EDITORIALS

To Be (B Read) or Not to Be (B Read), That Is the Question: Conflict of Interest and Radiographic Interpretation

Carrie A. Redlich, M.D., M.P.H.[1], and Kenneth D. Rosenman, M.D.[2]

[1]Yale University School of Medicine, New Haven, Connecticut; and [2]Michigan State University, East Lansing, Michigan

In this issue of *AnnalsATS*, Friedman and colleagues (pp. 1634–1641) address whether physician source of payment is associated with classification of radiographs for coal workers' pneumoconiosis (1). Although most pulmonologists are not directly involved in reading chest radiographs of workers at risk for pneumoconiosis, the study raises important issues relevant to broader clinical practice—the reliability of chest radiograph interpretations and the extent to which physician decisions are influenced by potential financial conflicts of interest (COIs); in this situation, the question of who is paying for the chest radiograph interpretation is raised.

Although clinicians are aware that mistakes can be made interpreting chest radiographs, such as missing a lung mass or other abnormalities, such oversights are usually attributed to human error. There is less awareness of the substantial variation in interpretations between different physicians (interreader) and variation in the same physicians over time (intrareader). For example, among nine radiologists participating in the National Lung Screening Trial, the multirater κ statistic for interreader agreement on the presence of at least one

noncalcified lung nodule was 0.38, whereas pairwise κ values between individual radiologists ranged from 0.13 to 0.60 (mean, 0.38) (2). Depending on the guidelines used to interpret κ values, the multirater κ would be characterized as fair (3), poor (4), or minimal (5), whereas the pairwise κ characterization would range from slight/poor/none to substantial/fair to good/moderate (3–5).

Most pulmonologists may not be familiar with the International Labor Organization system to classify chest radiographs for pneumoconiosis (6). In 1974, to reduce variability in classifications, the National Institute for Occupational Safety and Health developed a B reader certification program for physicians that requires passing a training course and a rigorous examination based on reading 125 radiographs with a recertification exam required every 5 years (6). The International Labor Organization classification system is required in the National Institute for Occupational Safety and Health Coal Workers' Surveillance Program as well as the U.S. Department of Labor (USDOL) Black Lung Benefits Program, and the Occupational Safety and Health Administration required medical surveillance for workers exposed to asbestos and silica. It is also widely used in other surveillance programs such as the U.S. Navy Asbestos Medical Surveillance Program and the Department of Energy Building Trades Medical Screening Program as well as in epidemiological studies, workers' compensation cases, and third-party asbestos and silica compensation suits.

Studies that have assessed the reliability of B reader interpretations of film and digital chest radiographs have found that interreader and intrareader κ values among B readers have ranged from 0.54 to 0.65 and 0.65 to 0.77 (7–14), generally considered good agreement. In addition, these studies have shown that major discrepancies in B reader interpretations were uncommon. Despite these findings, there have been concerns, based largely on media reports and anecdotal cases,

that in practice B reader interpretations can vary substantially and may be impacted by financial considerations.

The paper by Friedman and colleagues is the first well-designed study to address the COI issue in the classification of radiographs for pneumoconiosis (1). Previously, there have been newspaper reports and a single prior study, performed at the request of defendant attorneys, that involved six B readers who reinterpreted 551 chest radiographs previously read by plaintiff attorneys as showing asbestos-related changes. The authors found little evidence of asbestos-related changes and concluded that the earlier B reader interpretations done by physicians for the plaintiff attorneys were inaccurate (15). Methodological problems with this study of potential asbestos-related changes did not allow one to conclude the magnitude of the difference, but clearly there were differences in the interpretations between the B readers who had classified radiographs at the request of the attorneys of the plaintiffs and those who had classified radiographs at the request of the attorneys of the defendants.

By contrast, the extensive study by Friedman and colleagues included a much larger number of individuals (37,530), radiographs (63,780), and physician B readers (264) and a more objective study design. The data allowed for comparison between three groups of B reader physicians: those who performed B reader classifications predominately for employers, those who did so predominately for coal miners, and those predominantly hired by the USDOL, as determined by an independent review of records that referred to the physicians. Physician specialty (predominantly radiologists, pulmonologists, and internists) and years in practice were similar among the three groups. Physician B readers hired by an employer in the majority of cases classified 92.6% of the radiographs (17,048 of 18,403) as negative for pneumoconiosis compared with 24.8% (1,558 of 6,284) among those hired by a

miner in the majority of cases and 58.5% (16,822 of 28,753) for those hired by the USDOL in the majority of cases.

The authors documented a strong association between source of payment and the classification for pneumoconiosis. The odds of finding no pneumoconiosis increased substantially (adjusted odds ratio [OR], 1.46; confidence interval [CI], 1.44–1.47) per 10% increase in the proportion of times a physician was hired by the employer. Similarly, per 10% increase in times hired by the miner, the odds increased for classifying simple pneumoconiosis (adjusted OR, 1.51; CI, 1.49–1.52) and progressive massive fibrosis (adjusted OR, 1.28; CI, 1.26–1.30). What was particularly disturbing was that on almost 4,000 chest radiograph classifications, the B readers disagreed as to whether the radiograph was negative for pneumoconiosis versus showed findings of advanced pneumoconiosis (advanced simple pneumoconiosis or progressive massive fibrosis), much greater differences than would be expected from the literature cited above on interreader variability.

