INTER-READER VARIABILITY AMONG READERS USING ILO 1971 AND 1980 CLASSIFICATIONS OF THE PNEUMOCONIOSES

R. B. ALTHOUSE, M.S., B.S. • M. D. Attfield, Ph.D., B.Sc., F.S.S. • T. K. Hodous, M.D.
National Institute for Occupational Safety and Health, Morgantown, West Virginia

INTRODUCTION

The Coal Workers' X-ray Surveillance Program is a program mandated by the Occupational Safety and Health Act of 1969, which provides the opportunity for every working underground coal miner to have a chest radiograph taken at least once every five years.[1] One major objective of the program is to notify miners that their chest X-rays show signs of coal workers' pneumoconiosis (CWP), thereby allowing them the right to transfer to a low dust work environment. Broader objectives under the program provide for X-ray of all newly hired miners. Every participating miner is notified in writing of his X-ray status. The program is administered by NIOSH, Morgantown, WV.

Since the program inception in 1970, over 250,000 films have been processed. A diagram of the processing scheme is presented in Figure 1. Focusing on the X-ray processing scheme, it should be noted that a first reading is done at a facility close to the coal mine, usually by an 'A' reader. The films are then sent to NIOSH and distributed to 'B' readers selected on the basis of a randomized computer program. Thus all second readings are done by 'B' readers. The two results are summarized by a computer algorithm. If no consensus occurs, a third reading is requested from a 'B' reader. In order to become an 'A' reader, a person is required to correctly classify six X-rays from his/her own file, based on agreement with an expert panel. To become certified as a 'B' reader an examination given by NIOSH must be passed.

Reader variability between and within 'A' and 'B' readers has been an issue of concern since it reflects on the accuracy and consistency of data used for the transfer option, and on the perception of the program held by coal miners. In addition, it affects the ability to evaluate radiographic changes in individual coal miners, as well as overall estimates of prevalence and incidence over time.

Attfield documented inter-reader variability between and among 'A' and 'B' readers in a study in 1984.[2] The study was motivated by concerns brought to the attention of NIOSH, that first readings might be biased toward reading less abnormality since the choice of facility, and thus film reader, was made by the coal company. Findings of his study indicated that first readers, in fact, read higher levels of abnormality than second readers, but of more importance to this analysis, the study revealed high levels of variability, both between and among 'A' and 'B' readers.

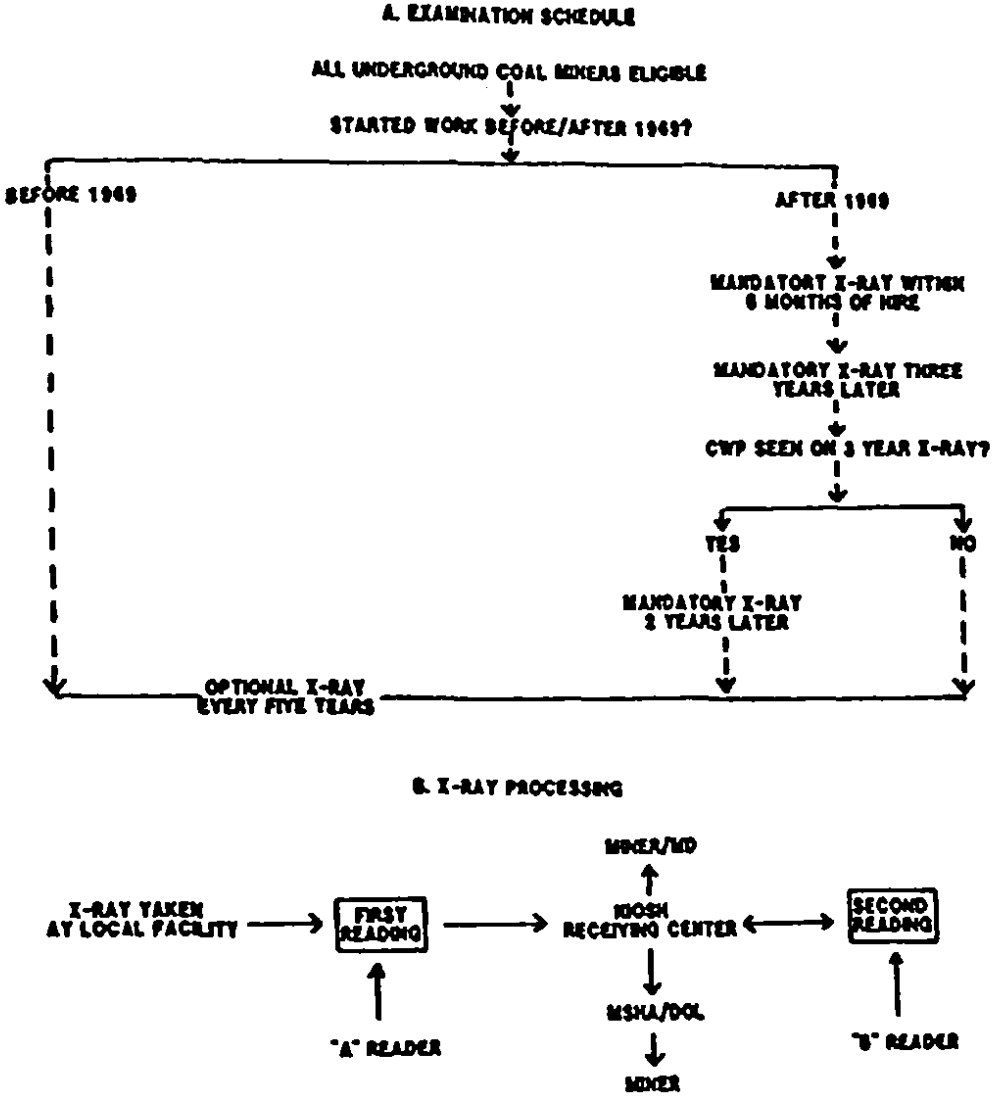
The chest X-rays reviewed by the Attfield study were taken in the CWXSP between 1978 and 1981. These films were classified using the 1971 ILO classification scheme.[3] During 1981 the use of the 1980 ILO classification was introduced into the program.[4] The change in classification required film readers to assess the overall profusion as well as the type of opacities in a different manner. Figure 2 presents an extracted section from each classification related to opacity type and profusion. In the 1971 scheme each reader assessed separately, and in a sequential manner the profusion of rounded, irregular, and combined opacities. The 1980 scheme requires the reader to specify the overall profusion, and then designate which opacity type, rounded or irregular, appears dominant.

