



Intramodality and Intermodality Comparisons of Storage Phosphor Computed Radiography and Conventional Film-Screen Radiography in the Recognition of Small Pneumoconiotic Opacities

A. Scott Laney, PhD; Edward L. Petsonk, MD, FCCP; and Michael D. Attfield, PhD

Background: Digital radiography systems are replacing traditional film for chest radiographic monitoring in the recognition of pneumoconiosis.

Methods: To further investigate previous findings regarding the equivalence of film-screen radiographs (FSRs) and storage phosphor computed radiographs (CRs), FSRs and CRs from 172 underground coal miners were classified independently by seven National Institute for Occupational Safety and Health-approved B readers, using the International Labor Office (ILO) classification of radiographs of pneumoconiosis.

Results: More CRs were classified as “good” quality compared with FSRs (prevalence ratio [PR], 1.5; 95% CI, 1.4-1.6; P , .001). B readers showed good overall agreement on scoring small opacity profusion using CRs vs FSRs (weighted κ , 0.58; 95% CI, 0.54-0.62). Significantly more irregular opacities (compared with rounded) were classified using CR images compared with FSR (PR, 1.3; 95% CI, 1.1-1.6; P = .01). Similarly, the smallest sized opacities (width < 1.5 mm, p and s type) were reported more frequently using CR vs FSR images (PR, 1.3; 95% CI, 1.1-1.5; P < .001). Interreader and intrareader agreement was lower with respect to the classification of shape and size than for small opacity profusion. Overall, interreader and intrareader variability did not differ significantly using CR vs FSR.

Conclusions: Under optimal conditions, using standardized methods and equipment, reader visualization of small pneumoconiotic opacities does not appear to differ meaningfully, whether using CR or FSR. Variability in ILO classifications between imaging modalities appears to be considerably lower than variability among readers. The well-documented challenge of reader variability does not appear to be resolved through the use of digital imaging alone, and additional approaches must be evaluated.

CHEST 2011; 140(6):1574-1580

Abbreviations: CR = storage phosphor computed radiograph; FSR = film-screen radiograph; ILO = International Labor Office; NIOSH = National Institute for Occupational Safety and Health; PR = prevalence ratio

Occupational lung diseases remain important causes of morbidity and mortality worldwide.¹⁻⁴ Health monitoring using chest radiographs is one of the recommended measures for the protection of workers in dusty trades.⁵ The International Labor Office (ILO) classification system provides a standardized approach for the classification of abnormalities associated with pneumoconiosis and related diseases on chest radiographic films.⁶ In many countries, conversion from the use of film to digital chest radiographic imaging is well underway. Health professionals face a number of technical and logistic challenges in the application of digital

chest radiographs for the recognition and classification of pneumoconioses, including appropriate and standardized means of image acquisition and viewing. The ILO has recently accepted a digitized version of the set of standard images used in classifying digital radiographs displayed on medical grade monitors (I. Fedotov, MD, PhD, personal communication, August 2010). However, to be useful in worker health protection programs, the implemented digital systems must be at least as effective as traditional film-screen radiography for the recognition and classification of pneumoconiotic opacities.⁷⁻¹⁰

The objective of this study was to further investigate the equivalence of film-screen and storage phosphor computed radiography in the recognition and classification of pneumoconiosis. Specifically, we explored several previously observed intermodality differences (with respect to the designation of small opacity shape and size, and the proportion of miners demonstrating high opacity profusion) and extended earlier comparisons of reader variability in the classification of pneumoconiosis between the two radiographic modalities.

