ENVIRONMENTAL RESEARCH 6, 60–67 (1973)

The Radiographic Categorization of Coal Workers' Pneumoconiosis by Lay Readers

WILLIAM L. PETERS, ROBERT B. REGER, AND WILLIAM KEITH C. MORGAN

Appalachian Laboratory for Occupational Respiratory Diseases, National Institute for Occupational Safety and Health, U. S. Public Health Service, Morgantown, West Virginia 26505 and Department of Medicine, Medical Center, West Virginia University, Morgantown, West Virginia 26505

Received May 5, 1972

An investigation was carried out to show whether lay readers could be trained to interpret chest radiographs for coal workers' pneumoconiosis. It was shown, first, that lay readers can be trained for this purpose, and second, that they are no more prone to interobserver variation than are physicians. Once again, the extent of the interobserver variation among experienced readers in the categorization of the films for coal workers' pneumoconiosis was shown to be greater than is commonly realized.

Necessary for the diagnosis of coal workers' pneumoconiosis (CWP) are a history of exposure to coal dust and certain relatively distinctive features in the chest radiograph. The severity of the condition is assessed by the extent, profusion and character of the opacities that are present in the chest film. Thus, the chest radiograph is universally accepted as the sole practical method of diagnosing the condition during life. Moreover, the measurement of progression in CWP depends on a comparison of chest films taken at intervals. The validity for the radiographic classification of CWP relates to the fact that there is a linear relationship between the coal dust content of the lungs and radiographic category (Rossiter, 1972). Furthermore, since the passage of the 1969 Federal Coal Mine Health and Safety Act, miners with category 2 or more CWP have the right to transfer to an area of the mine where the respirable dust is below 2 mg/m^3. It should, therefore, be apparent that the chest radiograph is of critical importance in the diagnosis, investigation, and prevention of CWP.

Because of the importance of the chest film in the field of occupational respiratory disease, it is often assumed that the interpretation of pneumoconiosis requires long experience, great sagacity, and extraordinary intellectual abilities. The ideal interpreter is assumed to have the wisdom of Solomon, to be as observant and as logical as Sherlock Holmes, and to show as fine an appreciation of shadows and tones as Albrecht Dürer. The mystique concerning radiographic interpretation of pneumoconiosis persists despite the repeated demonstration of marked inter- and intraobserver variation (Fletcher and Oldham, 1949; Reger and Morgan, 1970). In reality, the categorization of pneumoconiosis depends on guessing, or in putting it more euphemistically, estimating the number of dots or opacities present on a chest film. Moreover, it is a task accompanied by more than its fair share of tedium. We, therefore, decided to see whether lay personnel could be trained to read chest films for pneumoconiosis.

MATERIALS AND METHODS

The chest radiographs used in this study were a sample of those in the Interagency Study of CWP. In 1969 the U. S. Public Health Service and the Bureau of Mines of the Department of the Interior commenced a large epidemiological study of the prevalence and progression of CWP in U. S. coal miners. Thirty-one mines in widely different regions were selected for inclusion in the study. Each miner employed at these 31 mines was offered a medical examination consisting of posteroanterior and lateral chest films, and some simple tests of pulmonary function (Seaton *et al.*, 1972). The medical examinations will be repeated at 3- and then 5-year intervals.

The radiographs taken in this study were independently interpreted by three readers using the UICC classification (UICC committee, 1970). All were physicians and all had long experience in classifying and interpreting chest films for pneumoconiosis. In regard to major categories, namely, 1, 2, and 3 simple CWP and stages A, B, and C complicated pneumoconiosis, a consensus reading was accepted. In the occasional event when all disagreed (less than 5% of the films), a fourth reader was enlisted and consensus thus obtained.

Two lay readers, one a secretary and the other an X-ray technician, were trained by one of the three physician readers (WKCM). After 2–3 months practice, during which time they simultaneously but independently categorized all films being interpreted by their instructor, they were asked to independently categorize all the chest radiographs from seven Appalachian coal mines. Both lay readers referred to standard films while they were interpreting the films. No attempt was made to train them to recognize conditions other than pneumoconiosis, their sole purpose being to assess the extent, profusion, and size of the large and small opacities present on the film.

