8)
Asian	2 (0.6)	1 (0.6)	3 (0.6)
Hispanic	4 (1.1)	1 (0.6)	5 (1.0)
Smoking, n (%)			
Never smoker	120 (34.6)	99 (64.7)	219 (43.8)
Ex-smoker	197 (56.8)	42 (27.4)	239 (47.8)
Smoker	30 (8.6)	12 (7.9)	42 (8.4)
Pack-Years, mean (SD), range	39(33); 0.1-180	23(19); 0.3-75	36(31);0.1-180
Pack years, n (%)			
Low <10	38 (16.7)	16 (29.6)	54 (19.2)
Medium 10-20	32 (14.1)	8 (14.8)	40 (14.2)
High >20	138 (60.8)	26 (48.1)	164 (58.4)
Missing (Ever Smokers)	19 (8.4)	4 (7.5)	23 (8.2)
BMI mean (SD), range	28(5); 18-55	28(6); 18-58	28(5);18-58
BMI, n (%)			
<25	81 (23.3)	51 (33.3)	132 (26.4)
25-29	147 (42.4)	50 (32.7)	197 (39.4)
≥30	119 (34.3)	52 (34.0)	171 (34.2)
Spirometry, n (%)			
Normal	179 (51.6)	98 (64.0)	277 (55.4)
Normal/Borderline Obstructive	11 (3.2)	9 (5.9)	20 (4.0)
Obstructive	12 (3.4)	9 (5.9)	21 (4.2)
Restrictive	104 (30.0)	32 (20.9)	136 (27.2)
Mixed	32 (9.2)	3 (2.0)	35 (7.0)
Missing	9 (2.6)	2 (1.3)	11 (2.2)
Beryllium Sensitized, n (%)			
Yes	7 (2.0)	1 (0.7)	8 (1.6)
No	334 (96.3)	150 (98.0)	484 (96.8)
Missing	6 (1.7)	2 (1.3)	8 (1.6)
Beryllium exposure, n (%)			
Low, background	117 (33.7)	33 (21.6)	150 (30.0)
Occasional	87 (25.1)	39 (25.5)	126 (25.2)
Frequent, direct	21 (6.1)	-	21 (4.2)
Missing	122 (35.1)	81 (52.9)	203 (40.6)
Asbestos exposure, n (%)			
Low, background	181 (52.2)	72 (47.1)	253 (50.6)
Occasional	20 (5.8)	-	20 (4.0)
Frequent, direct	24 (6.9)	-	24 (4.8)
Missing	122 (35.1)	81 (52.9)	203 (40.6)

Table 2. Distribution of ILO abnormalities by reader and odds of detecting abnormality by readers 1-3 compared to B-reader

ILO abnormalities	ILO Reader 1 (Non-B reader) N (%); OR (95%CI)	ILO Reader 2 (Non-B reader) N (%); OR (95%CI)	ILO Reader 3 (Non-B reader) N (%); OR (95%CI)	ILO Reader 4 (B-reader) N (%); OR (95%CI)
	N=500 CXR films			
Pleural	46 (9.2); 1.53 (0.95-2.45)	50 (10.0); 1.68 (1.05-2.68)	41 (8.2); 1.35 (0.83-2.19)	31 (6.2); 1.0
Parenchymal Only (≥1/0)	62 (12.4); **2.30 (1.45-3.64)**	64 (12.8); **2.38 (1.51-3.77)**	47 (9.4); **1.68 (1.04-2.72)**	29 (5.8); 1.0
Parenchymal/Pleural (≥1/0)	26 (5.2); **3.73 (1.51-7.52)**	19 (3.8); **2.43 (1.05-5.60)**	10 (2.0); 1.26 (0.49-3.69)	8 (1.6); 1.0
Total Abnormal Profusion (≥1/0 to 3/+)	88 (17.6); **2.67 (1.78-4.01)**	83 (16.6); **2.49 (1.65-3.75)**	57 (11.6); **1.61 (1.04-2.84)**	37 (7.4); 1.0
Profusion				
0/1	37 (7.4); N/A	23 (4.6); N/A	26 (5.2); N/A	30 (6.0); N/A
1/0;	56 (11.2); N/A	33 (6.6); N/A	27 (5.4); N/A	17 (3.4); N/A
1/1;	14 (2.8); N/A	22 (4.4); N/A	23 (4.6); N/A	13 (2.6); N/A
1/2 - 3/+	18 (3.6); N/A	28 (5.6); N/A	7 (1.4); N/A	7 (1.4); N/A
Other Findings:				
Calcified Granulomas (cg)	203 (40.6); **5.65 (4.04-7.88)**	339 (67.8);**17.39 (12.39-24.41)**	279 (55.8); **10.43 (7.47-14.55)**	54 (10.8); 1.0
Emphysema (em)	50 (10.0); **11.0 (4.35-27.83)**	21 (4.2); **4.34 (1.62-11.60)**	15 (3.0); **3.06 (1.10-8.40)**	5 (1.0); 1.0
Effusion (ef)	5 (1.0); **0.19 (0.07-0.51)**	2 (0.4); **0.08 (0.02-0.32)**	2 (0.4); **0.08 (0.02-0.32)**	2 (0.4); 1.0
Honeycomb lung (ho)	1 (0.2); **0.50 (0.05-5.52)**	1 (0.2); **0.50 (0.05-5.52)**	-	2 (0.4); 1.0
Not abnormal	366 (73.2); N/A	367 (73.4); N/A	402 (80.4); N/A	432 (86.4); N/A