MATERIALS AND METHODS

The data for this study were derived from the National Institute for Occupational Safety and Health (NIOSH)-administered Enhanced Coal Workers' Health Surveillance Program.¹¹ The current investigation explores several findings from our previous study of 1,401 male underground coal miners.⁷ In that study, film-screen radiographs (FSRs) and storage phosphor computed radiographs (CRs) were acquired on the same day, and each image was later interpreted by two B readers, for a total of four classifications for each miner (two FSRs and two CRs). For the present analysis, we selected chest images from 172 of the 1,401 miners from the original study for further investigation. Images were selected for the present study based on the B reader classifications from the earlier study. Images were included in this study if one or more of the four original classifications indicated small opacity profusion > 0/0. The NIOSH Institutional Review Board approved the project, and written informed consent was obtained from each study participant (HSRB-06-DRDS-01). Information on the methods and procedures used in the Enhanced Coal Workers' Health Surveillance Program, including image acquisition and processing and participant recruitment, has been published previously.⁷

Image Classification

For this study, FSR and CR images from 172 miner participants were classified independently by seven B readers. Each reader interpreted both of the paired FSR and CR images from an individual miner, but the images were presented at separate reading sessions, and readers were blinded to the results of their own or other readers' previous interpretations. This yielded a total of 344 classifications from each of the seven readers (172 FSRs and 172 CRs) for a total of 2,408 total readings (1,204 for each modality). For intermodality comparisons, complete information for small opacity profusion was required from all seven readers in both modalities. Because of missing values due to lack of recording or because

readers defined a film as unreadable, small opacity profusion was available for 2,370 of the 2,408 total readings (1,185 CR and 1,185 FSR matched pairs). The methods and technical specifications for displaying images have been described previously.⁷ In addition, each of the readers in this study had participated in the previous study of 1,401 miners and had provided FSR and CR classifications for some of the images included in this study. Thus, for each of the 172 image pairs, classifications were available from the earlier study by either another or the same reader. To enable the assessment of within-reader variability by modality, results from the previous classifications by the same reader were used. For example, during the previous study, reader 1 had classified 49 of the 172 image pairs selected for this analysis. This yielded four classifications by reader 1 available for this study from that subset of 49; the first classification using FSR (FSR₁), the first CR classification (CR₁), the second FSR classification (FSR₂), and the second CR classification (CR₂).

Data Analysis

κ Statistics were used to examine intermodality and intra-modality agreement. For contingency tables larger than 2 × 2, Cicchetti-Allison-weighted κ values were used.¹² The overall weighted κ value was obtained by the aggregation of the small opacity profusion classifications across all readers for the 2,370 eligible images. *P* values presented are Mantel-Haenszel χ^2 . The SAS statistical software package, version 9.1 (SAS Institute; Cary, North Carolina) was used for all analyses.

RESULTS

Information on image quality is presented in Table 1 by imaging modality. More images were classified as "good" (ILO technical quality category 1) by CR compared with FSR (prevalence ratio [PR], 1.5; 95% CI, 1.4-1.6). The corresponding κ value (κ) was 0.4 (95% CI, 0.35-0.45). Additionally, significantly more images were classified as good or acceptable with no major defects (ILO technical quality categories 1 and 2) by CR compared with FSR (PR, 1.2; 95% CI, 1.1-1.2; and $\kappa = 0.23$; 95% CI, 0.17-0.29). The low κ values were largely driven by the higher proportion of films classified as category 3 (some defects) by FSR. Fifteen FSR images were not classified because of unacceptable film quality. For 14 of these 15, the classifications of the miners' CR images showed small opacity profusion of 0/0 (one was classified as 0/1).

Profusion of Small Opacities by Modality

The distribution of the profusion of small opacities for the 2,370 readings with complete information for small opacity profusion is shown by modality in Figure 1. The majority of readings showed profusion category 0/0 by FSR and CR.

More radiographs showing small opacities ($\geq 0/1$) were reported for FSR compared with CR (37.2% compared with 31.0%; PR, 1.2; 95% CI, 1.1-1.3; *P* = .001). In addition, intermodality agreement was moderate to good with respect to 0/0 classifications with $\kappa = 0.59$ (95% CI, 0.54-0.64). The higher percentage

Manuscript received March 16, 2011; revision accepted May 2, 2011.

