Comparison of Digital Radiographs with Film Radiographs for the Classification of Pneumoconiosis[1]

Alfred Franzblau, MD, Ella A. Kazerooni, MD, Ananda Sen, PhD, Mitchell M. Goodsitt, PhD, Shih-Yuan Lee, MS, Kenneth D. Rosenman, MD, MPH, James E. Lockey, MD, MS, Cristopher A. Meyer, MD, Brenda W. Gillespie, PhD, E. Lee Petsonk, MD, Mei Lin Wang, MD

Rationale and Objectives. Digital chest imaging has replaced film chest radiographs in many centers, but the International Labour Organization classification system, which is the most widely used system for recognition and classification of dust-related abnormalities, is predicated on film chest radiographs. The purpose of this study was to evaluate the equivalency of digital chest radiographs (including both hard copy and soft copy) with film radiographs for the recognition and quantification of abnormalities consistent with pneumoconiosis using the International Labour Organization classification system.

Materials and Methods. Digital chest images and film images, obtained from 107 subjects with a range of parenchymal and pleural abnormalities, were classified in random order by six B readers.

Results. Readings of film and soft copy images were equivalent for small opacity profusion; readings of hard copy images had significantly greater prevalence of small opacities compared to film and soft copy. The prevalence of large opacities differed significantly among all three image formats: hard copy greater than film greater than soft copy. However, film and soft copy readings for large opacities did not differ significantly when images demonstrating the coalescence of small opacities that had not yet become a large opacity were grouped with large opacities. The prevalence of pleural abnormalities differed significantly among all three image formats: film greater than hard copy greater than soft copy.

Conclusions. Film and soft copy images can be recommended for the recognition and classification of dust-related parenchymal abnormalities using International Labour Organization classifications. The role of digital radiography in reading for pleural abnormalities requires additional investigation.

Key Words. Chest radiographs; pneumonoconiosis; digital radiographs.

Acad Radiol 2009; 16:669–677

[1] From the Department of Environmental Health Sciences, University of Michigan School of Public Health, 109 South Observatory Street, Ann Arbor, 48109-2029 (A.F.); Department of Radiology, University of Michigan Medical School, Ann Arbor, MI (E.A.K, M.M.G.); Center for Statistical Consulting and Research, University of Michigan, Ann Arbor, MI (A.S.); Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI (S.-Y.L. B.W.G.); Department of Internal Medicine, Michigan State University Medical School, East Lansing, MI (K.D.R.); Departments of Environmental Health (J.E.L.) and Radiology (C.A.M.), University of Cincinnati Medical School, Cincinnati, OH; National Institute for Occupational Safety and Health, Morgantown, WV (E.L.P., M.L.W.). Received September 28, 2008; accepted December 8, 2008. Support for this research was provided by grant #S2200-22/23 from the Association of Schools of Public Health (ASPH) and the Centers for Disease Control and Prevention (CDC)/National Institute for Occupational Safety and Health (NIOSH). **Address correspondence to:** A.F. e-mail: afranz@umich.edu

doi:10.1016/j.acra.2008.12.020

Pneumoconioses are fibrotic or granulomatous lung diseases related to the inhalation of certain inorganic dusts (1–3). Workers in selected occupations have been at risk of pneumoconiosis since antiquity. Silica, asbestos, and coal have been the major dust hazards, though there are many other less commonly encountered materials that can present an inhalation hazard for pneumoconiosis.

In the 1930s, the International Labour Organization (ILO) based in Geneva, Switzerland, became involved in the development and evolution of a system for standardizing the classification of posteroanterior (PA) chest radiographs for pneumoconiosis (4). After multiple revisions, the ILO system remains the most widely used method for classifying chest radiographic abnormalities related to inhalation of pathogenic dusts (5–7).

The ILO system is predicated on use of film-screen radiographs (FSR) (5). In recent decades, many medical centers have introduced various forms of digital radiographic imaging into clinical practice. It has become difficult to obtain FSRs in many areas in the United States and it is anticipated this trend will continue.

Few studies have compared digital radiographic technology and FSR for the identification and quantification of abnormalities resulting from dust inhalation using the ILO system (8,9). The situation is further complicated by the fact that digital images can be viewed in two formats: laser-printed onto transparent medical imaging film as hard copy (HC) digital images and viewed with traditional view boxes like FSR images, or soft copy (SC) digital images that are viewed on a computer workstation monitor.

The goal of the present investigation was to assess and compare the impact of image format (FSR, SC, and HC) on the recognition and quantification of dust-related abnormalities when ILO classifications are performed by physicians experienced in interpreting radiographs for pneumoconiosis.

MATERIALS AND METHODS

The study had three phases: 1) recruitment of subjects and capturing of images; 2) reading of images (FSR, SC, HC) by a panel of six readers certified by the National Institute for Occupational Safety and Health in the use of the ILO system (ie, B readers); and 3) data cleaning and statistical analyses. Subjects gave written informed consent that was approved by the Medical Institutional Review Board of the University of Michigan and complied with Health Insurance Portability and Accountability Act regulations.

