

## AN ANALYSIS OF X-RAY READER AGREEMENT: DO FIVE READERS SIGNIFICANTLY INCREASE READER CLASSIFICATION RELIABILITY OVER THAT OF THREE READERS?

JOHN LEFANTE, Ph.D. • Janet Hughes, Ph.D.  
• Robert Jones, M.D. • Hans Weill, M.D.

Tulane University School of Medicine, Department of Medicine  
Pulmonary Diseases Section, New Orleans, LA, USA

### ABSTRACT

Five experienced readers, working independently and without exposure or other subject data, applied the ILO 1980 classification to the chest radiographs of 1,168 workers currently employed in the manufacture of man-made mineral fibers. To examine the effect of using five instead of three readers, we determined the difference in profusion of small opacities when using the median readings of all five readers, compared to the medians produced by each of the ten panels combining three of the readers. The distribution of small opacity profusion differed among individual readers, but most of the films were placed in low categories: fewer than one percent were above category 1/1. Readers also differed in their assessments of film quality, and showed differing sensitivities to radiographic effects of age and smoking. Despite these individual differences, the addition of two readers usually had little effect on median profusion judgments. For eight of the ten panels, five or fewer percent of films were classified differently; for all ten panels, fewer than two percent of films were reclassified beyond an adjacent category on the twelve-part scale. Eight panels were sensitive to radiographic effects of smoking, and three to effects of age; the five-member panel was also sensitive to smoking. We conclude that, when experienced readers are used, enlarging a panel from three to five members is unlikely to affect median small opacity profusion.

### INTRODUCTION

The ILO 1980 Classification of Radiographic Appearances of Pneumoconioses is subject to inter- and intra-observer variability.<sup>1</sup> It has therefore been accepted practice to use several experienced readers—at least three—in order to minimize the chance of systematic bias.<sup>3</sup> The median reading for the readers has been used to summarize each film, since it is not affected by the values of the extreme readings, as would be the mean reading.<sup>2</sup> Although it is intuitively true that the accuracy of the summarized readings should increase with more readers, there is a question whether a significant proportion of readings would change if the number of readers was increased beyond three. In a study of a working population engaged in man-made mineral fiber manufacturing, ILO 1980 classifications of a set of chest X-rays by five readers were available for comparison with all possible subsets of three readers.

### METHODS

The study materials consisted of the chest radiographs of 1,755 employees in seven plants. All five readers attempted to classify all films according to the ILO 1980 Classification. The readers concluded that a different percentage of these 1,755 films were of unreadable quality (0% to 8%). A total of 1,603 films were judged readable, and were

classified by all five readers. At the time of this analysis complete smoking history was available for 1,168 subjects, and we will only consider this subset to assess the effect of age and smoking status on the classification of low-level opacities. The ages of the workers considered have a mean of 41 years, a standard deviation of 11 years, and a range from 19 to 76 years. Sixty-eight percent of the workers were either current or ex-smokers with pack years having a mean of 28, a standard deviation of 25, and a range from 0.2 to 168.

### RESULTS

Most individual readings were concentrated in the lower profusion categories, as indicated in Table I. All readers classified between 82% and 98% of the films as 0/0, with only between 1% and 10% over 0/1. Two readers classified 0.5% of the films  $\frac{1}{2}$  or higher.

Table II considers the distribution of individual readings relative to the entire group. The entries in the table represent the percentage of times each individual reading deviates from the median reading for all five readers. For example, Reader A classified 3.5% of the 1,168 films one category higher than the classifications based on the median of all five readers. If we consider the number of times individuals rated

Table I  
Distribution of Small Opacities Profusion in %, by Reader (N = 1,168)

Reader	Profusion				
	0/0	0/1	1/0	1/1	$\geq 1/2$
A	90.6	3.7	1.2	4.0	0.5
B	82.1	7.5	3.6	6.3	0.5
C	89.4	8.5	1.3	0.8	0.0
D	92.1	6.0	1.5	0.4	0.0
E	98.0	1.2	0.7	0.1	0.0
Median	94.8	3.6	1.1	0.5	0.0