This analysis was undertaken to determine if the change in classification affected the levels of reader variability noted previously. In order to do this, readings of chest X-rays evaluated by first and second readers on the 1980 ILO classification of the pneumoconioses, were compared to those read in the Attfield study. Comparisons between first and second readers for each classification were made for the profusion of opacities read, the types of opacities read, and within each reader group, for the range of prevalence of abnormal opacities which they reported.

METHODS

X-rays of readable quality for miners with 10 or more years of underground mining tenure were selected from films read during 1981–1987, by the 1980 ILO classification. These restrictions led to selection of a sample of 10,249 X-rays comparable to the 14,886 read during 1978–1981 by the 1971 classification. The tenure restriction of ten years eliminates the large weighting toward normal readers seen in miners with less than 10 years tenure. Only the first and second reading were considered for each film. General variability between and among first and second readers under each classification were compared in three ways:

1. Agreement based on profusion of small opacities for all first and second readers in each ILO group.
2. Agreement based on type of small opacities for all first and second readers in each group.
3. Range of variation within each group of readers based on prevalence of opacities, determined by the difference in highest and lowest ranked readers.



Broad processing scheme for X-rays in the CWXSP.

Tabulation of small opacities was based on the 12 point scale, ranging from category 0/− to 3/+. In all cases 0/− was combined with 0/0. These were combined to a 4 point major category scale in some cases. Prevalence comparisons were based on profusion of opacities, category 0/1 or greater.

Agreement was computed using the simple crude percentage. Since it seems reasonable to expect that the simple agreement of readings on the same films would be influenced by the underlying distribution of films, the kappa statistic[5] was used to adjust for the amount of agreement one would expect simply by chance. Kappa represents the excess agreement, having adjusted for expected chance agreement. It is computed as:

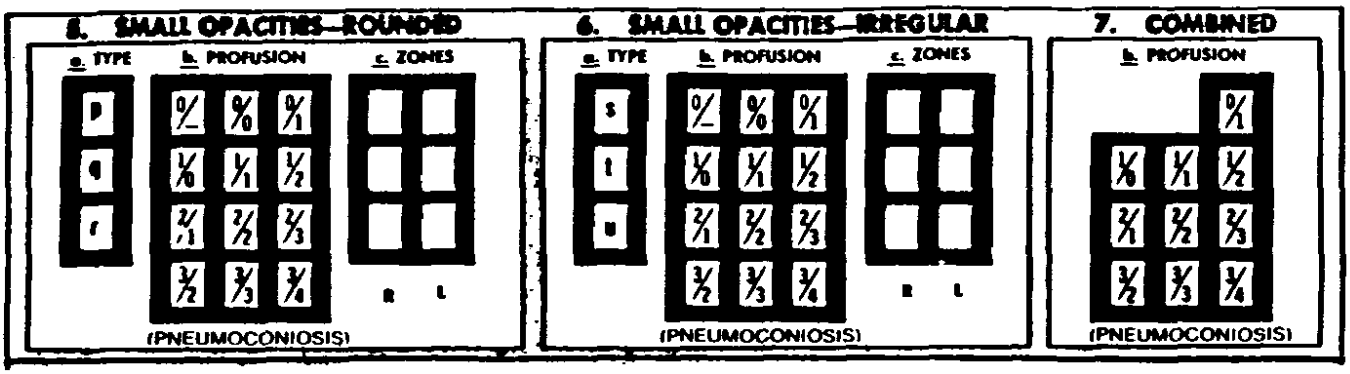
$$\text{kappa} = PC - PE / 1 - PE$$

where PC is crude agreement, PE is expected agreement, derived from marginal computations as in the chi-square test. A value of 0 represents no agreement, other than chance, while a value of 1 represents complete agreement.

For comparison of the two classification systems, the readings from the 1971 ILO were converted to a similar format to the 1980 ILO. A more detailed description is given elsewhere in these proceedings.[6] Briefly, for PRIMARY TYPE, if profusions for both rounded and irregular opacities were recorded by the 1971 ILO:

1. The larger of the two profusions was taken as the single profusion and the PRIMARY type was set to the type with the larger profusion.
2. If profusions were equal for each type the PRIMARY type was randomly assigned.

1971 Classification of Small Opacities



1980 Classification of Small Opacities

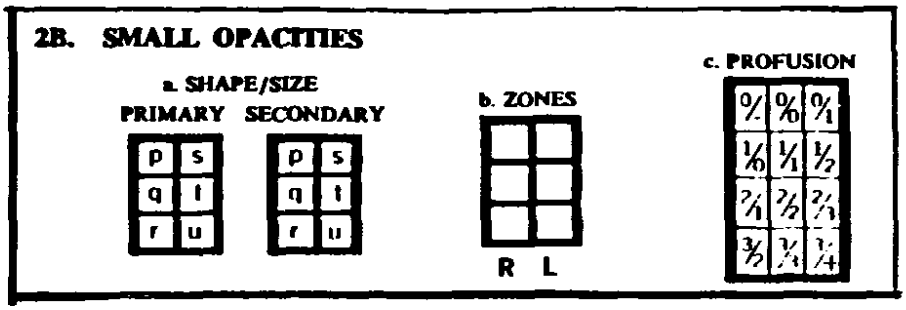


Figure 2

For the above situation SECONDARY TYPE was assigned as follows:

1. The SECONDARY type was set to the smaller of the profusion types.
2. The SECONDARY type was assigned the type not assigned to PRIMARY where profusions were equal.

If only one profusion type was recorded, that profusion was assigned to both PRIMARY and SECONDARY type.

RESULTS

Profusion of Opacities

Tabulations from the 1980 ILO group showed that 188 first readers and 20 second readers participated in evaluating the 10,249 X-rays. These values are close to the reported 196 first readers for the 1971 ILO, but indicate a decrease of 39% from the 33 second readers reported. Under 1980 ILO classification both first and second readers reported lower prevalence of category 0/1+ CWP, 12% for first and 4% for second readers. By the 1971 ILO classification scheme these values were 27% and 22% respectively. The lower prevalences under the 1980 ILO may be a reflection of an actual decrease in CWP, however, the differences between first and second readings shows a slight increase from 5% to 8% between the classification schemes. Figure 3 displays agreement between first and second readers further broken down into categories of the 12 point scale. Increases in reader agreement are seen in category 0/0 from 93 to 97% and in category 0/1 from 91 to 98%. Using the 1980 scheme decreases in agreement were evident in categories 1/0 and 1/2, from 30 to 27% and 30 to 16% respectively.