Affiliations: From the Surveillance Branch, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Morgantown, WV.

Funding/Support: The authors have reported to *CHEST* that no funding was received for this study.

Correspondence to: A. Scott Laney, PhD, Surveillance Branch, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, 1095 Willowdale Rd, Mail Stop HG900.2, Morgantown, WV 26505-2888; e-mail: alaney@cdc.gov

© 2011 American College of Chest Physicians. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians (<http://www.chestpubs.org/site/misc/reprints.xhtml>).

DOI: 10.1378/chest.11-0629

Table 1—Image Technical Quality for 1,204 Classifications by Imaging Modality

Quality	Film-Screen (FSR)	Digital (CR)
1: Good	536 (45.5)	792 (65.8)
2: Acceptable, no defects	410 (34.1)	332 (27.6)
3: Acceptable, some defects	228 (18.9)	74 (6.1)
4: Unacceptable	15 (1.2)	0 (0)
–: Missing/no indication	15 (1.2)	6 (0.5)

Data are presented as No. (%) and are 172 images classified by seven B readers, for a total of 1,204 classifications for FSR and CR each. CR = storage phosphor computed radiograph; FSR = film-screen radiograph.

of abnormality for FSR was concentrated mostly in the borderline and low categories (0/1, 1/0, and 1/1; 30.5% FSR vs 24.6% CR). In contrast, the percentage of radiographs classified as $\geq 1/2$ by modality was very similar (6.7% for FSR compared with 6.4% for CR; PR, 1.0; 95% CI, 0.8-1.4; $P = .80$) and exhibited a high level of agreement ($\kappa = 0.65$ [95% CI, 0.56-0.74]).

Taken together, when all readings were aggregated, the overall weighted κ was 0.58 (95% CI, 0.54-0.62), denoting moderate to good intermodality agreement. The distribution of the profusion of small opacities by modality is presented for each reader in Figure 2.

Small Opacity Shape and Size by Modality

The frequencies and percentages of the designations of primary small opacity shape and size are presented in Table 2. Both shape and size designations, when each was treated dichotomously, differed by modality. Significantly more irregular opacities (compared with rounded) were classified using CR images compared with FSR (PR, 1.3; 95% CI, 1.1-1.6; $P = .01$). Similarly,

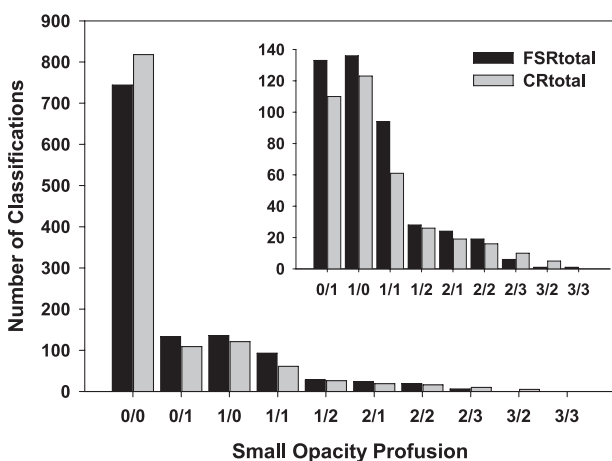


FIGURE 1. Profusion of small opacities for 1,204 classifications by imaging modality. One hundred seventy-two images classified by seven B readers, for a total of 1,204 classifications for FSR and CR each. Inset panel is data presented rescaled for viewing purposes. CR = storage phosphor computed radiograph; FSR = film-screen radiograph.

the smallest sized opacities (width < 1.5 mm, p and s type) were reported more frequently using CR vs FSR images (PR, 1.3; 95% CI, 1.1-1.5; $P < .001$). At the individual level, readers tended to be similar in their overall between-modality results, with a slight tendency to classify more small and irregular opacities using CR images compared with FSR (see also Figure 2, primary shape and size column). However, both interreader and intrareader variability was high with respect to the classification of shape and size when assessed at the six category levels (two shape \times three size categories), as demonstrated by the low levels of agreement presented in Table 3.

