

Observations on the Results of Multiple Readings of Chest Films In Coal Miners' Pneumoconiosis¹

Benjamin Felson, M.D., W. K. C. Morgan, M.D., Leonard J. Bristol, M.D.,
Eugene P. Pendergrass, M.D., Edgar L. Dessen, M.D., Otha W. Linton,
and Robert B. Reger

ABSTRACT—The authors studied the influence of multiple readings on the problem of detecting and evaluating coal workers' pneumoconiosis (CWP). There was an unexpected disagreement between three groups of readers, based on (a) poor film quality, (b) lack of experience with the classification systems employed, and (c) lack of familiarity with the roentgen manifestations of CWP. It is hoped that future surveys may be improved by films of higher quality and by a heightened awareness of the roentgen signs and symptoms of CWP among physicians.

INDEX TERMS: Pneumoconiosis • Radiographs, interpretation

Radiology 109:19-23, October 1973

THE PURPOSE of this report is to present the results of our investigation into the influence of multiple readings on the problem of detecting and evaluating coal workers' pneumoconiosis (CWP). Three significant events have made this study possible:

(a) Reports of the effect and implications of multiple interpretations of chest films by Yerushalmy *et al.* in the 1950's (2, 6, 7, 17-20). They recommended dual readings to improve accuracy in recognizing and evaluating pulmonary lesions; however, their material did not include subjects with pneumoconiosis.

(b) Multiple readings have been employed routinely in the diagnosis and categorization of pneumoconiosis in Great Britain for many years (1, 14) but have only recently been adopted in the U. S. There have been few attempts in either country to evaluate the use of multiple readings (3, 4, 12).

(c) A large group of chest films subjected to multiple readings and analysis, comprising a chest roentgenographic survey of about 60,000 U. S. coal miners, has recently been made available.

BACKGROUND

The Federal Coal Mine Health and Safety Act of 1969 (PL 91-173) gave every underground coal miner in the U. S. the right to have a free chest roentgenogram, and approximately 60,000 miners took advantage of this opportunity in 1971 and 1972. The examination was designed (a) to determine whether the miner was eligible for disability benefits and/or transfer to a less dusty

job, and (b) to provide an overall evaluation of the prevalence of pneumoconiosis among coal workers. The findings in the 55,730 miners who have received dual readings to date are described and analyzed in this report.

Public Law 91-173 empowered the National Institute for Occupational Safety and Health (NIOSH) of the U. S. Public Health Service (USPHS) to implement the chest examination provisions of the law for active miners. The law also specified that the films should be interpreted in terms of either the 1967 UICC²/Cincinnati classification or the 1968 extended International Labour Office (ILO) classification (15). It further stated that a miner diagnosed as having either (a) complicated pneumoconiosis, (b) simple pneumoconiosis at the category 2 or 3 level, or (c) category 1 simple pneumoconiosis developing with less than 10 years of exposure in a mine could be transferred to an area of the mine in which the respirable coal dust level was below 2 mg/m³. A miner with complicated pneumoconiosis also had the option of applying for disability benefits.

NIOSH elected to have single standard 35.6 × 43.2-cm (14 × 17-in.) postero-anterior chest films made at a radiographic facility or mobile x-ray unit in or near the miner's community. Each facility was required to submit six films for certification. If NIOSH consultants considered them technically satisfactory, the facility was approved for this part of the program.

METHOD

A multiple reading method, the so-called ABC

¹ Presented at the Fifty-eighth Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, Ill., Nov. 26-Dec. 1, 1972.

² Union Internationale Contre le Cancer.

system, was devised by NIOSH for use in interpreting the films. The first interpretation was made by a local physician, the A reader, who was often the physician in charge of the facility at which the roentgenogram was obtained. The NIOSH decision to involve local physicians in the program was based upon (a) the requirement that the first round of roentgen examinations be accomplished within 18 months after the law was passed, (b) a lack of federal facilities to carry out the examinations, and (c) a desire to integrate the project with the health care resources that ordinarily served the miners and their community.