The agreement between first and second readers on the same films are presented in Table I. Overall crude agreement within major categories increased from 78% to 87% from the 1978 to the 1980 group. This overall increase is most likely due to the higher proportion of normal X-rays seen in the 1980 group. However the kappa statistic which adjusts for chance ageement is virtually unchanged. Further-

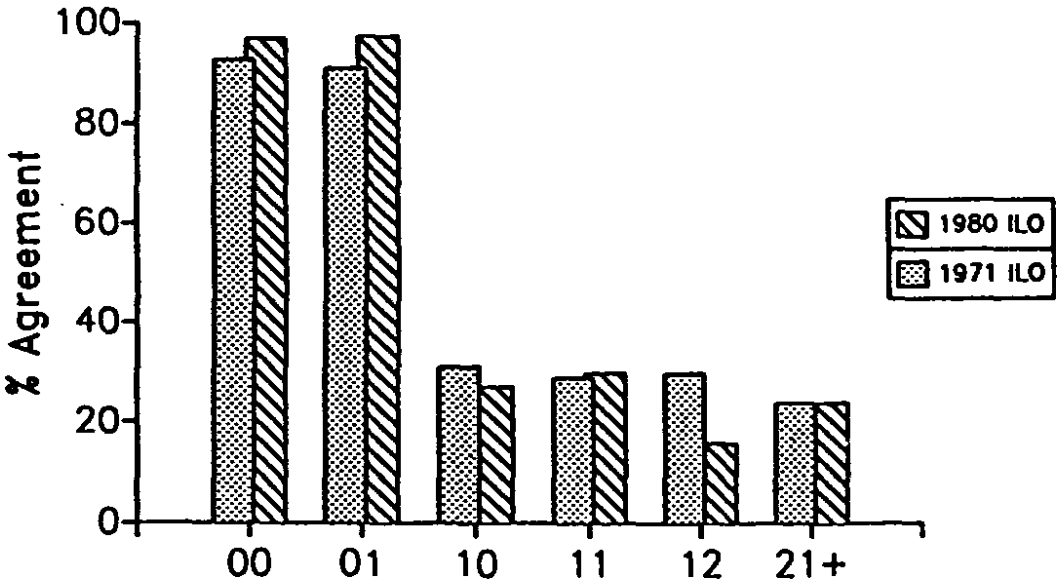


Figure 3. Agreement in profusion between reader 1 and reader 2 by minor category (agreement ± 1 minor category).

Table I

Agreement between 1st and 2nd Readers within Major Categories

1971 ILO		Reader 2			
		CAT 0	CAT 1	CAT 2+	TOTAL
Reader 1	CAT 0	10,793	927	30	11,750
	CAT 1	1,705	690	105	2,500
	CAT 2+	215	192	229	636
	TOTAL	12,713	1,809	364	14,886

(10+ years mining experience required.)

PC = 78%

PE = 69.6%

Kappa = .29

1980 ILO		Reader 2			
		CAT 0	CAT 1	CAT 2+	TOTAL
Reader 1	CAT 0	8,669	336	9	9,014
	CAT 1	757	251	24	1,032
	CAT 2+	82	56	65	203
	TOTAL	9,508	643	98	10,249

PC = 87%

PE = 82%

Kappa = .28

more the bias toward higher readings by first readers seen in the results using the 1971 ILO scheme continues to be seen. In the earlier study, 15% of readings classified as category 0 by the first readers were classified as category 1 or more by the second reader, while 8% of films classified as category 0 by the second reader were classified as category 1 or more by the first reader. This results in a 7% bias towards higher readings compared to the first readers. A similar calculation of findings from the 1980 classification group shows the bias reduced slightly to 5%. Both groupings show higher values in the lower diagonal elements than the comparable upper diagonal cells indicating that the change in classification scheme has not altered the tendency of first readers toward higher readings.

Type of Opacities

The agreement between readers in reading opacity type showed some parallel to profusion readings. On both classifications, first readers reported higher levels of rounded opacities as a percentage of all opacities than second readers, for primary type. The percentage for first readers increased slighty, from 74% to 75%, while that for second readers decreased from 63% to 56%. The difference in percentage between first and second readers showed a roughly 1.5 fold increase. A further breakdown into types of small opacities shown in Table II indicates that the change is due mostly to a shift in the second readers toward reading fewer rounded and more irregular type oapcities. Under both classification schemes first readers read the same percentages of rounded and irregular opacities, 74 and 26 percent respectively, although under the 1980 scheme a shift is seen towards reading more 's' and fewer 't' type irregular opacities. Second readers showed a much different pattern, with a large shift from fewer 'p' type rounded to more 't' type irregular opacities. This shift appears due to temporal changes in reading levels of rounded opacity types which have been occurring since 1978, discussed in more detail in Attfield, et al.,[6] rather than a change due specifically to the ILO classification scheme. The major point is that regardless of the underlying trends, the variability between first and second readers in the reporting of opacity types has not diminished under the 1980 ILO classification, but in fact has increased, compared to readings under the 1971 ILO scheme.