Within-Reader Variability

Intermodality agreement differed substantially from reader to reader, ranging from $\kappa = 0.39$ to $\kappa = 0.72$ (Fig 2). However, in the analysis of within-reader variability that included readings from the earlier study (described in the “Materials and Methods” section), readers by and large maintained a consistency with respect to self-agreement, irrespective of modality or reading session. For each reader, the level of intramodality agreement using FSR tended to be similar to that using CR, and similar to intermodality agreement (Fig 3). The deviations from the within-reader mean of the six κ values were relatively small. In only two of the 42 κ values presented in Figure 3 did the 95% CIs not intersect the mean value; one represented an intramodality comparison (reader 3, FSR₁ vs FSR₂), whereas one represented an intermodality comparison (reader 2, FSR₁ vs CR₁).

DISCUSSION

This report extends our previous comparative studies of traditional film-screen and digital radiography in the recognition of small opacities of pneumoconiosis. These data clearly demonstrate that classification of radiographs itself has far greater inherent variability than differences due to imaging modality. However, this observation is not entirely unexpected because interreader differences in the classification of radiographs has long been known to be an inherent source of variation.¹³⁻¹⁷ Although we were unable to identify major intermodality differences, we did observe some modest differences between CR and FSR with respect to (1) image quality, (2) the proportion of radiographs classified as 0/0, and (3) the classification of opacity size and shape.

The largest difference observed between modalities was in the rating of image quality. Across all images, readers clearly ranked image quality higher for CR than for film (PR, 1.5) and this is consistent with previous

Table 2—Shape and Size of Opacities

Shape and Size	Film-Screen	Digital
Rounded	328 (74.4)	245 (66.4)
p	108 (24.5)	107 (29.0)
q	160 (36.3)	102 (27.6)
r	60 (13.6)	36 (9.8)
Irregular	113 (25.6)	124 (33.6)
s	72 (16.3)	88 (23.9)
t	36 (8.2)	34 (9.1)
u	5 (1.1)	2 (0.5)

Data are presented as No. (%) and are 172 images classified by seven B readers, for a total of 1,204 classifications for FSR and CR each. Primary shape and size designations are standard International Labor Office designations. Readers did not record shape and size for images with 0/0 profusion.

quality over time could potentially address that specific issue, but are outside the scope of the present study.

The second difference between modalities that we observed was in the proportion of 0/0 classifications: more CR images than FSR images were classified as 0/0. This, too, was consistent with previous findings. However, the actual magnitude of the effect was quite small (PR, 1.1; 95% CI, 1.0-1.2) and logistic regression models (data not shown) indicated that the difference was entirely accounted for by readers 2 and 5. The modality differences in small opacity profusion were confined to small opacity profusion categories 1/1 and below. It may be that the superior CR image quality allows for greater confidence in definitively ruling out the presence of scant small opacities. However, because the magnitude of this association is borderline and driven by a minority of readers, any conclusions drawn from this observation should be weighted accordingly.

A third difference we observed between modalities was the greater proportion of small and irregularly shaped opacities shown by CR compared with FSR images. CR appears to lead to greater reporting of opacities that are smaller and more reticular, compared with FSR. If these results are confirmed, they have the potential to influence the selection of standard images

Table 3—Intrareader Simple κ Values and 95% CIs Comparing ILO Primary Shape and Size Designation

	κ Value	95% CI
Reader 1	0.30	0.12-0.48
Reader 2	0.19	0.03-0.36
Reader 3	0.48	0.14-0.64
Reader 4	0.39	0.14-0.64
Reader 5	0.18	-0.09-0.44
Reader 6	0.46	0.26-0.66
Reader 7	0.48	0.16-0.81

κ Values represent six-category agreement level of shape and size classification (p, q, r, s, t, u) between film-screen and digital modalities by reader for each image. ILO = International Labor Office.