The UICC classification of pneumoconiosis is based on Liddell's elaboration of the International Labour Organization's Classification (Liddell and May, 1966). In this classification, all categories are expressed as two digits separated by a stroke, e.g., 2/1. The first of the two digits represents the major category in which the film was placed while the second represents the category that was also considered but rejected. Thus, category 1/2 is a film that was placed in category 1 but in which category 2 was also considered; in contrast, category 2/2 means that no other category was considered. The usual four-point scale (0, 1, 2, and 3) is thereby extended to 12 points, 0/–, 0/0, 0/1, 1/0, 1/1, 1/2, 2/1, 2/2, 2/3, 3/2, 3/3, 3/4. It is assumed that the readings represent stages on a continuum that starts with no pulmonary dust deposition and ends with maximal dust deposition.

The three professional readers categorized 1471 chest radiographs. The lay readers interpreted 1,466 and 1,465 chest films, respectively, since around half-a-dozen films had been loaned to physicians in order to help in the diagnosis and follow up of nonoccupationally related chest conditions.

RESULTS

Before delving deeply into the extent of interobserver variation of the professional and lay readers involved in this study, it is informative to first obtain some general idea of the number of radiographs placed in each category by each of the

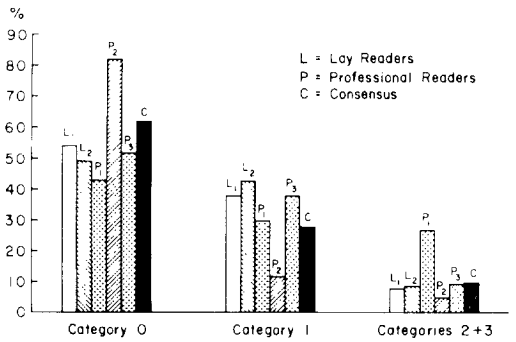

FIG. 1. Categorization of 1471 films by various readers.

five readers. The pertinent data are shown in Fig. 1. It is immediately apparent that the three physician readers varied greatly among themselves. The resemblance of the two lay readers to their trainer (P_3) is almost as striking. At first sight it seems that the two lay readers agreed closely with the consensus. This impression is somewhat misleading and does not necessarily mean that the lay readers were in agreement with the consensus more often than the professional readers. In actuality, as will become evident later, the lay readers agreed with the consensus less frequently than did the three physician readers.

Agreement in Regard to Presence or Absence of Pneumoconiosis

The percentage agreement in regard to what constitutes a positive and negative film for the lay and professional readers is shown in Table 1. All readers are compared to each other and to the consensus. The lack of unanimity between the professional readers is obvious; while P_1 and P_3 agreed 72% of the time, P_1 and P_2 agreed only 58% of the time. At first sight these differences seem excessive, until it is remembered that the majority of films used in this study were read by virtually all the readers as either category 0 or category 1, and that the difference between positivity and negativity may be one subcategory, viz., the difference between a reading of 0/1 and 1/0. It must also be borne in mind that the relatively high measure of agreement between P_3 and the consensus is a function of the fact that he "reads" in the middle of the other two.

TABLE 1
PERCENTAGE AGREEMENT (POSITIVE AND NEGATIVE)

	L_1	L_2	P_1	P_2	P_3	C
L_1	—	70.1	67.3	64.9	72.7	72.4
L_2			69.5	60.1	73.9	70.6
P_1			—	57.5	71.9	79.3
P_2				—	63.7	77.7
P_3					—	85.5
C						—

A "trainer effect" on the lay readers is evident from the data. They agreed with their trainer 73 and 74%, respectively, as to positive and negative films. Furthermore, the lay readers agreed among themselves on seven out of 10 occasions.

Agreement in Regard to Major Categories

Table 2 shows that when major categories of simple CWP are considered, all two-way contrasts are slightly lower than those in which positivity and negativity only were considered. This is to be expected since the subdivision of all positive films into their various categories allows more opportunity for a divergence of opinion. The data, nonetheless, reflect very similar findings to those included in Table 1. The extent of the divergence between the professional readers is appreciable, and again P_3 is closer to the consensus than the other professionals. A significantly higher measure of agreement exists between the lay readers and P_3, in contrast to the lay readers and P_1 and P_2. Again the influence of P_3 as a trainer is evident. It is noteworthy that agreement was present between the lay readers and P_1 and P_2 as often, and in one instance, more often, than it was between the three professionals.