It should be noted that the study demonstrates association, not causation. It is possible that attorneys for coal miners and employers may recontact and reuse B readers who are more likely to read a radiograph in the interests of their client rather than being influenced by the source of payment. Nevertheless, the marked variation in classification, particularly of the almost 4,000 radiographs noted above, and the strong association with potential COIs speak to the need to eliminate COI and reduce interreader variability in the Black Lung Benefits program.

Variability in the classification of radiographs for pneumoconiosis raises concerns related not only to the Black Lung Benefits program but also to the many other settings where B readings are performed, especially as the findings are commonly used to identify work-related lung disease and/or determine benefits. As noted by the authors, the black lung program is a "*microcosm of the larger workers' compensation system*" where similar COI concerns exist. However, data in the United States on potential financial COIs in the workers' compensation system are lacking and any evaluation would need to be performed separately on each of the unique state-based systems,

Friedman and colleagues discuss a number of potential changes to reduce potential COIs, from who pays and how much is paid for B reader classification, having a B reader panel, and decertification of B readers who provide "unreasonably inaccurate classifications" (1). Having more pulmonologists become certified B readers, so as to expand the limited pool of B readers in the United States (currently 207) (16), may also be useful.

In addition to B reader settings, the findings are relevant to broader clinical practice and the reliability of radiographic interpretations. Ongoing work on artificial intelligence technology, specifically deep convolutional neural network methods, has been shown to achieve human-level performance for certain lung disease diagnoses and has emerged as a possible new solution to address both interreader and intrareader variability (17) and could potentially be used to perform B reader interpretations in the future. ∎

References

1 Friedman LS, De S, Almberg KS, Cohen RA. Association between financial conflicts of interest and International Labor Office classifications for black lung disease. *Ann Am Thorac Soc* 2021;18:1634–1641.

2 Singh SP, Gierada DS, Pinsky P, Sanders C, Fineberg N, Sun Y, *et al.* Reader variability in identifying pulmonary nodules on chest radiographs from the national lung screening trial. *J Thorac Imaging* 2012;27:249–254.

3 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.

4 Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley; 1981.

5 McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–282.

6 Halldin CN, Hale JM, Weissman DN, Attfield MD, Parker JE, Petsonk EL, *et al.* The National Institute for Occupational Safety and Health B Reader Certification Program-An update report (1987 to 2018) and future directions. *J Occup Environ Med* 2019;61:1045–1051.

7 Sen A, Lee SY, Gillespie BW, Kazerooni EA, Goodsitt MM, Rosenman KD, *et al.* Comparing film and digital radiographs for reliability of pneumoconiosis classifications: A modeling approach. *Acad Radiol* 2010;17:511–519.

8 Impivaara O, Zitting AJ, Kuusela T, Alanen E, Karjalainen A. Observer variation in classifying chest radiographs for small lung opacities and pleural abnormalities in a population sample. *Am J Ind Med* 1998;34:261–265.

9 Lawson CC, LeMasters MK, Kawas Lemasters G, Simpson Reutman S, Rice CH, Lockey JE. Reliability and validity of chest radiograph surveillance programs. *Chest* 2001;120:64–68.

10 Musch DC, Landis JR, Higgins ITT, Gilson JC, Jones RN. An application of kappa-type analyses to interobserver variation in classifying chest radiographs for pneumoconiosis. *Stat Med* 1984;3:73–83.

11 Musch DC, Higgins ITT, Landis JR. Some factors influencing interobserver variation in classifying simple pneumoconiosis. *Br J Ind Med* 1985;42: 346–349.

12 Naidoo RN, Robins TG, Solomon A, White N, Franzblau A. Radiographic outcomes among South African coal miners. *Int Arch Occup Environ Health* 2004;77:471–481.

13 Welch LS, Hunting KL, Balmes J, Bresnitz EA, Guidotti TL, Lockey JE, *et al.* Variability in the classification of radiographs using the 1980 International Labor Organization Classification for Pneumoconioses. *Chest* 1998;114:1740–1748.

14 Muller JG, Lieske JM, Hernandez JE, Kubiak G, Rudolph WG, Jan MH. Variability in interpretation among B-readers in the U.S. Navy Asbestos Medical Surveillance Program. *Mil Med* 2008;173:375–380.

15 Gitlin JN, Cook LL, Linton OW, Garrett-Mayer E. Comparison of "B" readers' interpretations of chest radiographs for asbestos related changes. *Acad Radiol* 2004;11:843–856.

16 Martin M, Cohen B, Weissman D, Halldin C, Storey E, Wolfe A. Become a NIOSH-certified B reader. Morgantown, WV: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH); 2018 [accessed 2021 Aug 4]. Available from: https://doi.org/10.26616/NIOSHPUB2019111.

17 Pham HH, Le TT, Tran DQ, Ngo DT, Nguyen HQ. Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing* 2021;437:186–194.