To qualify as an A reader, the physician had to either participate in one of the two-day seminars conducted periodically for NIOSH by the American College of Radiology (ACR) or submit to NIOSH six films interpreted properly according to the UICC/Cincinnati or ILO classification. Most qualified by attending one of the seminars, but many were certified on both counts. The A readers were drawn from several medical disciplines, including radiology, internal medicine, chest disease, industrial medicine, and general practice. Of the 175 A readers, 139 were radiologists and 36 were from other disciplines.

After the A reader had recorded his interpretation on a special form, the radiograph and report were sent to the Appalachian Laboratory for Occupational Respiratory Diseases (ALFORD), where the findings were recorded. All films were then forwarded to B readers, 24 radiologists with long experience with pneumoconiosis drawn from three radiology departments, two of which were at university hospitals. The B reader then returned the films to ALFORD together with his interpretations.

If both the A and B readers interpreted the film as normal, or if they agreed on the major degree of profusion of small lesions (simple CWP) and the presence and extent of large opacities (complicated CWP), the reading was accepted as final. If there was disagreement regarding any of these parameters, the film was sent to a C reader for arbitration. Films in which there was disagreement only in the type (irregular or rounded), size, or zone distribution of small opacities were not submitted for a C reading. There were seven C readers, all experienced in pneumoconiosis interpretation, who served as staff radiologists in one of the three radiology departments from which the B readers were recruited, though the B and C readers of a given film were never from the same institution. Three of the C readers were members of the USPHS Panel of Radiology Consultants to NIOSH. The C reader was given the A and B

readings but was not told which was which. The film and its C reading were returned to ALFORD for recording and storage. ALFORD notified the miner and his personal physician of the results, including the need for further medical attention. The Bureau of Mines was contacted if job transfer rights were involved. The data used in this paper came from ALFORD reports for the period ending July 31, 1972.

RESULTS

The findings of the A and B, A and C, and B and C readers in regard to small opacities, *i.e.*, simple CWP, are compared in TABLES I-III. The A and B readers agreed in about three-fourths of their readings (68.9% of negative readings and 5.5% among the categories of positive readings) (TABLE I). They disagreed in about one-fourth of the films, of which 21.9% were read as negative by either A or B and 4.0% were read as positive by both but at different levels (*e.g.*, category 2 *vs.* category 1). In general, the A readers showed a much stronger tendency to read a film as positive than did the B readers. If films read as normal by both readers were excluded, roughly 17,000 films remained. Of these "positive" films (as indicated by one or both readers), one reader disagreed with the other in 82.2% and the two agreed in only 17.7%.

The collective comparison between the A and C readers is shown in TABLE II. Of the 14,594 films on which the A and B readers disagreed, the C readers found 225 unreadable. Of the remaining 14,369 readings, there was agreement in 3,552 (24.4%) and disagreement in 10,817 (75.3%). In only 10.1% did both A and C readers agree that the chest was normal. In 9,336 films only one reader gave a normal reading; in 93.5% of these, it was the C reader who recorded the negative reading. The tendency of the A reader to ascribe higher categories of involvement was again noted, especially when one reader read the film as 0. Our statistics do not presently allow a comparison of the level of agreement for A readers who were radiologists and those from other medical disciplines.

The B and C readings are compared in TABLE III. Again, these were only the films in which the A and B readers disagreed. There was 69.8% agreement and 30.2% disagreement between the two groups. Of the 3,106 cases in which only one reading was negative, 52% were C readings and 48% were B readings. Among the positive readings, the B readers tended to list a higher category than the C readers, but this was not as striking as the discrepancy between the A and C readers.

TABLE I: COLLECTIVE EXPERIENCE, A vs. B*

		Reader A					
		Category of Simple CWP					
		0	1	2	3	Total	
READER B	Category of Simple CWP	0	38,425 (68.9%)	8295 (14.9%)	1,717 (3.1%)	44 (0.1%)	48,481 (87.0%)
		1	1,651 (3.0%)	1,913 (3.4%)	941 (1.7%)	90 (0.2%)	4,595 (8.3%)
		2	375 (0.7%)	656 (1.2%)	1,064 (1.9%)	209 (0.4%)	2,304 (4.2%)
		3	22 (0.1%)	45 (0.1%)	192 (0.4%)	91 (0.2%)	350 (0.7%)
	Total		40,473 (72.6%)	10,909 (19.6%)	3,914 (7.1%)	434 (0.8%)	55,730 (100%)

* All percentages are based on 55,730 cases. This table relates to the distributions of small opacities, regardless of whether or not large opacities were also present.