Recruitment of Subjects and Questionnaire

The study was performed between 2002 and early 2007, and subject recruitment was from early 2003 through early 2005. Subjects had to be at least 18 years old, capable of giving informed consent, and capable of coming to the University of Michigan Medical Center to complete the required radiographic studies and a short written questionnaire. Pregnant women were not eligible to avoid potential radiation exposure to the fetus. Subjects were recruited from two sources: patients seen or referred to the University of Michigan or patients listed in the Michigan or Ohio Silicosis Registries (10,11).

A primary objective of subject recruitment was to achieve a final study group that had adequate representation of all major ILO small opacity profusion categories (ie, 0, 1, 2, and 3). More subjects were sought in the two lower ILO major profusion categories (ie, 0, 1) because it was considered important to characterize differences at the critical 0/1 and 1/0 boundary. Other recruitment objectives were to include adequate numbers of subjects with a range of pleural abnormalities such as found with asbestos exposure and to have a reasonable balance of increased profusion of both "rounded" and "irregular" small opacities so that our findings would be applicable to all forms of pneumoconiosis. An important, but secondary recruitment objective was to recruit subjects with large pneumoconiotic opacities. These recruitment objectives were based partly on use of the κ statistic to assess inter-rater and intrarater agreement (note: results of κ analyses will be presented elsewhere). The value of κ depends, in part, on the prevalence of "positive" findings: κ approaches zero when the true prevalence approaches zero or 100% (12). Hence, recruitment was designed so that the prevalence of key "positive" findings would be bounded away from these extremes.

All subjects completed a survey that included questions in the following areas: demographics (age, gender, race/ethnicity); smoking history (13); occupational and dust exposure history; and medical history, particularly related to lung and pleural disease. Height and weight were measured.

Chest Radiographic Imaging Methods and Quality Control

For each subject, a standard PA FSR image and a PA digital radiographic (DR) image were obtained on the same day. One DR image was lost in the Picture Archiving and Communication System and was not recoverable, and one FSR was lost in the radiology file room, each on different subjects. Therefore, the final study group included 106 images for FSR, HC, and SC, respectively, but these were based on 107 subjects.

All radiographic images were obtained at the University of Michigan Medical Center. Protocols for acquiring the DR and FSR images for this study were developed to ensure that uniform techniques were employed and to maintain detailed records of the examinations. The protocol for capturing DR chest images was as follows.

1. Images were captured on the flat-panel amorphous Selenium digital detector of the Hologic DR 1000C system (Hologic, Inc., Bedford, MA) with the following settings: 120 kVp, 160 or 250 mA, 72" (183 cm) source-to-image distance with left and right (lung field) phototimer sensors. A 10:1, 70 grid lines/cm, 165 cm focus, scatter rejection grid was employed. The Hologic (direct ray) detector has a pixel size of 139 μm. The image array size is 2560 × 3072 pixels (35 cm × 43 cm).
2. The kVp, mA, exposure time, and mAs were recorded on the patient record sheet.

3. The digital image was sent to the thoracic radiology reading room for routine clinical interpretation and was archived.
4. Subject identifiers were stripped out of the DICOM header of DR image files, a unique subject identifier number was inserted into the header, and the deidentified DICOM file was stored for research purposes.
5. The deidentified digital image was printed on a Fuji FM-DPL high-quality laser printer (FUJIFILM Medical Systems USA, Inc., Stamford, CT) using Fuji film. Images were printed using the standard look-up table recommended by Fuji for printing hard copies (LUT Gamma #17, linear-tone, blue base, OD range: 0.15 to 3). Hard copies were printed in 100% format (ie, full size, no reduction).

The final digital images were generated using the standard image processing settings for PA chest radiographs on the Hologic DR system employed in the University of Michigan radiology department at the time of the study. These Hologic-specific settings were: density = 0.2; adaptive = OFF; fixed latitude = 75; adaptive latitude = 110; IT weight = 0.0; DRC filter = 0; look-up table = 3; collimate = OFF; index method = 2; levels = 6; region 1 = 2.0; region 2 = 2.0; and IT param = 218. An investigation to determine the optimal numerical processing algorithm for diagnosing lung disease from Hologic DR chest images was beyond the scope of this study.

The protocol for capturing FSR images was as follows.

1. Standard PA chest film/screen technique was employed: 125 kVp, 150 mA, wall unit, 72" (183 cm) source-to-image distance, all three phototimer sensors, Agfa UVC film (Agfa-Gevaert Group, Wilmington, DE) in Agfa UV Super Rapid Screen Cassette, Normal "0" density setting (the speed of the screen-film system was 200). Either a 10:1 or a 12:1, 40 grid lines/cm, 152–183 cm focal range (for each), scatter rejection grid was employed.
2. The kVp, mA, exposure time, and mAs were recorded on the patient record sheet.
3. Exposed film was developed in an automatic processor using the standard 90-second cycle and using Agfa/DuPont HSG chemistry, and the film was then placed in the patient's film folder.
4. After clinical interpretation, patient identifiers on the film were covered with opaque tape, and a unique subject identifier number was written on the tape.