Table 3. Results of inter-reader concordance analysis of interpretation of CXR according to ILO guidelines

ILO abnormalities	ILO Reader 1 vs. ILO-B Reader 4 *Kappa Statistic(95%CI)*	ILO Reader 2 vs. ILO-B Reader 4 *Kappa Statistic (95%CI)*	ILO Reader 3 vs. ILO B-Reader 4 *Kappa Statistic (95%CI)*	All Readers *Kappa Statistic (95%CI)*
Ever-abnormal (Y/N)	**0.35 (0.24-0.46)**	**0.31 (0.20-0.43)**	**0.32 (0.19-0.45)**	**0.43 (0.34-0.53)**
Type of abnormality				
Parenchymal only	0.26 (-0.01-0.54)	0.25 (-0.02-0.53)	**0.29 (0.01-0.56)**	**0.38 (0.30-0.46)**
Parenchymal and Pleural	0.33 (-0.04-0.70)	**0.42 (0.02-0.84)**	**0.55 (0.06-1.0)**	**0.48 (0.38-0.58)**
Pleural only	**0.31 (0.04-0.59)**	**0.40 (0.11-0.70)**	**0.40 (0.11-0.70)**	**0.48 (0.40-0.56)**
Profusion Score groups				
Group 0 (0/-,0/0)	0.38 (-0.03-0.79)	0.41 (-0.04-0.86)	0.37 (-0.10-0.84)	**0.49 (0.28-0.71)**
Group 1 (0/1)	0.02 (-0.27-0.32)	0.22 (-0.09-0.54)	0.09 (-0.22-0.41)	**0.13 (0.04-0.23)**
Group 2 (1/0)	0.10 (-0.20-0.39)	-0.01 (-0.34-0.32)	0.002 (-0.33-0.34)	**0.13 (0.04-0.23)**
Group 3 (1/1)	0.20 (-0.21-0.61)	0.14 (-0.23-0.51)	0.19 (-0.18-0.57)	**0.23 (0.13-0.33)**
Group 4 (1/2-3/+)	**0.47 (0.04-0.90)**	0.26 (-0.11-0.63)	0.42 (-0.13-0.97)	**0.39 (0.30-0.49)**
Other abnormalities				
Calcified granulomas (cg)	**0.18 (0.11-0.25)**	**0.07 (0.03-0.10)**	**0.14 (0.09-0.18)**	**0.19 (0.15-0.23)**
Emphysema (em)	**0.13 (0.01-0.25)**	0.14 (-0.05-0.33)	0.09 (-0.09-0.27)	0.24 (-0.23-0.71)

Kappa statistic interpretation (inter-reader agreement: (Landis and Koch, 1977)

<0 poor;
0-0.20 slight;
0.21-0.40 fair;
0.41-0.60 moderate
0.61-0.80 substantial;
0.81-1.0 almost perfect