Size of large or conglomerate opacities (complicated CWP) was omitted as a parameter in the present study. There were 1,181 instances in which the A or B reader or both indicated the presence of complicated CWP. In 35.6% the two agreed; in 41.8% the A readers recorded it while the B readers did not; the reverse occurred in 22.5%. Again, the A readers overread as compared with the B readers.

When A and C readers were compared in regard to complicated CWP, there were 943 instances in which one or both gave a positive reading. The two agreed in 35%, while the A readers alone gave a positive reading in 45% and only the C reading was positive in 18%. When B and C readers were compared in respect to the 697 films which one or both read positive for large opacities, there was agreement in 52%; the films were read as positive by B alone in 25% and by C alone in 22%. Thus the level of positivity of the B and C readings of large opacities was approximately the same.

DISCUSSION

Several factors may be responsible for the disappointing level of divergent interpretations among the A, B, and C readers in this first round of roentgen examinations, notably (a) inherent inter-observer disagreement, (b) films of poor quality, (c) a basic lack of experience with the UICC/Cincinnati and ILO classification systems on the part of many of the A and B readers, and (d) a general lack of familiarity with the roentgen manifestations of CWP.

In the 1950's, Yerushalmy *et al.* studied the

TABLE II: COLLECTIVE EXPERIENCE, A vs. C*

		Reader A					
		Category of Simple CWP					
		0	1	2	3	Total	
READER C	Category of Simple CWP	0	1,468 (10.1%)	7,257 (49.7%)	1,441 (9.9%)	32 (0.2%)	10,198 (69.9%)
		1	537 (3.7%)	1,369 (9.4%)	801 (5.5%)	53 (0.4%)	2,760 (19.0%)
		2	65 (0.4%)	332 (2.3%)	628 (4.3%)	201 (1.4%)	1,226 (8.4%)
		3	4 (0.1%)	8 (0.1%)	86 (0.6%)	87 (0.6%)	185 (1.3%)
		U/R	59 (0.4%)	102 (0.7%)	59 (0.4%)	5 (0.1%)	225 (1.5%)
	Total		2,133 (14.6%)	9,068 (62.2%)	3,015 (20.7%)	378 (2.6%)	14,594 (100%)

* All percentages are based on 14,594 cases.

TABLE III: COLLECTIVE EXPERIENCE, B vs. C*

		Reader B					
		Category of Simple CWP					
		0	1	2	3	Total	
READER C	Category of Simple CWP	0	8,558 (58.6%)	1,316 (9.0%)	309 (2.1%)	15 (0.1%)	10,198 (69.9%)
		1	1,315 (9.0%)	988 (6.8%)	424 (2.9%)	33 (0.2%)	2,760 (18.9%)
		2	146 (1.0%)	355 (2.4%)	564 (3.9%)	161 (1.1%)	1,226 (8.4%)
		3	5 (0.1%)	20 (0.1%)	80 (0.5%)	80 (0.5%)	185 (1.3%)
		U/R	117 (0.8%)	75 (0.5%)	28 (0.2%)	5 (0.1%)	225 (1.5%)
	Total		10,141 (69.5%)	2,754 (18.9%)	1,405 (9.6%)	294 (2.0%)	14,594 (100.0%)

* All percentages are based on 14,594 cases.

results of multiple readings of groups of chest films by expert panels of radiologists and chest physicians (2, 4, 5, 17-20). They found inter-individual disagreement on all films (both negative and positive) to be less than 5%, whereas among the abnormal films it rose to 30%. Intraindividual disagreement in the positive group was about 20%. These figures applied not only to recognizing lesions but also to evaluating their significance. Incorporating a rereading of the same films by the same reader or different readers lowered the false-

negative figures (positive films read as negative) significantly but had the opposite effect on the number of false-positives (negative films read as positive). It was also noted that the inconsistency of the readings increased in proportion to the number of categories or terms used in describing the lesions. (Such intra- and inter-observer disagreements are not confined to radiology, of course.) Yerushalmy's conclusions certainly apply to the present study. In fact, we had even more categories to consider and a gradation from one category to the next to judge, adding to the difficulty of interpretation and the likelihood of differences of opinion.