Standard ILO Radiographic Images

All B readers used their own set of 22 ILO standard films when reading FSR and HC images, in accordance with ILO guidelines (5).

To allow for side-by-side comparison of SC images with ILO standard images, it was necessary to develop a digital version of the ILO standard films. With permission from the ILO, we accomplished this by digitizing and archiving the entire set of ILO standard films. The films were digitized using a VIDAR Diagnostic Pro Scanner (VIDAR Systems Corporation, Herndon, VA), using standard chest processing. The films were digitized to 12 bits (4096 gray levels) with a spatial resolution of 142 μm. The digital images were then reviewed on a high-quality display and two participants in the study adjusted the window width and level of each image on the monitor so that the resulting image approximated the corresponding FSR standard image on a view box. A list of the window level and width values for each standard film is available from the authors. The digitized ILO standard images were supplied to each of the participating B readers.

Reading of Images

The six physicians who interpreted images for this study were all B readers. Their durations of certification ranged from 4 to 24 years, with a mean of 13 years.

To ensure that all B readers employed high-quality display monitors when reading SC images, they were required to use a high-resolution physician-quality diagnostic workstation similar to that employed in the radiology department at the University of Michigan at the time of the study (ie, SIEMENS SIENET Magic View 1000 system). This system consists of:

1. A Sun UltraSPARC 60 workstation with 2x DOME dual-head Md5px frame buffer.
2. Software version VB33C including display functions such as: window level and width (brightness and contrast) adjustment, continuous pan/zoom, user selectable zoom factor, full resolution and full view image presentation, a "magnifying glass" tool with independent window/level adjustments, and a region-of-interest measurement tool with mean pixel value and standard deviation readings.
3. SIMOMED 2 K Grayscale CRT monitors (four monitors per workstation) with the following characteristics:

Resolution (pixels) 2048 × 2560
Maximum luminance:
10 cm × 10 cm area 600 cd/m^2
Fully white screen 350 cd/m^2

The specifications most critical for this study were the 2k-monitor resolution, maximum brightness levels, and window/level adjustment, zooming, panning, and magnifying glass features. It was also critical that the brightness and contrast of the monitors were adjusted to satisfy the DICOM Grayscale Standard Display Function (14,15). We achieved

Table 1
Reading Order of Image Formats by B Readers

	Round 1			Round 2		
Reader	Cycle 1	Cycle 2	Cycle 3	Cycle 1	Cycle 2	Cycle 3
1	Hard*	Film†	Soft‡	Hard	Soft	Film
2	Soft	Film	Hard	Film	Soft	Hard
3	Film	Soft	Hard	Film	Hard	Soft
4	Hard	Soft	Film	Soft	Hard	Film
5	Film	Hard	Soft	Hard	Film	Soft
6	Soft	Hard	Film	Soft	Film	Hard

* Hard = digital radiograph laser-printed on film, and the hard copy viewed on a standard radiographic viewbox.
† Film = traditional film-screen radiograph viewed on a standard radiographic viewbox.
‡ Soft = digital radiograph viewed as soft copy on a diagnostic quality workstation.

this on the SIMOMED 2K monitors by adjusting the monitor settings so that the luminance of the 20% gray step of the Society of Motion Picture and Television Engineers pattern was set at 6.0 ± 0.2 cd/m^2 and the 90% gray step was set at 152.6 ± 0.5 cd/m^2 (16). Other monitors that were employed in the study at the other sites were adjusted similarly.
B readers were able to control the displayed brightness and contrast of each individual image via the window level and width settings on the workstation. They also could change the magnification when reading the soft copy images.

There were two rounds of readings; a round consisted of each reader classifying all of the images in each of the formats (FSR, HC, and SC). The three image formats were presented to the readers in random order (Table 1). Within each image format the order of presentation of the individual radiographs was also randomized. Readers classified images according to the 2000 revision of the ILO classification system (5). There were at least 30 days between readings of each set of images for each reader.

Statistical Analyses

Models evaluated marginal rating differences across image formats with and without controlling for potential confounders. Both generalized estimating equations and mixed model approaches were adopted for statistical analyses to account for the clustering effect induced by multiple ratings made on the same subject using different image formats. The generalized estimating equation approach was implemented using the SAS procedure GENMOD to analyze outcomes that are both continuous and discrete. In these models *image format*, *reader*, and *round*, have been used as fixed effects, and the *subject id* was used as a *random effect*. To make the analysis robust (ie, not too dependent on the model assumptions), a compound symmetry/sphericity working correlation matrix was used to model the dependence between

