Alternative Approaches to B Reader Quality Assurance

Commissioned Paper

Ralph T. Shipley, M.D.

The B reader program certifies experts in the interpretation of chest radiographs of the pneumoconioses. The program began in the 1970s as a means to identify physicians to participate in national programs for epidemiological research and for compensation of coalminers and others with disabilities related to dust inhalation (1). It is based on the International Labour Office (ILO) "Classification of Radiographs of the Pneumoconioses," and a "Revised Edition 2000" has recently been published. The goals of the Classification are to standardize classification methods and facilitate international comparisons of data collected on pneumoconiosis for screening and surveillance, epidemiological investigations, and clinical purposes (2).

The ILO Classification provides a standard way of describing and quantifying the changes seen on chest radiographs of workers exposed to dusty environments. It was designed to be used as an epidemiological tool to facilitate international comparability of pneumoconiosis statistics (3), but wider uses have been found for it, including medico-legal and clinical applications.

INITIAL GOALS AND HISTORY

For over 70 years, the ILO has published systems, periodically revised, for the classification of radiographs of the pneumoconioses (4). The early editions were designed to classify changes of silicosis only and used a four-point scale of severity. In 1959, standard films were incorporated. In 1971, the gradation of severity was expanded to a 12-point scale, and provisions were added to include pleural and parenchymal changes related to asbestos inhalation. In 1980, an expanded set of standard films illustrating the types and profusion of small opacities, large opacities, and pleural changes was added. The 2000 revision included additional symbols, simplified the pleural classification, and added another standard film option, the "Quad Set." This consists of 14 films, including nine of the most commonly used standards from the Complete Set, plus five films which are "composite reproductions of quadrant sections from the other radiographs in the Complete Set (2)." The Quad Set is wholly compatible with the Complete Set, and was intended both to reduce the cost of purchase of the standard films and to improve compliance with the requirement for direct comparison of the radiographs being classified to the standard films, by reducing the number of films that must be handled.

B reader certification is accomplished by passing a practical examination which consists of classifying 125 radiographs according to the ILO system. This test is administered by the National Institute for Occupational Safety and Health (NIOSH) in Morgantown, WV and selected other locations. To maintain certification, every four years B readers must pass a recertification test comprising 50 radiographs. The correct answers to the test have been determined by an expert panel of readers (5).

READER VARIABILITY

Inter- and intra-reader variation in radiographic interpretation has been an ongoing issue with the B reader program, despite the introduction of revised classification systems and standard films (6-19). Reader variability occurs throughout the Classification including small and large rounded opacities, but most of the current controversy involves pleural thickening and the reading of small irregular opacities at low profusion levels.

The causes of reader variability are many, and include film quality, reader training and experience, and bias, as well as the variability inherent in the act of individuals interpreting chest radiographs. Variability in classifying radiographs is accentuated by the use of the detailed 12-point scale. Any approach to B reader quality assurance should begin with an analysis of the causes of reader variability.

Film quality is an important component of inter-reader variation (20). Light films promote over-reading, whereas dark films promote under-reading (21). Digital radiography and soft copy interpretation at a workstation will affect this issue, because with this technology it is possible for the reader to control density and contrast independently of the exposure factors used to acquire the image. Classification of digital images, compared to hard copy radiography interpretation, is reviewed elsewhere in these proceedings (see Franzblau paper included in this publication); one preliminary study found no significant differences in the two approaches (22). However, technical defects due to underinflation, mottle, scatter, and positioning are not solved with digital imaging, and proper comparison with the ILO standard films may be logistically difficult unless an electronic edition of the standard films is published.

The training and experience of the reader strongly influences inter- and intra-reader variability. Inexperienced readers tend to over-read when compared to recognized experts (18). Experience with both the wide variation of normal in chest radiography as well as the typical patterns of involvement by pneumoconiosis may reduce the misinterpretation of non-pneumoconiotic opacities as pneumoconiosis. A survey of a group of candidate B readers attending a training course in 1990 indicated that 70% were reading between zero and 10 films for pneumoconiosis per month (1). This finding raised the question of whether some readers are classifying insufficient numbers of films to maintain proficiency.