The method employed by Yerushalmy *et al.* to determine the "correct" answer was unanimous agreement among the experts, either on individual reading or at a panel session. This was not the case in our series. We accepted the C reading as the final verdict. However, a recent study by Reger and Morgan (12) indicated that inter-observer error varied between 19 and 33% among positive readings for CWP by four "experts," two of whom were members of the NIOSH Radiology Consultants Panel. Since the experts so often disagree among themselves, there is certainly no assurance that the C reading is the valid one. However, the fact that the C reader was the most experienced of the three readers added some credibility to his interpretations.

A higher rate of disagreement between the A readings and the B and C readings was found as expected. Some consolation can be gleaned if a discrepancy of one category is accepted as agreement, in which case the level of agreement rises considerably. For example, if the two readings on a film are 0 and 1, 1 and 2, or 2 and 3 and this difference is accepted as agreement, the concordance between A and B readers rose from 74.5% to 95.9%, while that between A and C readers rose from 24.4% to 88.8% and that between B and C readers rose from 69.8% to 96.3%. Even so, the level of disagreement between the A and B readings and the C readings was disappointingly high and far beyond that anticipated from inherent problems of interobserver variation. Obviously, other factors probably contributed to this situation.

It has been shown that film quality exerts a significant effect on the categorization of pneumoconiosis (10, 11, 16). In a recent study, Reger *et al.* showed that experienced readers tend to compensate for under- and overpenetrated films, while a neophyte may not (13). Because of the necessity of completing the first round at an early date and because of the difficulty of finding in-

dividual miners and persuading them to return for retakes, the A and B readers were requested to interpret all films unless they were absolutely unreadable. Despite this, more than 10% of some of the film batches submitted were subsequently considered unreadable by C readers. As time passed, it became apparent that all readers would have to become more critical if the effect of poor film quality was to be reduced. Thus the fact that the quality of many of the films in this study was clearly below standard may well have been a cause of some of the disparities in reading. At present, the role played by film quality in this regard is not measurable.

There was a widespread lack of familiarity with the present pneumoconiosis classification systems among both the A and B readers. For example, only the three members of the USPHS Panel of Radiology Consultants to NIOSH had had prior experience with the UICC/Cincinnati classification. A crash training program in classifying had to be set up, and the ACR was authorized by NIOSH to do this. Nevertheless, although the weekend seminars and home study material prepared by the ACR were acknowledged to be extremely helpful by most A and B readers, they were not always effective in raising the proficiency in classification to the required level. Some of the difficulties in coding appeared to lessen with time as inexperienced readers became more familiar with the intricacies of the system.

Another problem was the absence of a set of standard films already classified by groups of experts which A and B readers could compare to the films they were reading. These standard sets have only recently become available, too late to be of use to the A, B, and C readers.

Another cause of difficulty was unfamiliarity of many of the A readers with the basic roentgen characteristics of CWP. Some of them did not have a clear concept of the many roentgen details of CWP and were also unaware of the wide variation in appearance of the healthy chest (these shortcomings also applied to some of the B readers). The difficulty in distinguishing category 1 CWP from a normal lung is emphasized by the frequency with which the C reader recorded a chest as normal while the A or B reader called it positive. The A readers had a very strong tendency toward the positive side in these instances. Kerr has been critical of the lack of training provided by our medical schools in occupational diseases in general and the pulmonary aspects in particular, as well as the limited number of publications on the subject other than in journals of occupational health (9).

SUMMARY AND CONCLUSIONS

The extent of disagreement in the A, B, and C readings for this group of miners' films in the first round of medical examinations is felt to be unacceptable. It is ironic that a system designed to determine proper compensation for respiratory disease is based on a test with a level of reliability lower than that which is acceptable for many other diagnostic techniques.

Several ways of improving the interpretation of future chest films might be suggested. Most obvious to NIOSH and ACR task force participants is the need to improve the quality of the films. Too many roentgen facilities fail to produce routine films of consistently good diagnostic quality (8), whether for pneumoconiosis or other purposes. This subject will be dealt with in more detail by other participants in this project, and an additional report on this aspect will appear shortly.