Table 2
Subject Characteristics

	Frequency	Percent
Gender		
Male	86	80
Female	21	20
Body mass index (kg/m^2)		
<25 (normal)	28	26
≥25 to <30 (overweight)	45	42
≥30 (obese)	34	32
Ever smoked		
No	39	36
Yes	68	64
Current smoking		
No	97	91
Yes	10	9
History of dust exposure		
No	47	44
Yes	60	56
Dust exposure type (n = 60)*		
Silica	34	57
Asbestos	28	47
Other/unknown	12	20
	Mean (S.D.)	Median (range)
Age (y)	64.7 (11.9)	65 (31–91)
Body mass index (kg/m^2)	28.5 (5.2)	28.1 (19.5–48.8)
Pack-years		
All subjects (n = 107)	19.5 (24.1)	12 (0–96)
Ever smoked (n = 68)	30.7 (23.8)	23.5 (1–96)

SD, standard deviation.
* Some subjects reported more than one type of dust exposure, so the numbers add up to more than 60.

the measurements made on the same subject (the only two exceptions being parenchymal abnormalities and large opacities or "ax" where we had to use an independent correlated structure because of convergence issues). Variables that could potentially influence ratings (eg, age, gender, body mass index [BMI], and pack-years of smoking), were included in models as covariates. Models of outcomes other than image quality were also adjusted for median image quality. Mixed model analyses were performed using SAS/STAT GLIMMIX for dichotomous outcomes and MIXED for continuous outcomes. All analyses were performed using SAS/STAT for Windows version 9.1 (17).

RESULTS

Of the 107 subjects, 80% (n = 86) were male, 56% (n = 60) had a history of pathogenic dust exposure, 64% (n = 68) were current or former smokers, and 74% (n = 79) were overweight or obese (ie, BMI ≥25 kg/m^2). The mean subject age was 64.7 years (range, 31–91) (Table 2).

Table 3
Results of ILO Classifications Overall and by Image Format

Outcome variable	Overall		Film		Hard Copy		Soft Copy		
	n	%	*n*	%	*n*	%	*n*	%	X^2 (*P* Value)
1. Image quality (*n* = 3816*)									
1	1130	29%	398	31%	301	24%	431	34%	76.62
2	2282	60%	774	61%	778	61%	730	57%	(<.0001)
3	382	10%	98	8%	175	14%	109	9%	(6 df)
4 (unreadable)	22	1%	2	0%	18	1%	2	0%	
2A. Any parenchymal abnormalities (*n* = 3794*)									11.97
No	1216	32%	443	35%	358	29%	415	33%	(.0025)
Yes	2578	68%	827	65%	896	71%	855	67%	(2 df)
2Ba. Shape/size of primary small opacities (*n* = 2578)									1.85
Round (p, q, r)	829	32%	281	34%	280	31%	268	31%	(.3958)
Irregular (s, t, u)	1749	68%	546	66%	616	69%	587	69%	(2 df)
2Bc. Small opacity profusion									
0	1529	40%	543	43%	455	36%	531	42%	16.55
1	1158	31%	385	30%	392	31%	381	30%	(.0111)
2	852	22%	265	21%	306	25%	281	22%	(6 df)
3	255	7%	77	6%	101	8%	77	6%	
2C. Large opacities									
0	3216	85%	1076	85%	1036	83%	1104	87%	14.12
A	228	6%	78	6%	79	6%	71	6%	(.0284)
B	271	7%	93	7%	101	8%	77	6%	(6 df)
C	79	2%	23	2%	38	3%	18	1%	
2C. Large opacities									9.09
No (0)	3216	85%	1076	85%	1036	83%	1104	87%	(.0106)
Yes (A or B or C)	578	15%	194	15%	218	17%	166	13%	(2 df)
2C. Large opacities or ax									7.87
No (0)	3026	80%	1020	80%	969	77%	1037	82%	(.02)
Yes (A or B or C or ax)	768	20%	250	20%	285	23%	233	18%	(2 df)
3A. Pleural abnormalities									30.26
No	2585	68%	795	63%	868	69%	922	73%	(<.0001)
Yes	1209	32%	475	37%	386	31%	348	27%	(2 df)
3C. Costophrenic angle obliteration									6.37
No	3546	93%	1169	92%	1183	94%	1194	94%	(.0413)
Yes (right and/or left)	248	7%	101	8%	71	6%	76	6%	(2 df)
3D. Diffuse pleural thickening									4.52
No	3620	95%	1199	94%	1201	96%	1220	96%	(.1043)
Yes (right and/or left)	174	5%	71	6%	53	4%	50	4%	(2 df)

df = degrees of freedom.

* Images were obtained for each of the three modalities in 107 subjects and were classified on two separate occasions by six B Readers. The number of images assessed for film quality is greater than for subsequent outcomes. For a small number readings, image quality was rated unreadable (*n* = 22). These readings provide no data for subsequent outcomes.

