COMPARISON OF DIGITAL RADIOGRAPHS WITH FILM-SCREEN RADIOGRAPHS FOR CLASSIFICATION OF PNEUMOCONIOSIS

Alfred Franzblau, University of Michigan, Ann Arbor, MI, USA
Ella A. Kazerooni, University of Michigan, Ann Arbor, MI, USA
Ananda Sen, University of Michigan, Ann Arbor, MI, USA
Mitch Goodsitt, University of Michigan, Ann Arbor, MI, USA
Shih-Yuan Lee, University of Michigan, Ann Arbor, MI, USA
Kenneth Rosenman, Michigan State University, East Lansing, MI, USA
James Lockey, University of Cincinnati, Cincinnati, OH, USA
Cristopher Meyer, University of Cincinnati, Cincinnati, OH, USA
Brenda Gillespie, University of Michigan, Ann Arbor, MI, USA
Lee Petsonk, NIOSH, Morgantown, WV, USA
Mei Lin Wang, NIOSH, Morgantown, WV, USA

Comparison of DR with FSR

- The International Labour Organization (ILO) system for classifying chest radiographic changes related to inhalation of pathogenic dusts is predicated on film-screen radiography (FSR).

- Digital radiography (DR) has replaced FSR in many centers, but there are few data to indicate whether DR is equivalent to FSR in identifying and quantifying interstitial and pleural abnormalities.

- DR images can be printed and viewed on film ('hard copy'-HC) or can be viewed on a monitor at a computer workstation ('soft copy'-SC).

Comparison of DR with FSR

- The goal of the present investigation was to assess the impact of chest radiograph image format (FSR, SC, or HC) on ILO classifications performed by experienced readers in individuals with abnormalities of the lung parenchyma and/or pleura that may result from dust inhalation.

- We compared the reliability of classifications across three image formats (FSR, SC & HC), and we also compared the prevalence of findings across image formats.

Study Methods

- The study had three phases:
  - recruitment of subjects and capturing of FSR and DR images
  - reading of images (FSR, SC, & HC) by the panel of readers
  - data cleaning and statistical analyses

Subject Recruitment

- Primary subject recruitment objectives:
  - adequate representation of all major ILO small opacity profusion categories (i.e., "0", "1", "2" and "3"), but with a somewhat heavier emphasis on ILO major categories "1" and "0"
  - adequate numbers of subjects with pleural abnormalities
  - reasonable balance of increased profusion of both 'rounded' and 'irregular' small opacities

Subject Recruitment

- Secondary recruitment objective:
  - to recruit subjects with large pneumoconiotic opacities

Subject Recruitment

- Other data elements:
  - Questionnaire (demographics, smoking history, work/dust exposure history, medical history)
  - Measured height and weight
  - Performed standard PA FSR and PA DR image on the same day
  - No clinical examinations or pulmonary function tests

Image Capture Methods

- FSR and DR images were captured from 107 subjects

  - DR images were captured on a flat-panel amorphous Selenium digital detector of a Hologic DR system

  - FSR images used standard PA chest film/screen technique (125 kVp, 150 mA, wall unit, 183 cm SID, all 3 phototimer sensors, Agfa UVC film in Agfa UV Super Rapid Screen Cassette, Normal "0" density setting - the speed of the screen-film system was 200)

Image Readings

- One DR image was lost in the PACS, and one FSR radiograph was lost in the radiology file room (these involved different subjects). Therefore, the final study group included 106 images for FSR, HC, and SC, but these were based on 107 subjects.

- Six B readers classified each image on each subject (FSR, HC & SC) presented in random order using the 2000 version of the ILO system.

Image Format Reading Order

Reader	Round 1			Round 2		
	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

The reading order of images was also randomized within each cycle.

106 subjects x 6 readers x 3 formats x 2 rounds = 3816!

Standard Images

- All B readers used their own set of ILO standard films when reading FSR and hard copy DR images, in accordance with ILO guidelines.

- To allow for side-by-side comparison of soft copy DR images with ILO standard images, (with permission of the ILO) we digitized an entire set of ILO standard films. The digitized ILO standard images were supplied to each of the 6 B readers.

Work Station Requirements

- To insure that all B readers employed high quality display monitors for their readings of the SC images, a set of workstation requirements was developed and distributed to each B reader.

Data Recording & Checking

- B readers recorded their findings on a scannable version of the 2000 revision of the ILO form created by the National Institute for Occupational Safety and Health (NIOSH).