for future revisions of the ILO classification system that will incorporate digital radiography. Most studies using the ILO classification system have focused on three outcomes: the profusion of small opacities, the presence of large opacities, and the presence of pleural abnormalities. To our knowledge, no studies have reported intrareader and interreader variation in small opacity shape and size designations. In this study, shape and size designation differed substantially by reader, with much lower levels of interreader agreement than for profusion. This poses a challenge in drawing conclusions regarding observed differences between modalities in opacity shape and size. However, the increased recognition of the smallest opacities (both rounded and irregular) when viewing CR images compared with FSR was a consistent finding across all readers, which reinforces it as a true effect.

Despite considerable effort to uncover differences between FSR and CR, we were unable to identify any important differences between these imaging modalities when they were used in the classification of the pneumoconioses. Readers rated CR image quality higher than that of FSR, consistent with previous studies, which should further support the adoption of digital radiographs for this purpose.^{7,9,10,18,19} However, use of digital images does not appear to reduce the well-documented challenge of variability between readers. The results of this study plainly demonstrate that reader variability is far greater than the variability that can be attributed to imaging modality. Reader perception and observation is a dynamic process, and variability both within and between readers in the classification of radiographs of pneumoconiosis using the ILO system has long been recognized.

A variety of methods have been used to arrive at a “true” classification while minimizing reader variability. Some of these methods include the following: (1) requiring readers to demonstrate proficiency in the interpretation of radiographs of pneumoconioses (NIOSH B reader examinations), (2) using an aggregate measure from multiple readers and readings per radiograph as a “final determination,” (3) requiring readers to use standardized methods and practices, and (4) implementing quality control programs with feedback to the readers.²⁰

For the current study, seven experienced B readers interpreted each of the study radiographs. When combined as a group, the global weighted κ value across all readers for the full 12-category small opacity profusion score was 0.58, indicating good intermodality agreement (n = 2,370). In addition, the overall distribution of small opacities by modality was similar when taken as a whole.

By using previous classifications from an earlier study, we were able to assess intrareader variability by