The lack of agreement between the lay readers and the consensus category is outwardly a little surprising. Part of the explanation lies in the fact that lay readers did not participate in the original readings from which the consensus was derived. In addition, a more subtle consideration has to be taken into account. When all the interrelationships between readers are considered relative to the consensus, a situation may result which tends to obscure the true facts. For example, although L_1 and P_3 place almost the same proportion of films in category 0, and although both agree closely in this regard with the consensus, they need not necessarily be in agreement about the same films. Thus, some of the films that L_1 calls category 0 are interpreted as category 1 by P_3, while in a roughly equal number the converse is true. The net result is that although both place roughly the same proportion of film in category 0, they are not necessarily the same films. Similarly, while L_1 and P_3 might agree about a certain film, the consensus interpretation might agree with neither. Alternatively, while L_1 might agree with the consensus, P_3 might not. If these agreements and disagreements

TABLE 2
PERCENTAGE AGREEMENT (MAJOR CATEGORIES)

	L_1	L_2	P_1	P_2	P_3	C
L_1	—	63.2	50.9	60.4	64.7	64.7
L_2		—	50.1	56.0	66.0	63.2
P_1			—	49.6	54.4	67.6
P_2				—	58.9	74.9
P_3					—	82.3
C						—

TABLE 3
PERCENTAGE AGREEMENT (MINOR CATEGORIES)

	L_1	L_2	P_1	P_2	P_3
L_1	—	38.3	34.3	24.3	40.1
L_2		—	36.8	35.6	44.9
P_1			—	23.6	33.7
P_2				—	30.3
P_3					—

occur inopportunely as described above, they tend to exaggerate the differences between the consensus verdict and the lay readers.

Agreement in Regard to Minor Categories of CWP

A subsample of 588 films was used in this analysis. On the basis of the data from Tables 1 and 2, we were surprised that any agreement at all resulted when using the full 12-point scale of the UICC classification.

Surprisingly and gratifyingly, the lay readers agreed with their trainer (P_3) more than 40% of the time. The proportion of times they agreed with the other two expert panel members was significantly lower (Table 3). Moreover, the percentage of agreement between the professional panel members was indeed low. In general, these data are merely an extension of Table 2 and are of little importance except that they show the same general pattern of extreme variability.

Rather than considering disparities between readers in terms of major and minor categories of CWP, a different and it is felt more realistic approach would be to compare the readers according to the degree of divergence that might be expected; thus, a difference of one subcategory either up or down the scale would seem a reasonable and indeed acceptable difference. For example, if P_3 reads a film as 2/1, then 1/2 and 2/2 would be considered as acceptable readings. Using this approach, the 12-point scale is reduced to a sliding 4-point scale which, in many instances, bridges the gap between categories, i.e., 0/1 to 1/0, 1/2 to 2/1, etc.

When this approach is used, the variability between readers decreases (Table 4). Note that although the same general pattern exists in the two-way contrasts presented, an acceptable measure of agreement is present much more often. In only two instances, viz., L_1 vs P_2, and P_1 vs P_2, was agreement lower than anticipated.

TABLE 4
PERCENTAGE "SATISFACTORY" AGREEMENT (MINOR CATEGORIES)

	L_1	L_2	P_1	P_2	P_3
L_1	—	82.4	75.9	62.9	82.2
L_2		—	75.9	76.1	85.5
P_1			—	57.3	78.6
P_2				—	77.0
P_3					—

TABLE 5
CATEGORIZATION OF PMF, BY READER

Reader	Total films considered	Films classified as PMF, number	% Total
L_1	1466	60	4.1
L_2	1465	28	1.9
P_1	1471	54	3.7
P_2	1471	30	2.0
P_3	1471	34	2.3
C	1471	34	2.3

Agreement in Regard to Complicated Pneumoconiosis

Less of an effort was made to teach the lay readers how to assess complicated pneumoconiosis since it is felt that the differential diagnosis of larger opacities requires a more sophisticated medical background. The number of films classified by each of the readers is shown in Table 5. The reason why L_1 read appreciably more complicated pneumoconiosis than all the other readers other than P_1, is related to the fact that many of the large opacities which she observed were probably nonpneumoconiotic and were produced by other conditions such as tuberculosis or pleural thickening. In addition the fact that she was a trained X-ray technician with some knowledge of radiology may have an effect on her performance. The slightly lower number of subjects diagnosed by L_2 as having complicated pneumoconiosis is a consequence of her failing to recognize the ill-defined type of complicated pneumoconiosis.

DISCUSSION

For reasons given earlier, the categorization of the chest radiograph for pneumoconiosis is still critical in the prevention of industrial pulmonary disease. It is important, nevertheless, to realize the limitations of chest radiography and even more importantly to the limitations of those who interpret the chest film. The belief that all radiographs fall neatly into a particular category is widespread among radiologists and almost universal among the legal fraternity.