Table III presents data related to the final question, the range of variability among first and second readers. There was an overall decline in variability from 94% and 58% under the 1971 ILO to 68% and 31% under the 1980 ILO classification for first and second readers respectively. These reductions may in fact reflect actual levels of decrease in CWP, but the important element here is that first readers continue to read a range of disease at least twofold that of second readers.

DISCUSSION

The issue of variability in X-ray readings has implications, both as it relates to the CWXSP and to film readings in other research areas. In the CWXSP, the reduction of reader variability is of ongoing importance. One might expect that the change from the 1971 to the 1980 classification, given the focus on an overall evaluation of profusion of opacities, followed by a determination of primary and secondary opacity types, would increase agreement in readings between first and second readers.

Our findings indicate that this has not occurred. In summary, we found that first readers continue to read higher levels of prevalence of category 0/1+ opacities than second readers. The differences have increased rather than diminished. Reader agreement under the 1980 classification has increased in categories 0/0 and 0/1 only, decreasing in categories 1/0 and 1/2. Furthermore variability related to readings of small opacity types has increased. Among readers who read 50 or more films under each classification, the range of prevalence read by first readers as compared to second remains unacceptably high.

CONCLUSION

The findings of this study similar to those of Attfield[2] in-

Table II

Types of Opacities Reported as Primary Type for First and Second Readers

OPACITY TYPE	CLASSIFICATION			
	1971 ILO		1980 ILO	
	RDR1 (%)	RDR2 (%)	RDR1 (%)	RDR2 (%)
p	50.6	25.4	42.3	9.2
q	20.9	34.2	31.0	37.9
r	2.6	3.3	1.1	2.6
Total	74.0	62.9	74.4	49.7
s	14.2	19.7	18.1	19.7
t	11.0	16.7	7.1	30.4
u	0.8	0.7	0.3	0.3
Total	26.0	37.1	25.5	50.3

Table III

Range of Prevalence Category 0/1+ among First and Second Readers

ILO CLassification	First Reader	Second Reader
1971	1%–95%	3%–61%
1980	5%–68%	0%–31%

dicate that a quality control system is needed to minimize this problem in the CWXSP. This need exists in relation to both first and second readers. A system of feedback, notifying readers of their standing relative to all readers is one possibility. Specific details of the system could conceivably vary for 'A' and 'B' readers. A recently published paper documents unacceptably high reader variability in 'B' readers involved in asbestos medical surveillance.[7] The successful solution to this problem in the CWXSP might well serve as a model for other medical surveillance programs in which X-ray readers participate.

REFERENCES

1. Occupational Safety and Health Reporter: Health Standards for Underground Coal Mines (compilation of regulations 1970–1980). Bureau of National Affairs, Inc., Washington, DC., (1978).
2. Attfield, M.D., Althouse, R.B., and Reger, R.B.: An Investigation of Inter-Reader Variability Among X-ray Readers Employed in the Underground Coal Mine Surveillance Program. *Annals of the American Conference of Govt. Industrial Hygienists*. Vol. 14:401–409 (1986).
3. International Labour Office: ILO U/C International Classification of Radiographs of Pneumoconiosis, 1971. Geneva, ILO, 1972 (Occupational Safety and Health Series No. 22, revised).
4. International Labour Office: International Classification of Radiographs of Pneumoconiosis, (revised edition, 1980). Geneva, ILO, 1980. (Occupational Safety and Health Series No. 22 (Rev. 80)).
5. Fleiss, J.L.: *Statistical Methods for Rates and Proportions*, pp. 146–147. John Wiley and Sons, New York (1973).
6. Attfield, M.D., Hodous, T.K., Althouse, R.B.: A Comparison of the Profusion and Type of Small Opacities Reported with the 1980 and 1971 ILO Classification Using Readings from the Coal Workers' X-ray Surveillance Program. In these proceedings.
7. Ducatman, A.M., Yang, W.N., Forman, S.A.: 'B-Readers' and Asbestos Medical Surveillance. *Journal of Occupational Medicine* 30:8, pp. 644–647. (1988).

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

II

Pittsburgh, Pennsylvania, USA—August 23–26, 1988
Pittsburgh, Pennsylvanie, Etats-Unis—23–26 aout 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

CDC
CENTERS FOR DISEASE CONTROL

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)

November 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 II