Inexperience of physicians with the disease and its classification has lessened as a result of Public Law 91-173 and the NIOSH program, and many more physicians have come to have a heightened awareness of the effects of CWP and related occupational respiratory diseases. NIOSH has indicated that it will continue to create and sponsor physician education projects to overcome the difficulties encountered in this first effort. Whether this will eventually produce more experts in the roentgen aspects of CWP remains to be seen. A pneumoconiosis film-reading proficiency test which may be used in the future as a qualifying method for A, B, and C readers is almost ready for trial.

Efforts are being made to devise a better system to interpret films taken in the second round. It is of paramount importance that an effective method be found to accurately interpret large numbers of chest films, not only those of coal workers but those of men exposed to other types of dust such as asbestos and beryllium who will soon be studied in a similar manner as the result of new legislation.

REFERENCES

1. Ashford JR: The classification of chest radiographs for coalworkers' pneumoconiosis. Scientific Department, National Coal Board, 1959
2. Cochrane AL, Garland LH: Observer error in interpretation of chest films. International investigation. *Lancet* 2: 505-509, 13 Sep 1952
3. Fletcher CM: The clinical diagnosis of pulmonary emphysema—an experimental study. *Proc Roy Soc Med* 45: 577-584, 1952
4. Fletcher CM, Oldham PD: The problem of consistent radiological diagnosis in coal miners' pneumoconiosis. Experimental study. *Brit J Industr Med* 6:168-183, Jul 1949
5. Garland LH: On the scientific evaluation of diagnostic procedures. *Radiology* 52:309-327, Mar 1949
6. Garland LH: On the reliability of roentgen survey procedures. *Amer J Roentgen* 64:32-41, Jul 1950
7. Garland LH, Miller ER, Zwerling HB, et al: Studies on the value of serial films in estimating the progress of pulmonary disease. *Radiology* 58:161-175, Feb 1952
8. Jacobson G, Bohlig H, Kiviluoto R: Essentials of chest radiography. *Radiology* 95:445-450, May 1970
9. Kerr LE: Coal workers' pneumoconiosis in an affluent society. *Public Health Rep* 85:847-852, Oct 1970
10. Liddell FDK: The effect of film quality on reading radiographs of simple pneumoconiosis in a trial of x-ray sets. *Brit J Industr Med* 18:165-174, Jul 1961
11. Pearson NG, Ashford JR, Morgan DC, et al: Effect of quality of chest radiographs on the categorization of coalworkers' pneumoconiosis. *Brit J Industr Med* 22:81-92, Apr 1965
12. Reger RB, Morgan WKC: On the factors influencing consistency in the radiologic diagnosis of pneumoconiosis. *Amer Rev Resp Dis* 102:905-915, Dec 1970
13. Reger RB, Smith CA, Kibelstis JA, et al: The effect of film quality and other factors on the roentgenographic categorization of coal workers' pneumoconiosis. *Amer J Roentgen* 115: 462-472, Jul 1972
14. Stradling P, Johnston RN: Reducing observer error in 70-mm. chest-radiography service for general practitioners. *Lancet* 1:1247-1250, 18 Jun 1955
15. UICC/Cincinnati classification of the radiographic appearances of pneumoconioses. A cooperative study by the UICC committee. *Chest* 58:57-67, Jul 1970
16. Wise ME, Oldham PD: Effect of radiographic technique on readings of categories of simple pneumoconiosis. *Brit J Industr Med* 20:145-153, Apr 1963
17. Yerushalmy J: Reliability of chest radiography in diagnosis of pulmonary lesions. *Amer J Surg* 89:231-240, Jan 1955
18. Yerushalmy J: Reliability of chest radiography. [In] Naclerio EA, ed: *Bronchopulmonary Diseases. Basic Aspects, Diagnosis and Treatment.* New York, Hoeber, 1957
19. Yerushalmy J, Harkness JT, Cope JH, et al: Role of dual reading in mass radiography. *Amer Rev Tuberc* 61: 443-464, Apr 1950
20. Zwerling HB, Miller ER, Harkness JT, et al: The clinical importance of lesions undetected in a mass radiographic survey of the chest. *Amer Rev Tuberc* 64:249-255, Sep 1951

Department of Radiology
University of Cincinnati College of Medicine
Cincinnati, Ohio 45229