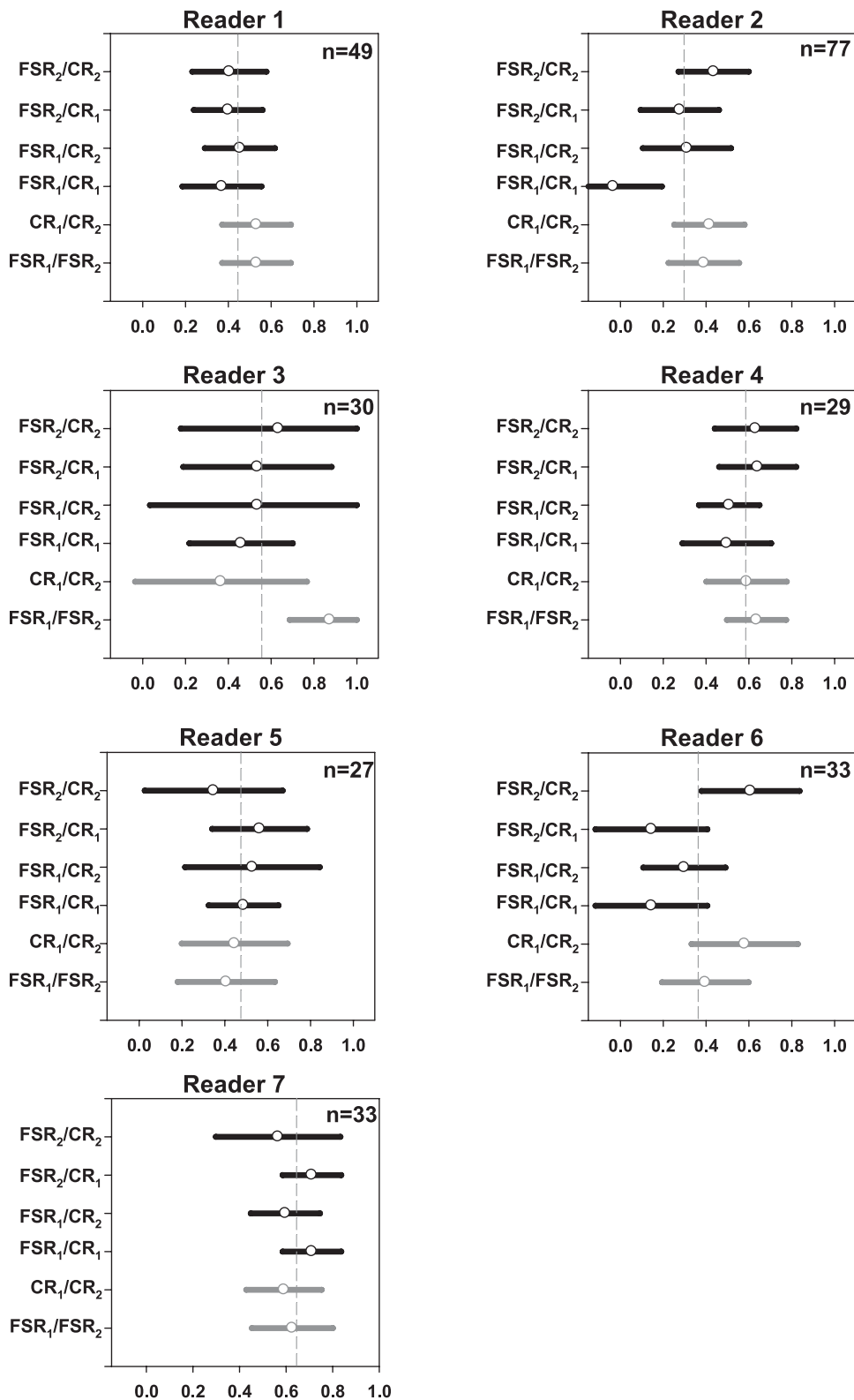


FIGURE 3. Intrareader weighted κ values and 95% CIs comparing small opacity profusion between imaging modalities. Each image was interpreted by the same reader twice by each imaging modality. FSR₁ represents first film-screen reading, FSR₂ the second film-screen reading, CR₁ the first digital reading, and CR₂ the second digital reading for each image. Gray values are within modality κ values. Gray dashed line indicates the mean of κ values for each reader. The total number of image pairs with four classifications is presented in the upper right quadrant for each reader.

modality. Unfortunately, use of digital images did not appear to improve reader variability. The ongoing challenge of reader variability stresses the importance of integrating quality control into all pneumoconiosis reading programs. Recommended approaches include presenting readers, in a blinded fashion, with well-characterized images, and providing feedback on their performance in relation to the quality control films and to other readers.²⁰

CONCLUSIONS

Overall, the current study highlights two important findings. First, with adequate attention to equipment and methods, there appears to be little meaningful difference between FSR and CR imaging modalities with respect to the visualization of small pneumoconiotic opacities. Second, additional procedures are required to reduce the variability of ILO classifications among readers, irrespective of imaging modality. Future studies should address the effect on reader performance of various recommended quality control programs.

ACKNOWLEDGMENTS

Author contributions: *Dr Laney*: contributed to the design and analysis of the study and the writing of the manuscript.

Dr Petsonk: contributed to the design and analysis of the study and the writing of the manuscript.

Dr Attfield: contributed to the design and analysis of the study and the writing of the manuscript.