Conditions other than dust inhalation may be associated with the appearance of small opacities. Male sex, cigarette smoking, obesity, age, underinflation, and other factors can produce the appearance of irregular opacities on chest radiographs, generally at low levels of profusion (23-26). When utilizing the Classification and the standard films in epidemiologic studies, readers are not generally asked to distinguish between the small irregular opacities considered a result of interstitial fibrosis (e.g., asbestosis) and those thought to arise from airway inflammation or other causes. The ILO does state that when the Classification is used for some clinical purposes, the reader may be instructed to "classify only those appearances which the reader believes or suspects to be pneumoconiotic in origin (2)." It is a common clinical exercise when reading chest radiographs to differentiate non-fibrotic "increased markings" from interstitial fibrosis. When this differentiation cannot be made with confidence using the routine radiograph, high resolution computerized tomography (HRCT) may be recommended (see Rose and Lynch paper included in this publication).

This combination of fibrotic and non-fibrotic irregular opacities in one profusion level may contribute to variation in classification. For epidemiologic studies, readers are generally instructed to use the classification as a "pure" tabulation of radiographic appearances and to classify irregular opacities regardless of presumed etiology. In other settings, readers who anticipate that a Classification indicating irregular opacities will be construed as showing parenchymal asbestosis may choose to not classify those opacities which they judge are unlikely to be related to asbestos inhalation. The ILO has recognized in the Revised Edition 2000 that the Classification is used differently in epidemiological than in clinical studies (2). Differences between readers may be amplified by the fact that some observers consider the standard films themselves to be ambiguous in this sphere. Several of the standard films for irregular opacities show changes that are typical of interstitial fibrosis (e.g., 3/3 s/s), whereas others show an increase in irregular opacities that is not particularly typical of interstitial fibrosis (e.g., 1/1 s/t).

One reader might record "increased markings" as "s" or "t" opacities potentially due to pneumoconiosis, whereas another reader might interpret the same opacities as age or smoking related and not classify them. To further add complexity, many believe that occupational dust exposure produces irregular opacities on chest radiographs without interstitial fibrosis, perhaps by causing inflammatory or fibrotic changes in airways.

The argument has been made that reader variability will be minimized if the reader does not interpret the findings but merely records them (27). Consistent with that view, the 1980 edition of the Classification instructed the reader to classify all appearances that "might be due to pneumoconiosis," and the 2000 revision retains that instruction for classifying radiographs in epidemiologic studies (2, 3). However, the 2000 edition of the Classification acknowledges that, in some clinical settings, medical readers may classify only those appearances, which they believe or suspect to be dust-related. To avoid confusion, individuals who utilize and interpret reports of ILO Classifications should be familiar with the published guidelines and the protocol used in generating the specific results.

In the absence of calcification and using standard radiographic approaches, the chest radiograph is neither sensitive nor specific for pleural plaques. Extra-pleural fat deposition and muscle shadows produce thickening of the pleural stripe that can mimic asbestos-related pleural plaques (28), while plaques may be present and not visible on the PA radiograph (10, 29, 30). Oblique views may help increase certainty of plaques and aid discrimination from pleural fat (7), but are not part of the ILO Classification. The Revised Edition 2000 adds a requirement that pleural thickening be at least 3mm to be classified, and there is evidence that this minimum threshold will improve reader agreement (31). Readers who classify an area of pleural thickening because it might be plaque would be expected to disagree with readers who interpret a pleura opacity as fat and therefore do not classify it. Specification of the role of etiologic judgments in reading protocols should assist in the interpretation of situations in which a reader, for example, classifies pleural shadows as changes that "might be pneumoconiosis," but records under comments that the shadows are deemed due to "adipose tissue."

Bias is an important component of inter-reader variability. Knowledge of the exposure history, prevalence of abnormality in a given set of radiographs for classification, employment of the reader by plaintiff or defense counsel, and sympathy for the plight of either the worker or the industry, among other issues, all have the potential to influence a reader's tendency to under- or over-read. The local reading environment, such as location on the east or west coast versus mid-continent has been noted to affect the outcome of Classifications (11).