Table II  
Distributions of Individuals' Readings, by Distance from Median Reading (N = 1,168)

Reader	Number of Sub-Categories From the Median						$\geq +3$
	-3	-2	-1	0	+1	+2	
A	0.0	0.0	1.0	91.4	3.5	2.1	2.0
B	0.0	0.1	0.3	83.6	7.3	4.3	4.4
C	0.1	0.2	0.9	92.0	6.2	0.3	0.3
D	0.0	0.4	1.8	93.0	4.1	0.7	0.0
E	0.0	0.6	3.4	95.9	0.1	0.0	0.0

films more than one category above or below the median, we see that Readers A and B tended to read higher than the others. Readers C and D read on both sides of the median, with slightly greater readings above. Reader E read consistently lower than the others.

In order to assess the effect of adding two readers to an original three, all ten possible panels of three readers were considered. For each of these panels of three, the median reading for each film was recorded. Table III demonstrates, for each combination, the percentage of films that changed their classification when the medians were based on all five readers. Since a change of more than one category is considered important, the combinations of readers in Table III

are ranked in descending order according to the percentage of films that changed more than one subcategory when two readers were added. In the worst case, adding two readers made any change in only 7.4% of the films (when A, B, and C are the original three), and a change greater than one category in only 1.6%. In the majority of combinations (70%), fewer than 0.5% changed by more than one category. The combinations that are affected the most by adding two readers are the ones that include Readers A and B, the two highest readers based on Table II. The addition of two more moderate readers will affect a greater proportion of film classifications. Nevertheless, even in these cases only 1.5% of the classifications changed by two or more categories, indicating that the addition of two readers had little effect.

The possible effects of age and smoking status on the classification of low-level opacities was investigated for each reader. Exposure information on the workers was not available at the time of analysis and was not included. Obviously, this very important explanatory variable should be considered when possible, since any detectable effects due to age and smoking alone could be confounded with an exposure effect. The limited attempt here is to see how often detection of the radiographic effects of age or smoking are changed by changing the size or composition of panels. Table IV presents the result of logistic regressions on age, smoking (ever versus never), pack years, and finally pack years after adjusting for age, for each reader. Readers A and E, high and low readers respectively, detected no effects (with the exception of an age effect for Reader E). Readers B, C, and D detected almost every effect. Therefore, no discernible relationship to over- or under-reading exists. Table V

presents the result of logistic regressions for each three-reader panel and for the five-reader panel. Eight panels were sensitive to the radiographic effects of smoking, as shown by a significant relationship between small opacity profusion and either smoking category or pack-years. Three panels were similarly sensitive to the effects of age. The five-member panel was also sensitive to smoking, indicating that the addition of two readers is unlikely to eliminate this possible source of confounding with dust effects.

### CONCLUSIONS

When using three experienced readers (the widely accepted minimum for epidemiologic research), the effect of including two additional readers seems to be negligible. Despite using readers with significant inter-observer variability, only a small proportion of films were changed by more than one category, after adding two readers.

Table III  
Rate of Change in Median Profusion Category After Adding Two Readers,  
for Each Possible Combination of Three Readers

Readers	Percent Films Showing Any Change	Percent Films Changing More Than One Sub-Category
A B C	7.4	1.6
A B D	5.7	1.5
A B E	5.0	1.4
B D E	3.2	0.4
B C D	4.6	0.3
C D E	3.3	0.3
A C D	2.4	0.3
A D E	2.7	0.3
B C E	3.4	0.3
A C E	2.2	0.2

Table IV  
Sensitivities of Individual Readers to Age and Smoking Effects on  
Small Opacity Profusion (Logistic Regression, N = 1,168)