A total of 3816 readings were completed for the study (106 images × 3 formats × 6 readers × 2 rounds; Table 3). The subject recruitment goals related to the prevalence of key outcomes (ie, small opacities, large opacities, and pleural abnormalities) were achieved. Small opacity profusion categories "0" and "1" together represented a majority (73%) of the scores for FSR images, as was intended by design. Among FSR images marked positive for small opacities, there was a reasonable representation of both small rounded (34%) and small irregular opacities (66%). Fifteen percent of readings of FSR images indicated the presence of large opacities, and 37% indicated the presence of pleural abnormalities. Table 3 also displays the results for readings based on HC and SC images, and statistical comparisons among all three image formats based on χ^2 statistics.

Tables 4 and 5 display unadjusted pairwise comparisons of the marginal prevalence of outcomes by image format. There were no significant differences in image quality ratings

Table 4
Pairwise Comparisons of Prevalence of Findings by Image Format (FSR, HC, and SC) based on Dichotomous Outcomes

Classification Comparison	FSR vs. HC*	FSR vs. SC*	HC vs. SC*
1A: Film quality (category 1 vs. 2, 3, 4)	0.67 (0.49–0.92)	1.11 (0.85–1.45)	1.66 (1.39–1.96)
1.A: Film quality (categories 1, 2 vs. 3, 4)	0.47 (0.31–0.73)	0.89 (0.56–1.41)	1.87 (1.53–2.30)
2.A: Parenchymal abnormalities (yes/no)	0.75 (0.65–0.86)	0.91 (0.80–1.04)	1.22 (1.09–1.35)
2.C: Large opacities (yes/no)	0.86 (0.75–0.98)	1.18 (1.03–1.36)	1.38 (1.20–1.58)
2.C: Large opacities with ax (yes/no)	0.83 (0.74–0.93)	1.07 (0.98–1.17)	1.29 (1.16–1.44)
3.A: Pleural abnormalities (yes/no)	1.30 (1.10–1.53)	1.53 (1.31–1.78)	1.18 (1.04–1.33)
3.C: Costophrenic angle obliteration (yes/no)	1.45 (0.99–2.11)	1.36 (0.93–1.99)	0.94 (0.79–1.12)
3.D: Diffuse pleural thickening (yes/no)	1.35 (0.94–1.95)	1.45 (0.99–2.12)	1.07 (0.84–1.37)

FSR, film-screen radiographs; HC, hard copy; SC, soft copy.
* Estimate of odds ratio (95% confidence interval) based on logistic regression with generalized estimating equations accounting for clustering by subject.

Table 5
Pairwise Comparisons of Mean Scores by Image Format (FSR, HC, and SC) based on Ordinal Outcomes

Classification Comparison	FSR vs. HC*	FSR vs. SC*	HC vs. SC*
1.A: Film quality (4-point scale)	−0.166 (.0002)	0.013 (.7366)	0.179 (<.0001)
2.B: Small opacities (12-point scale)	−0.381 (<.0001)	−0.028 (.6771)	0.353 (<.0001)
2.B: Small opacities (4-point scale)	−0.136 (<.0001)	−0.015 (.5596)	0.122 (<.0001)
2.C: Large opacities (4-point scale)	−0.051 (.0142)	0.041 (.0098)	0.092 (<.0001)

FSR, film-screen radiographs; HC, hard copy; SC, soft copy.
* Estimate of difference (*P* value) based on continuous models with generalized estimating equations accounting for clustering by subject.

between FSR and SC, whereas the quality ratings for the HC images were significantly higher (ie, worse) than ratings for both FSR and SC images. Results of corresponding models that also included adjustments for age, gender, BMI, pack-years of smoking, round, individual B readers, and image quality were similar (Tables 6, 7).

The prevalence of parenchymal abnormalities (yes/no) was not significantly different between FSR and SC images (Table 4), whereas HC images were judged to show parenchymal abnormalities more frequently. A similar pattern emerged when considering either the 12-point or the 4-point ILO scales (Table 5). Inclusion of covariates in models for the shapes of primary and secondary small opacities (ie, rounded and irregular) had little effect on the magnitude, direction or statistical significance of the mean effects of image format (Table 7).

The contribution to the total model log-likelihood was partitioned into components by reader, round, and image format. The relative percentages were calculated for models of parenchymal abnormalities (yes/no – 86.5%, 2.3%, and 11.2%), small opacities on the 12-point scale (70.8%, 3.2%, and 26.0%), and small opacities on the 4-point scale (74.7%, 3.8%, and 21.5%), respectively. Overall, although some statistically significant differences between formats were observed, the variance between formats was small compared to the variance between readers.

In models of large opacities, all pairwise comparisons of image formats differed significantly, with the following ranking of the formats for the prevalence of large opacities: HC greater than FSR greater than SC (Tables 4 and 5). The symbol "ax" is used in the ILO classification scheme to identify the presence of a coalescence of small opacities that has not yet become a large opacity. The dichotomous model for large opacities was also analyzed by combining ax and large opacities as positive outcomes (there is no size classification for ax, so it is not possible to include ax in a continuous model for large opacities (ie, the 4-point scale). Categorizing ax with large opacities had little impact on the odds ratios for FSR versus HC or HC versus SC, but the odds ratio in the model for FSR versus SC was closer to the null and was no longer statistically significant (Table 4).