The inherent qualities of chest radiography and the ILO Classification, and the human element of interpretation make variation in reading inevitable (32). The extensive literature detailing inter-reader variability indicates that not all the disagreement can be from bias related to the implications of a positive or negative reading. The great variation of normal in chest radiographs, the asymmetric distribution of small opacities within lung zones, and the contribution of non-pneumoconiotic factors lead to variation in interpretation. Numerous studies have been published in which experienced readers with no financial incentives have significant disagreements. To improve precision in epidemiological studies using the chest radiograph, it is recommended to employ three readers representative of general reading practices (that is, they should not fall at the extremes of the range of variability between readers) and use the median reading (9, 16). In clinical situations, the radiographic reading should be interpreted in the context of a complete clinical evaluation (33).

STANDARD FILMS

To enhance calibration among readers, the ILO guidelines require a side-by-side comparison to

the standard set of radiographs for each chest image being classified (34). However, the standard film set is expensive and can be cumbersome to handle. As a consequence, compliance with use of the standards during classifications has been inconsistent, although the availability of a lower cost Quad Set of standards, with fewer films, may encourage their use. An additional source of variability derives from the different published versions of the Standard Radiographs, which were produced using differing copy techniques and resulted in dissimilar appearances for the Standard Radiographs from the various sets. As of this writing, the ILO has not provided an approved digitized version of the Standard Radiographs for use with soft copy interpretation, and until such images are available, NIOSH recommends that readers "should continue to use traditional film screen radiographs and standards (35)."

The standards are mid-category examples, so that each exemplifies the center of each major category from 0/0 to 3/3. Unfortunately, the standard films are copies of radiographs, the originals of which used dated techniques. Their quality is variable, and several contain excessive contrast. In addition, the two 0/0 standards are really quite normal, closer to 0/- in the 12-point scheme, in the author's opinion. Thus, the gap between the standard for normal and the standard for mild involvement is greater than it appears, and this leads to greater difficulty in distinguishing 0/1 and 1/0. Boundary standards, or radiographs that illustrate the boundary between categories rather than the mid-categories, may improve reader agreement (36, 37). There is evidence that the mid-category standards bias the reader toward mid-category classification (38).

The pleural standards require revision. The standard for diffuse thickening does not conform to the Revised Edition 2000, which requires blunting of the costophrenic angle. In fact, this standard most likely illustrates pleural fat. The standard image for noncalcified, circumscribed plaque shows an en face plaque, but there are other appearances of plaque that should be illustrated.

FUTURE DIRECTIONS

It is not clinical practice for workers with a question of occupational lung disease to undergo open lung biopsy, so that radiologic-pathologic correlation studies are limited (39). For non-pneumoconiotic interstitial lung disease where biopsy is not possible or appropriate, current clinical practice is to integrate computerized tomography/high-resolution computerized tomography (CT/HRCT) with clinical assessment for diagnosis. Thus, the best "gold standard" readily available for investigation of pneumoconiosis is CT/HRCT.

Because the chest radiograph is inexpensive and easy to acquire, its use in surveys of workers at risk for pneumoconiosis will continue despite its well-documented limitations. Outside the B reader program, it is common experience that the advent of CT scanning has improved our ability to interpret chest radiographs. For these reasons, I propose two areas for inquiry: The categorization of small irregular opacities and the accurate detection of pleural plaques.

While there has been radiographic-HRCT correlation (40, 41), there has been little interest in using HRCT to "go backward" to improve chest radiographic interpretation. We know that "increased markings" with coarsening of bronchovascular shadows can be seen on chest radiographs that does not reflect interstitial fibrosis. On the other hand, interstitial fibrosis usually produces peripheral, subpleural irregular opacities that are confirmed with HRCT. A set of guidelines for distinguishing these opacities on the chest radiograph could be drawn up and tested on a cohort of workers who have both chest x-rays (CXR) and HRCT available for review. The HRCT images could be used as the gold standard to determine the accuracy of the guidelines for interpretation. The inter- and intra-reader variability could be compared to conventional classifications. Non-fibrotic opacities could be studied to identify any correlation with chronic bronchitis or other clinical condition that might not be fibrosis but still be occupationally related.

Similarly, classification of pleural thickening should be studied. A set of guidelines could be drawn up to distinguish pleural fat from pleural plaques on the chest radiograph and tested on a suitable cohort of workers who have both CXR and CT/HRCT available for review.