Results

Table 4. Logistic regression models for known parenchymal and pleural risk factors/predictors to validate ILO readings by reader

Predictor	ILO Reader 1 OR (95% CI)	ILO Reader 2 OR (95% CI)	ILO Reader 3 OR (95% CI)	ILO Reader 4 OR (95% CI)
Spirometry				
Normal	1.0	1.0	1.0	1.0
Obstructive	3.10 (0.88-10.93)	3.06 (0.93-10.13)	1.32 ((0.33-5.34)	1.91 (0.45-8.09)
Restrictive	**3.66 (1.98-6.76)**	1.78 (1.00-3.16)	**2.82 (1.50-5.28)**	1.39 (0.63-3.07)
Mixed	**2.36 (1.01-5.55)**	1.57 (0.69-3.59)	1.69 (0.69-4.13)	**4.19 (1.69-10.39)**
Smoking	1.15 (0.73-1.82)	1.23 (0.80-1.90)	1.17 (0.73-1.90)	1.11 (0.63-1.96)
Asbestos exposure				
Rare	1.0	1.0	1.0	1.0
Occasional	0.91 (0.33-2.54)	0.66 (0.23-1.86)	1.42 (0.51-3.94)	0.68 (0.17-2.71)
Frequent	1.85 (0.71-4.84)	1.62 (0.65-4.02)	2.48 (0.97-6.35)	1.90 (0.65-5.56)
Age	**1.08 (1.04-1.11)**	**1.04 (1.01-1.07)**	1.03 (0.99-1.07)	1.04 (1.00-1.09)
BMI	0.97 (0.92-1.03)	1.01 (0.96-1.06)	0.97 (0.92-.1.02)	0.94 (0.87-1.01)

The algorithm for ILO abnormalities analyzed separately i.e. parenchymal separately from parenchymal+pleural and pleural did not converge and no models were generated

Summary of Findings/Conclusions

Clinician readers with knowledge of specific industry and site-specific exposures were up-to over three times more likely to interpret films as consistent with ILO abnormalities when compared to ILO B-reader who was blinded to site-specific and exposure information. This increased sensitivity resulted in higher individual estimates of prevalence of ILO abnormalities (all types) based on clinician readings compared to ILO B-reader.

The agreement between these clinician readers for the cohort of 757 workers from one site was previously assessed to range from moderate to substantial for ever-abnormal films and for ordinal group profusion scoring (Mikulski et al., 2011). The agreement between each of the clinician readers and the ILO B-reader individually was lower than in the previously published study and ranged only from slight to fair in the same abnormality categories. The overall agreement between all four readers in this study confirmed these findings.

The higher than expected agreement between clinician readers resulted in estimates of prevalence of isolated ILO parenchymal and coincident parenchymal and pleural abnormalities up-to twice as high as those of the ILO B – reader when NIOSH recommended median profusion score was used to reconcile multiple readings (9.6% vs. 5.8% and 3.4% vs. 1.6%). Estimates of prevalence of pleural abnormalities based on two positive out of a total of three readings were also higher for clinician readers compared to ILO B-reader (8.2% vs. 6.2%)

Clinician readers were significantly more likely to report chronic granulomatous disease and emphysema and less likely to report effusion than the B-reader.

Clinician and ILO B-readings showed a similar pattern of increased likelihood of abnormal spirometry results associated with abnormal ILO reading. There was a suggestion of some degree of specificity for the association between abnormal ILO readings consistent with pneumoconiosis and restrictive and or mixed spirometric deficits. Clinician and ILO B-readings were also similar as regards associations between increased age and likelihood of ILO abnormality consistent with pneumoconiosis, with odds ratios ranging from 1.03 to 1.08 for increase in abnormal ILO reading associated with each year increase in age.

Concordance between clinician non-B readers and the ILO B-reader was highest for films with the highest and lowest profusion scores

The increase in sensitivity of clinician readers has been reported before (Gitlin et al., 2004) and cannot be fully explained by this study. The associations between abnormal spirometry and abnormal film characterization suggest some degree of validation, however without a true Gold Standard sensitivity and specificity cannot been quantified. Use of standardized, quantifiable methods of interpretation of interstitial opacities and patterns using CAT scan technologies may assist in addressing such questions.