Pleural abnormalities were most frequently recorded with FSR and least frequently with SC images: FSR greater than HC greater than SC (Tables 3 and 4). The impact of image format on the prevalence of pleural findings was explored further by dichotomizing pleural abnormalities according to thickness: ≤5 versus >5 mm. This re-parameterization did not change the results (not shown). The relative contributions to the model log-likelihood of reader, round and image format for pleural abnormalities (80.2%, 2.7%, and 17.1%, respectively) were similar to the results for parenchymal

Table 6
Adjusted Pairwise Comparisons of Prevalence of Findings by Image Format (FSR, HC, and SC) based on Dichotomous Outcomes

Classification Comparison	FSR vs. HC*	FSR vs. SC*	HC vs. SC*
1.A: Film quality (Category 1 vs. 2, 3, 4)	0.65 (0.46–0.91)	1.12 (0.84–1.49)	1.72 (1.43–2.08)
1.A: Film quality (Categories 1, 2 vs. 3, 4)	0.42 (0.24–0.71)	0.87 (0.50–1.54)	2.10 (1.63–2.70)
2.A: Parenchymal abnormalities (yes/no)	0.72 (0.60–0.86)	0.90 (0.78–1.04)	1.26 (1.09–1.46)
2.C: Large opacities (yes/no)	0.83 (0.70–0.99)	1.23 (1.04–1.46)	1.48 (1.24–1.76)
2.C: Large opacities with ax (yes/no)	0.79 (0.66–0.94)	1.12 (0.99–1.27)	1.43 (1.22–1.67)
3.A: Pleural abnormalities (yes/no)	1.28 (1.08–1.53)	1.59 (1.35–1.88)	1.24 (1.08–1.42)
3.C: Costophrenic angle obliteration (yes/no)	1.41 (0.99–2.00)	1.39 (0.98–1.97)	0.98 (0.80–1.22)
3.D: Diffuse pleural thickening (yes/no)	1.32 (0.97–1.80)	1.43 (1.04–1.98)	1.08 (0.84–1.40)

FSR, film-screen radiographs; HC, hard copy; SC, soft copy.

* Estimate of odds ratio (95% confidence interval) based on logistic regression with generalized estimating equations accounting for clustering by subject. All models are adjusted for age, gender, body mass index, pack-years of smoking, round and individual B readers. Models of outcomes other than image quality were also adjusted for median image quality. When we used mixed model analyses, the overall direction and significance of results was unchanged, suggesting that the parameter estimates are robust to assumptions about the covariance structure (not shown).

abnormalities. The image formats did not differ significantly in most models of costophrenic angle obliteration and diffuse pleural thickening, but the prevalence of these findings was low (<10%), and the power was reduced for these outcomes.

DISCUSSION

ILO classifications of HC images demonstrated a significantly higher prevalence of small parenchymal opacities compared to FSR and SC images from the same individual; the FSR and SC readings did not differ significantly. Recognition of large opacities differed significantly among all three image formats, being most prevalent for HC images, and least prevalent for SC images. However, a significant difference between FSR and SC readings was no longer apparent when ax was categorized with large opacities. All three image formats differed significantly for recognition of pleural abnormalities, with FSR having the highest prevalence and SC having the lowest prevalence of such findings. The relative contribution of image format to the log-likelihood of models was small compared to readers.

Zähringer et al (8) compared HC images obtained with a selenium drum detector (Thoravision: Philips Medical Systems, Hamburg, Germany) to traditional FSR from 50 miners (exposed to uranium and quartz dust – not asbestos) interpreted by four readers using the ILO system. More than 95% of readings were $\leq 1/0$; only two subjects had large opacities. They concluded that there was no difference between image formats for identification of parenchymal abnormalities, that pleural findings were more common with HC than FSR, and that the image quality of HC images was rated significantly better than FSR. Compared to the present study, the study by Zähringer et al (8) was limited in power (fewer subjects overall, few subjects with parenchymal abnormalities, and fewer readers), scope (no SC, few large opacities, no subjects with exposure to asbestos), and analyses (no adjustment for potential confounders).

Takashima et al (9) compared flat panel detector digital radiographs, storage phosphor computed radiographs, and conventional analog radiographs in the detection of small parenchymal opacities. The study involved 30 patients (only 10 had ILO profusion score $\geq 1/0$) and three readers. Digital images were viewed as hard copy only. Large opacities and pleural abnormalities were not considered as outcomes. Each reader compared the three images for each subject side-by-side at the same time. There were no formal statistical comparisons of the κ values across the three image formats. The authors concluded that the 12-point profusion scores for flat panel detector images and conventional radiographs did not differ significantly, but that computed radiographs differed from the other two image formats. This study was small, and therefore also limited in power (fewer subjects, few subjects with parenchymal abnormalities, and fewer readers), scope (no SC, no analyses related to pleural findings or large opacities), and analyses (no adjustment for potential confounders, no formal statistical comparison across image formats). Also, it is unclear how the side-by-side viewing methodology may have influenced results.