QUALITY ASSURANCE

The current system of quadrennial recertification is designed to periodically test the individual B reader's adherence to conformity with the level of reading defined by an expert panel. The underlying assumption is that the reader will read similarly in practice as he/she does in the testing situation. However, B readers are influenced by local standards (12) that may not apply to the testing situation, and experience suggests that bias may modify readers' interpretations when not in the testing environment (42).

A 1990 workshop was held to discuss the status of the B reader program (43). Among other ideas, proposals for quality assurance included (1) instituting a mandatory program of checks on readers, (2) initiating a core group of expert readers, (3) making provision for readers to voluntarily calibrate themselves with expert readers.

In order for mandatory checks of readings to identify bias, they would have to be random, not chosen by the individual B reader. Comparing the B reader's reports to the experts' could promote consistent reading patterns between recertifications, but if the B reader chose the radiographs to be monitored, bias could be maintained. The difficulty in mandatory audits is the wide variety of reading activity, which includes clinical, medico-legal, industrial, and governmental settings. While auditing would be easy in a large-scale surveillance project like the NIOSH Coal Workers X-ray Surveillance Project, it would be difficult to achieve in the day-to-day readings done for clinical and legal purposes.

Limiting the total number of B readers or appointing a core group of expert readers would have the effect of containing the variability problem to a smaller number of readers than the current group. The smaller number would ensure that each B reader would maintain a larger experience in interpreting radiographs for pneumoconiosis. The number of current B readers who read a low volume of films would suggest that reducing the total number would not have a detrimental effect on the program. However, the elitist nature of such a proposal, the intrusion upon local practices, and the financial implications will make this proposal unpopular and difficult to implement.

Voluntary calibration of readers has been successfully reported in Canada (44). This was accomplished by periodically circulating batches of radiographs to physicians reading films for pneumoconiosis. Their interpretations were sent to a central location, and feedback from an expert was provided in return. This led to improved agreement over time, indicating a learning effect. The logistical issues of mailing radiographs were recognized and addressed, but the advent of digital radiography and the ability to make images available on-line or on CD-ROM discs may make this approach even practical and economical. CT/HRCT correlation could be included to provide validation of the expert reading. Since the B reader would be well aware of which radiographs were being reviewed, the problem of bias would not be addressed. However, the feedback would be more frequent than recertification every 4 years, perhaps leading to greater uniformity.

SUMMARY

In summary, reader variability is multifactorial. If reducing bias is a priority, random review of B reader interpretations with feedback and a mechanism to enforce compliance with expert reading standards must be developed. Requiring B readers to maintain a minimum volume of reading would eliminate the problem of the outlier, low volume, B reader. Circulating CD-ROM discs or creating a

website with cases available for interpretation with feedback would be an inexpensive way to provide continuing education to B readers between recertification exams. Inter-reader agreement could improve as a result.

Further research into chest radiograph interpretation with HRCT correlation may reduce the variability related to combining fibrotic and non-fibrotic small irregular opacities in the Classification. Differentiating pleural fat from pleural thickening may be improved through further study, or we may determine that, in the absence of diaphragmatic plaques or calcification, the chest radiograph is not effective for this purpose. The pattern of use of the Revised Edition 2000 of the ILO Classification with its acknowledgement of both epidemiological and clinical approaches needs to be clarified.

Reader variability is inherent in chest radiographic interpretation. This is present where no financial incentives are involved, but accentuated when they are. Proposals for further study of methods to improve uniformity of interpretation have been presented. Effective quality assurance designed to eliminate bias will be difficult to implement. Assuring quality through innovative new methods and inquiry into improved reading practices through CT/HRCT correlation will help maintain and improve the stature of the B reader program.

REFERENCES

1. Wagner GR, Attfield MD, Kennedy RD, et al. The NIOSH B Reader Certification Program: An Update Report. *J Occup Med* 1992; 34:879-884.

2. International Labour Office: Guidelines for the Use of the ILO Classification of Radiographs of the Pneumoconioses, 2000 ed., Geneva, International Labour Office, 2002.

3. International Labour Office: Guidelines for the Use of ILO International Classification of Radiographs of Pneumoconioses, 1980 ed. Geneva, International Labour Office, 1980.