Financial/nonfinancial disclosures: The authors have reported to *CHEST* that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Other contributions: We acknowledge the contribution of the National Institute for Occupational Safety and Health Coal Workers' Health Surveillance Program under the leadership of Anita Wolfe, John Wood for programming support, and the B readers who participated in the study. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health. Mention of product names does not imply endorsement by the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

REFERENCES

1. Antao VC, Petsonk EL, Sokolow LZ, et al. Rapidly progressive coal workers' pneumoconiosis in the United States: geographic clustering and other factors. *Occup Environ Med.* 2005;62(10):670-674.
2. US Department of Health and Human Services. *Work-Related Lung Disease Surveillance Report, 2007, Volume 1.* Washington, DC: US Department of Health and Human Services; 2008. DHHS (NIOSH) Pub. No. 2008-143a.
3. Ross MH, Murray J. Occupational respiratory disease in mining. *Occup Med (Lond).* 2004;54(5):304-310.

4. Laney AS, Attfield MD. Coal workers' pneumoconiosis and progressive massive fibrosis are increasingly more prevalent among workers in small underground coal mines in the United States. *Occup Environ Med.* 2010;67(6):428-431.
5. Wagner GR. *Screening and Surveillance of Workers Exposed to Mineral Dust.* Geneva, Switzerland: WHO; 1996.
6. International Labour Office. *Guidelines for the Use of the ILO International Classification of Radiographs of Pneumoconioses.* Geneva, Switzerland: International Labour Office; 2002.
7. Laney AS, Petsonk EL, Wolfe AL, Attfield MD. Comparison of storage phosphor computed radiography with conventional film-screen radiography in the recognition of pneumoconiosis. *Eur Respir J.* 2010;36(1):122-127.
8. Levine BA, Ingeholm ML, Prior F, et al. Conversion to use of digital chest images for surveillance of coal workers' pneumoconiosis (black lung). *Conf Proc IEEE Eng Med Biol Soc 2009.* 2009:2161-2163.
9. Sen A, Lee SY, Gillespie BW, et al. Comparing film and digital radiographs for reliability of pneumoconiosis classifications: a modeling approach. *Acad Radiol.* 2010;17(4):511-519.
10. Franzblau A, Kazerooni EA, Sen A, et al. Comparison of digital radiographs with film radiographs for the classification of pneumoconiosis. *Acad Radiol.* 2009;16(6):669-677.
11. National Institute for Occupational Safety and Health. Enhanced Coal Workers' Health Surveillance Program (ECWHSP). <http://www.cdc.gov/NIOSH/topics/surveillance/ORDS/ecwhsp.html>. Published 2009. Accessed July 19, 2011.
12. Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *Am J EEG Technol.* 1971;11:101-109.
13. Ashford JR. The classification of chest radiographs for coalworkers' pneumoconiosis. A study of the performance of two readers over a period of six years. *Br J Ind Med.* 1960;17:293-303.
14. Felson B, Morgan WK, Bristol LJ, et al. Observations on the results of multiple readings of chest films in coal miners' pneumoconiosis. *Radiology.* 1973;109(1):19-23.
15. Fletcher CM, Oldham PD. The problem of consistent radiological diagnosis in coalminers' pneumoconiosis; an experimental study. *Br J Ind Med.* 1949;6(3):168-183.
16. Musch DC, Landis JR, Higgins IT, Gilson JC, Jones RN. An application of kappa-type analyses to interobserver variation in classifying chest radiographs for pneumoconiosis. *Stat Med.* 1984;3(1):73-83.
17. 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(6):1740-1748.
18. Fernandez JM, Ordiales JM, Guibelalde E, Prieto C, Vano E. Physical image quality comparison of four types of digital detector for chest radiology. *Radiat Prot Dosimetry.* 2008;129(1-3):140-143.
19. Ganten M, Radeleff B, Kampschulte A, Daniels MD, Kauffmann GW, Hansmann J. Comparing image quality of flat-panel chest radiography with storage phosphor radiography and film-screen radiography. *AJR Am J Roentgenol.* 2003;181(1):171-176.
20. National Institute for Occupational Safety and Health. Recommended practices for reliable classification of chest radiographs by B readers. <http://www.cdc.gov/niosh/topics/chestradiography/radiographic-classification.html>. Updated May 24, 2011. Accessed July 19, 2011.