In the present study, FSR and HC images were rated using the ILO standard films, whereas SC images were classified using a digitized version of the ILO standards. Hence, the present study was an assessment of the combined effect of reading SC images using the digitized version of the ILO standards created for this study. The equivalence of outcomes related to parenchymal abnormalities (when the symbol "ax" is considered equivalent to a large opacity) for FSR and SC suggests that the digitized version of the ILO standards

Table 7
Adjusted Pairwise Comparisons of Mean Scores by Image Format (FSR, HC, and SC) based on Ordinal Outcomes

	FSR vs. HC*	FSR vs. SC*	HC vs. SC*
1.A: Film quality (4-point scale)	−0.166 (.0002)	0.013 (.7379)	0.179 (<.0001)
2.B: Small opacities (12-point scale)	−0.420 (<.0001)	−0.026 (.6871)	0.393 (<.0001)
Adjusted for primary and secondary shape	−0.404 (<.0001)	−0.048 (.5652)	0.452 (<.0001)
2.B: Small opacities (4-point scale)	−0.148 (<.0001)	−0.014 (.5642)	0.134 (<.0001)
Adjusted for primary and secondary shape	−0.150 (<.0001)	−0.007 (.8232)	0.157 (<.0001)
2.C: Large opacities (4-point scale)	−0.058 (.0093)	0.041 (.0078)	0.099 (<.0001)

FSR, film-screen radiographs; HC, hard copy; SC, soft copy.

* Estimate of difference (*P* value) based on continuous models with generalized estimating equations accounting for clustering by subject. All models are adjusted for age, gender, BMI, pack-years of smoking, round and individual B readers. Models of outcomes other than image quality were also adjusted for median image quality. Models for 4-point and 12-point scales were additionally adjusted for primary and secondary shape of small opacities. When we used mixed model analyses, the overall direction and significance of results was unchanged, suggesting that the parameter estimates are robust to assumptions about the covariance structure (not shown).

created for this study may provide an approach for the implementation of the ILO classification for SC images.

Differences of less than half of a minor category were highly significant (eg, FSR vs. HC and SC vs. HC; Table 5). In fact, the study had 90% power to detect a mean difference between image formats of 0.17 units on the 12-point ILO scale.

The present study has a number of strengths. The study had adequate power to assess key outcomes, including ILO classifications for small opacities, large opacities, and pleural abnormalities. The patient population had a reasonable mix of irregular and rounded small opacities. The study images were obtained using well-defined standard methods and up-to-date equipment, and the readers read SC images on high-quality work stations. The study evaluated both HC and SC images. A digital version of the ILO standard images was created so that the readers could perform side-by-side comparisons when reading SC images, thereby demonstrating the feasibility of this approach. Models with and without adjustment for multiple potential confounders were assessed, but parameter estimates did not differ, which is not surprising because the study design was balanced by reader, round, format, and subject.

The present study also has a number of weaknesses. No independent gold standard was used, such as chest computed tomographic examinations or pathology, for determining the presence of parenchymal and pleural abnormalities. For example, although significant differences in the prevalence of pleural abnormalities among image formats were found, these results do not indicate which format is closer to the "truth." It is recognized that FSR has low sensitivity to pleural changes in comparison to autopsy results (18).

There are other challenges not addressed in the present study. HC images were printed in 100% format in the present study. In many centers, HC images are laser-printed in reduced format, such as a 66% scale HC (19). It has been shown that reduction of image size by 50% or more leads to loss of detection accuracy (20). The importance of the issue of HC image size is diminished by the findings that classifications of HC images differed significantly from readings of FSR and SC images for prevalence of parenchymal and pleural abnormalities; therefore, HC may not be a preferred choice for ILO classifications regardless of image size.

Another important issue that is beyond the scope of the present study is digital image processing. For a variety of reasons, digital radiographic images are processed numerically before interpretation (21,22). The choice of processing parameters is critical because inappropriate processing can potentially produce false-negative or false-positive results. The lack of standardization of numerical processing of digital radiographic images is somewhat analogous to variation in film characteristic curves and exposure techniques with FSR. However, the potential for image manipulation with digital radiographic image processing is much greater than with FSR, and can be harder to detect (eg, processing parameters may not be displayed explicitly in the final digital image, processing parameters are not consistent among equipment manufacturers). At present, there is no empirical basis for the choice of numerical processing parameters for chest digital images for optimal identification of parenchymal or pleural abnormalities consistent with pneumoconiosis. Research is needed to determine acceptable or optimal numerical image processing parameters for digital radiographic images for pneumoconiosis classification.