Reader	% > 0/0	Age	Smoking	Pack-Years	Pack-Years/Age
A	9.4	NS	NS	NS	NS
B	17.9	**	**	**	*
C	10.6	NS	*	**	**
D	7.9	**	**	**	**
E	2.0	*	NS	NS	NS

\*significant  $p \leq 0.05$

\*\*significant  $p \leq 0.01$

NS = not significant

Table V  
Sensitivities of All Three-Reader Panels to Age and Smoking Effects on Small Opacity Profusion

Readers	Age	Smoking	Pack-Years	Pack-Years/Age
A B C	NS	NS	*	NS
A B D	*	NS	*	NS
A B E	*	NS	NS	NS
B D E	*	*	**	*
B C D	NS	*	**	**
C D E	NS	**	**	**
A C D	NS	*	**	*
A D E	NS	*	NS	NS
B C E	NS	*	*	*
A C E	NS	NS	NS	NS
ABCDE	NS	*	NS	NS

\* $p \leq 0.05$

\*\* $p \leq 0.01$

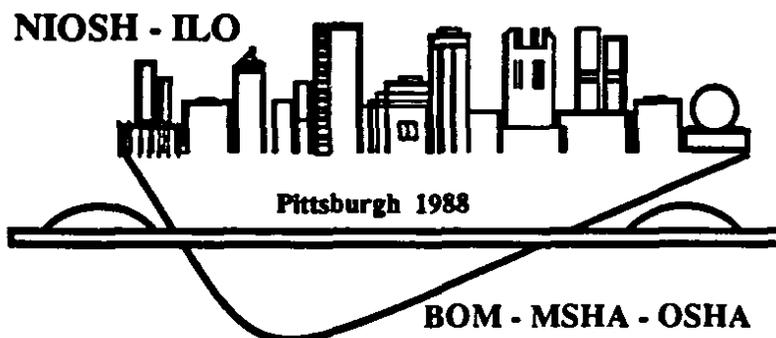
NS = not significant

**REFERENCES**

1. Gilson, J.C., Jones, R.N.: Radiography. In: Weill, H., Turner-Warwick, M. (eds): *Occupational Lung Diseases: Research Approaches and Methods*, pp. 35-59. Marcel Dekker, New York, (1981).
2. Rossiter, C.E., Browne, K., Gilson, J.C.: International Classification Trial of AIA Set of 100 Radiographs of Asbestos Workers. *Br. J. Ind. Med.* 45:538-543 (1988).
3. Weill, H., Jones, R.N.: The Chest Roentgenogram as an Epidemiologic Tool. *Arch. Environ. Health* 30:435-439 (1975).

*Proceedings of the VIIth International Pneumoconioses Conference*  
*Transactions de la VIIe Conférence Internationale sur les Pneumoconioses*  
*Transacciones de la VIIa Conferencia Internacional sobre las Neumoconiosis*

Part **I**  
Tome  
Parte



Pittsburgh, Pennsylvania, USA—August 23–26, 1988  
Pittsburgh, Pennsylvanie, Etats-Unis—23–26 août 1988  
Pittsburgh, Pennsylvania EE. UU—23–26 de agosto de 1988



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Public Health Service  
Centers for Disease Control  
National Institute for Occupational Safety and Health



## **Sponsors**

**International Labour Office (ILO)**  
**National Institute for Occupational Safety and Health (NIOSH)**  
**Mine Safety and Health Administration (MSHA)**  
**Occupational Safety and Health Administration (OSHA)**  
**Bureau of Mines (BOM)**

**September 1990**

## **DISCLAIMER**

**Sponsorship of this conference and these proceedings by the sponsoring organizations does not constitute endorsement of the views expressed or recommendation for the use of any commercial product, commodity, or service mentioned.**

**The opinions and conclusions expressed herein are those of the authors and not the sponsoring organizations.**

**DHHS (NIOSH) Publication No. 90-108 Part I**