4. Shipley RT. The 1980 ILO Classification of Radiographs of the Pneumoconioses. *Radiol Clin North Am* 1990; 30:1135-1145.

5. Morgan RH. Proficiency Examination of Physicians for Classifying Pneumoconiosis Chest Films. *AJR* 1979;132: 803-808.

6. Berge SR, Skyberg K. Radiographic Evidence of Pulmonary Fibrosis and Possible Etiologic Factors at a Nickel Refinery in Norway. *J Environ Monit* 2003; 5:681.

7. Lawson CC, LeMasters MK, LeMasters GK, et al. Reliability and Validity of Chest Radiograph Surveillance Programs. *Chest* 2001; 120:64.

8. Welch LS, Hunting KL, Balmes J, et al. Variability in the Classification of Radiographs Using the 1980 International Labor Organization Classification for Pneumoconioses. *Chest* 1998; 114:1740.

9. Mulloy KB, Coultas DB, Samet JM. Use of Chest Radiographs in Epidemiologic Investigations of Pneumoconioses. *Br J Ind Med* 1993; 50:273.

10. Frumkin H, Pransky G, Cosmatos I. Radiologic Detection of Pleural Thickening. *Am Rev Respir Dis* 1990; 142:1325.

11. Parker DL, Bender AP, Hankinson S. Public Health Implications of the Variability in the Interpretation of "B" Readings for Pleural Changes. *J Occup Med* 1989; 31:775.

12. Ducatman AM, Yang WN, Forman SA. 'B Readers' and Asbestos Medical Surveillance. *J Occup Med* 1988; 30:644.

13. Bourbeau J, Ernst P. Between- and Within-Reader Variability in the Assessment of Pleural Abnormality Using the ILO 1980 International Classification of Pneumoconioses. *Am J Ind Med* 1988; 14:537.

14. Weill H. Diagnosis of Asbestos-Related Disease. *Chest* 1987; 91:802.

15. Rossiter CE, Browne K, Gilson JC. International Classification Trial of AIA set of 100 Radiographs of Asbestos Workers. *Br J Ind Med* 1988; 45:538.

16. Weill H, Jones R. The Chest Roentgenogram as an Epidemiologic Tool. *Arch Environ Health* 1975; 30:435.

17. Amandus HE, Pendergrass EP, Dennis JM, et al. Pneumoconiosis: Inter-Reader Variability in the Classification of the Type of Small Opacities in the Chest Roentgenogram. *Radiology* 1974; 122:740.

18. Felson B, Morgan WKC, Bristol LJ, et al. Observations on the Results of Multiple Readings of Chest Films in Coal Miners' Pneumoconiosis. *Radiology* 1973; 109:19.

19. Reger RB, Morgan WKC. On the Factors Influencing Consistency in the Radiologic Diagnosis of Pneumoconiosis. *Am Rev Respir Dis* 1970; 102:905.

20. Musch DC, Higgins ITT, Landis JR. Some Factors Influencing Interobserver Variation in Classifying Simple Pneumoconiosis. *Br J Ind Med* 1985; 42:346-349.

21. Reger RB, Smith CA, Kibelstis JA, et al. The Effect of Film Quality and other Factors on the Roentgenographic Categorization of Coal Workers' Pneumoconiosis. *Radiology* 1972; 225:462.

22. Mannino DM, Kennedy RD, Hodous TK. Pneumoconiosis: Comparison of Digitized and Conventional Radiographs. *Radiology* 1993; 187:791.

23. Meyer JD, Islam SS, Ducatman AM, et al. Prevalence of Small Lung Opacities in Populations Unexposed to Dusts. *Chest* 1997; 111:404.

24. Dick JA, Morgan WKC, Muir DFC, et al. The Significance of Irregular Opacities on the Chest Roentgenogram. *Chest* 1992; 102:251.

25. Epstein DM, Miller WT, Bresnitz EA, et al. Application of ILO Classification to a population without Industrial Exposure: Findings to be differentiated from Pneumoconiosis. *AJR* 1984; 142:53.

26. Zitting AJ. Prevalence of Radiographic Small Lung Opacities and Pleural Abnormalities in a Representative Adult Population Sample. *Chest* 1995; 107:126-131.