In summary, ILO classification outcomes for small opacities and large opacities (when the symbol "ax" is considered equivalent to a large opacity) are similar using FSR and SC images (but not HC images). Both FSR and SC images can therefore be recommended for the recognition and classification of dust-related parenchymal abnormalities when the methods used are similar to those in this study. A digital standard set of ILO images has been generated and

demonstrated for use when reading SC images. Digital radiography (both HC and SC) differed significantly from FSR for the detection of pleural abnormalities; therefore, the role of digital radiography in reading for pleural abnormalities requires additional investigation.

ACKNOWLEDGMENT

We wish to thank James Good and University of Michigan in the Janis Huff Radiology Department for their diligence, support, and good cheer. Emmanuel Christodoulous, PhD, of the University of Michigan Radiology Department assisted with the digitization of the ILO standard radiographs. We are grateful to Drs. Kurt Hering and Igor Fedotov at the International Labour Organization for giving permission to digitize the ILO standard films for use in this study. Mary Jo Reilly at the Michigan Silicosis Registry and Ed Socie at the Ohio Silicosis Registry provided invaluable assistance with recruitment of subjects. We thank Bethany Baker for her commitment to this project, and for being the glue that helped to hold it together and keep it moving forward. Drs. Michael Attfield and Robert Castellan at NIOSH reviewed the manuscript and provided many thoughtful and helpful suggestions. Mention of commercial products is not an endorsement by the authors, ASPH, CDC, NIOSH, or the ILO. The findings and conclusions contained in this report are those of the authors and do not necessarily reflect those of ASPH, CDC, NIOSH, or the ILO.

REFERENCES

1. Harber P, Schenker M, Balmes J. Occupational and environmental respiratory diseases. St. Louis: Mosby, 1995.
2. Rosenstock L, Cullen MR, Brodkin CA, et al, eds. Textbook of clinical occupational and environmental medicine. 2nd ed. Philadelphia: Elsevier Saunders, 2005.
3. Petsonk EL, Parker JE. Coal workers' lung diseases and silicosis. In: Fishman AP, Fishman JA, Grippi MA, et al, eds. Fishman's pulmonary diseases and disorders. 4th ed. New York: McGraw-Hill, 2008. p. 967–980.
4. Musch DC. Interobserver variation in classifying radiographs for the pneumoconioses. Ann Arbor, MI: University of Michigan Doctoral Dissertation, 1981.
5. International Labour Organization (ILO). Guidelines for the use of ILO international classification of radiographs of pneumoconioses, Revised edition 2000. Geneva: International Labour Office, 2002.
6. Pham QT. Chest radiography in the diagnosis of pneumoconiosis. Int J Tuberc Lung Dis 2001; 5:478–482.
7. Mulloy KB, Coultas DB, Samet JM. Use of chest radiographs in epidemiological investigations of pneumoconioses. Br J Ind Med 1993; 50: 273–275.
8. Zähringer M, Piekarski C, Saupe M, et al. Comparison of digital selenium radiography with an analog screen-film system in the diagnostic process of pneumoconiosis according to ILO classification [in German]. Fortschr Röntgenstr 2001; 173:942–948.
9. Takashima Y, Suganuma N, Sakurazawa H, et al. A flat-panel detector digital radiography and a storage phosphor computed radiography: screening for pneumoconioses. J Occup Health 2007; 49:39–45.
10. Rosenman KD, Reilly MJ, Kalinowski DJ, et al. Silicosis in the 1990's. Chest 1997; 111:779–786.
11. Reilly MJ, Rosenman KD, Watt FC, et al. Silicosis surveillance - Michigan, New Jersey, Ohio, and Wisconsin, 1987–1990. MMWR 1993; 42:23–28.
12. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. J Clin Epidemiol 1988; 41:949–958.
13. American Thoracic Society. Epidemiology standardization project: recommended respiratory disease questionnaires for use with adults and children in epidemiological research. Am Rev Resp Dis 1978; 118:7–53.
14. Samei E, Rowberg A, Avraham E, et al. Toward clinically relevant standardization of image quality. J Digital Imaging 2004; 17:271–278.
15. Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. Med Phys 2005; 32:1205–1225.
16. Nawfel RD, Chan KH, Wagenaar DJ, et al. Evaluation of video gray-scale display. Med Phys 1992; 3:561–567.
17. SAS Institute Inc., SAS/STAT User's Guide, Version 9.1, Cary, NC: SAS Institute Inc., 2002.
18. Frumkin H, Pransky G, Cosmatos I. Radiologic detection of pleural thickening. Am Rev Resp Dis 1990; 142:1325–1330.
19. MacMahon H. Digital chest radiography: practical issues. J Thorac Imaging 2003; 18:138–147.
20. Schaefer C, Prokop M, Oestmann J, et al. Impact of hard-copy size on observer performance in digital chest radiography. Radiology 1992; 184: 77–81.
21. Prokop M, Neitzel U, Schaefer-Prokop C. Principles of image processing in digital chest radiography. J Thorac Imaging 2003; 18:148–164.
22. McAdams HP, Samei E, Dobbins J, et al. Recent advances in chest radiology. Radiology 2006; 241:663–683.