For the chest radiograph to continue to serve a useful function in the diagnosis and prevention of CWP necessitates the acceptance of the fact that all films fall on a continuum from no dust in the lungs to maximal dust, whatever that might be. Although a group of physicians may arbitrarily designate a series of films as standards, this is a subjective exercise and no other chest film will ever exactly correspond to any of the standards. At the best, all that can be hoped for is for the interpreter to make, between the standard and unknown radiograph, a reasonably close approximation in appearance, and hence in coal dust content. If the absolute approach to categorization is discarded, viz., the belief that an unknown chest radiograph can fit into one category only, the statistical (or epidemiological) approach substituted, the radiograph remains a most useful investigation of the pulmonary status of workers exposed to a variety of harmful dusts.

To be compelled to sit and categorize between 100–300 chest films a day is tedious in the extreme. Notwithstanding the tedium, it is obviously imperative that at least one physician should screen the films for significant disease other than CWP. However, the usual routine for most epidemiological studies is for at least three or four physicians to independently read the films so that the "maximal" accuracy is obtained and a majority verdict is available. If agreement cannot be reached by means of a majority verdict, a conference between the readers is held and a decision forced. The latter, assumed by many to be the most scientific approach, in reality results in the most opinionated and dogmatic person prevailing. Because of the above arguments, we decided to see whether it was possible to train lay personnel to categorize radiographs for CWP. Lay readers have many advantages over their professional colleagues, viz., they are more flexible and less set in their ways, they are more likely to report what they see rather than "to interpret" what they see, and finally they are less likely to become bored and hence are less likely to let their attention wander.

The results of this study show that lay readers can be trained to classify pneumoconiosis; moreover, their variability is no worse than is that of physician interpreters. The marked interobserver variation that has been revealed in this study among experienced readers is not a new observation; however, the extent of the problem is seldom appreciated. Since so much depends on the "accurate" interpretation of the chest radiograph, every effort should be made to limit interobserver variation. This can best be effected by testing the interpreters at periodic intervals against a set of standard films. The latter should include at least 150 radiographs and should include all categories of pneumoconiosis but with the greatest proportion in the range 0/1 to 2/1. Readers who deviate significantly from the standards or who show appreciable intraobserver variation should not be used. In our experience, intrareader variability for our professional readers is surprisingly small. They have been remarkably consistent in attaching the same diagnosis to the same film on repeat trials. As yet, this concept has not been tested using lay readers but we have no reason to believe that they will differ.

Bearing in mind that the Federal Coal Mine Health and Safety Act makes available to every coal miner in the U. S. serial chest radiographs at no charge and that there are an estimated 120,000 coal miners, it is apparent that the interpretation of these films is a major undertaking. The magnitude of the problem becomes even greater when it is realized that a series of films on the same man have to be compared in order to assess progression. Thus, we put forward the proposition that a substantial proportion of the chores of routine interpretation for determining progression could be taken over by lay readers. We do not expect this proposal to be popular with all sections of the medical community, in particular those who believe the radiographic categorization of CWP is accurate, objective, and absolute, however, we would call their attention to Alexander Pope's couplet:

> To observations which ourselves we make,
> We grow more partial for th' observer's sake.

ACKNOWLEDGMENTS

We thank Dr. Eugene Pendergrass, Emeritus Professor of Radiology, University of Pennsylvania, and Dr. George Jacobson, Professor and Head of the Department of Radiology, University of Southern California, and Miss Sandra Colebank and Mrs. Mary Dunn for interpretation of films.

REFERENCES

1. FLETCHER, C. M., AND OLDHAM, P. D. (1949). The problem of consistent radiological diagnosis in coalminers' pneumoconiosis. *Brit. J. Ind. Med.* **6**, 168.
2. LIDDELL, F. D. K., AND MAY, J. D. (1966). Assessing the radiological progression of simple pneumoconiosis. Medical Research Memorandum 4. National Coal Board. London.
3. REGER, R. B., AND MORGAN, W. K. C. (1970). On the factors influencing consistency in the radiologic diagnosis of pneumoconiosis. *Amer. Rev. of Resp. Dis.* **102**, 905.
4. ROSSITER, C. E. (1972). Relation between content and composition of coalworkers' lungs and radiological appearances. *Brit. J. Ind. Med.* **29**, 31.
5. SEATON, A., LAPP, N. L., AND MORGAN, W. K. C. (1972). The relationship of pulmonary impairment in simple coal workers' pneumoconiosis to the type of radiographic opacity. *Brit. J. Ind. Med.* **29**, 50.
6. UICC COMMITTEE. (1970). UICC/Cincinnati classification of the radiographic appearances of pneumoconioses. *Chest* **58**, 57.