27. Jacobsen M. The International Labour Office Classification: Use and Misuse. *Ann NY Acad Sci* 1991; 643:100-107.

28. Friedman AC, Fiel SB, Fisher MS, et al. Asbestos-Related Pleural Disease and Asbestosis: A Comparison of CT and Chest Radiography. *AJR* 1988; 150:269-275.

29. Schwartz DA. New Developments in Asbestos-Induced Disease. *Chest* 1991; 99:191-198.

30. Sargent EN, Boswell WD, Ralls PW, et al. Subpleural Fat Pads in Patients exposed to Asbestos: Distinction from Non-calcified Pleural Plaques. *Radiology* 1984; 152:273-277.

31. Greene R, Boggis C, Jantsch H. Asbestos-Related Pleural Thickening: Effect of Threshold Criteria on Interpretation. *Radiology* 1984; 152:560-573.

32. Henry DA. International Labor Office Classification System in the Age of Imaging: Relevant or Redundant. *J Thorac Imaging* 2002; 17:179.

33. Ross RM. The Clinical Diagnosis of Asbestosis in This Century Requires More Than a Chest Radiograph. *Chest* 2003; 124:1120-1128.

34. International Classification of Radiographs of the Pneumoconioses, Revised Edition 2000, the COMPLETE SET of Radiographs. Available at: http://www-ilo-mirror.cornell.edu/public/english/support/publ/xtextoh.htm#b1774

35. To B or not to B a NIOSH B Reader. Available at: http://www.cdc.gov/niosh/pamphlet.html

36. Hodous TK, Lu X, Denton RA. ILO Classification of the Standard Chest Films of the 1986 Chinese Roentgenodiagnostic Criteria of the Pneumoconioses. *In:* Proceedings of the VIIth International Pneumoconioses Conference, Pittsburgh, 1988, p 487.

37. David A. Present Use and Trends in the Development of the ILO Classification of Radiographs of the Pneumoconioses. *In:* Proceedings of the VIIth International Pneumoconioses Conference, Pittsburgh, 1988, p 479.

38. Hessel PA. The Middling Tendency in Reading Chest Films for Pneumoconiosis: An Important Source of Bias and Variability. *Am J Ind Med* 1985; 8:43.

39. Gamsu G, Salmon CJ, Warnock ML, et al. CT Quantification of Interstitial Fibrosis in Patients with Asbestosis: A Comparison of Two Methods. *AJR* 1995; 164:63-68.

40. Harkin TJ, McGuinness G, Goldring R, et al. Differentiation of the ILO Boundary Chest Roentgenograph (0/1 to 1/0) in Asbestosis by High-Resolution Computed Tomography Scan, Alveolitis, and Respiratory Impairment. *JOEM* 1996; 38:46-52.

41. Aberle DR, Gamsu G, Ray CS. High-Resolution CT of Benign Asbestos-Related Diseases: Clinical and Radiographic Correlation. *AJR* 1988; 151:883-891.

42. Ducatman AM. Variability in Interpretation of Radiographs for Asbestos Abnormalities: Problems and Solutions. *Ann NY Acad Sci* 1991; 643:108-120.

43. Attfield MD, Wagner, GR. A Report on a Workshop on the National Institute for Occupational Safety and Health B Reader Certification Program. *J Occup Med* 1992; 34:8875.

44. Muir DC, Julian JA, Roos JO, et al. Classification of Radiographs for Pneumoconiosis: The Canadian Pneumoconiosis Reading Panel. *Am J Ind Med* 1993; 24:139.

The NIOSH B Reader Certification Program:

Looking into the Future

Department of Health and Human Services
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health

CDC Workplace Safety and Health

National Institute for Occupational Safety and Health NIOSH

Disclaimer

Ordering Information

To receive documents or other information about occupational safety and health topics, contact NIOSH at

1-800-CDC-INFO (1-800-232-4636)
TTY: 1-888-232-6348
E-mail: cdcinfo@cdc.gov
or visit the NIOSH Web site at
www.cdc.gov/niosh

For a monthly update on news at NIOSH,
subscribe to NIOSH eNews by visiting
www.cdc.gov/niosh/eNews

DHHS (NIOSH) Publication No. 2009-140