- B readers returned completed forms to the University of Michigan for scanning and error checking using a SAS program.

Methods – Statistical Analyses

- Analyses compared the inter-reader reliability of ratings for image quality, parenchymal and pleural abnormalities across image formats (using the multi-rater kappa statistic).

- Analyses investigated marginal rating differences across image formats (i.e., compared prevalence of findings) controlling for potential confounders such as age, gender, smoking and BMI. A mixed model approach (GEE) was adopted for statistical analyses to account for the clustering effect induced by multiple ratings made on the same subject using different image formats.

Subject Characteristics

	N (%)
Gender – Male	86 (80)
Ever Smoked	68 (64)
Current Smoking	10 (9)
History of Dust Exposure	60 (56)
Dust Exp Type Silica	34 (57)
Asbestos	25 (42)
Other/unknown	12 (20)

Note: Some subjects reported more than one type of dust exposure

DRAFT

Subject Characteristics

	Mean (s.d.)	Median (range)
Age (years)	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 - ever smoked (n=68)	30.7 (23.8)	23.5 (1-96)

Image Quality Results

Film Quality	Combined n=3816	Film n=1272	Hard Copy n=1272	Soft Copy n=1272	X^2 p-value
1	29%	31	24	34	
2	60	61	61	57	<0.0001
3	10	8	14	9	
4 (UR*)	1	0	1	0	

*UR = unreadable

Parenchymal Results

DRAFT

	Combined n=3794	Film n=1270	Hard Copy n=1254	Soft Copy n=1270	X^2 p-value
2A. Any Parenchymal Abnormalities					
No	32%	35	29	33	0.0025
Yes	68	65	71	67	
2Ba. Shape/Size of Primary Small Opacities*					
Round (p, q, r)	32	34	31	31	0.3958
Irregular (s, t, u)	68	66	69	69	
2Bc. Small Opacity Profusion					
0	40	43	36	42	
1	31	30	31	30	0.0111
2	22	21	25	22	
3	7	6	8	6	

*only includes images with 'yes' in 2A

Parenchymal Results

	Combined n=3794	Film n=1270	Hard Copy n=1254	Soft Copy n=1270	X^2 p-value
2C. Large Opacities					
0	85%	85	83	87	0.0284
A	6	6	6	6	
B	7	7	8	6	
C	2	2	3	1	
2C. Large Opacities					
No (0)	85	85	83	87	0.0106
Yes (A or B or C)	15	15	17	13	
2C. Large Opacities with 'ax'	80	80	77	82	0.02
No (0)	20	20	23	18	
Yes (A or B or C or ax)					

Pleural Results

	Combined n=3794	Film n=1270	Hard Copy n=1254	Soft Copy n=1270	X^2 p-value

3A. Pleural Abnormalities
    No
    Yes
	
	68%
	32
		
		59
		41
			
			69
			31
				
				73
				27
					
					<0.0001

3A. Pleural Abnormalities
    No (none or 3-5 mm = 'a')
    Yes (width > 5mm)
	
	92
	8
		
		90
		10
			
			92
			8
				
				95
				5
					
					<0.0001

3C. Costophrenic angle
    Obliteration
    No
    Yes (right and/or left)
	
	93
	7
		
		92
		8
			
			94
			6
				
				94
				6
					
					0.0413

3D. Diffuse Pleural
Thickening
    No
    Yes (right and/or left)
	
	95
	5
		
		94
		6
			
			96
			4
				
				96
				4
					
					0.1043

Reliability: Multi-rater Kappa Values for Inter-rater Agreement by Image Format

	Film	Hard Copy	Soft Copy
1.A: Film Quality (4-point ordinal scale)	**0.30^a**	**0.20^b**	**$0.23^{a,b}$**
1.A: Film Quality (Category 1 versus 2,3&4)	0.32	0.29	0.28
1.A: Film Quality (Categories 1&2 versus 3&4)	**0.38^a**	**$0.24^{a,b}$**	**0.16^b**
2.A: Any Parenchymal Abnormalities (yes/no)	0.62	0.56	0.65
2.B: Small Opacities (12-point scale)	0.29	0.24	0.27
2.B: Small Opacities (4-point scale)	0.48	0.45	0.46
2.C: Large Opacities (4-point scale)	0.48	0.53	0.48
2.C: Large Opacities (yes/no)	0.62	0.72	0.64
3.A: Pleural Abnormalities (yes/no)	0.46	0.50	0.48
3.C: Costophrenic Angle Obliteration (yes/no)	0.56	0.49	0.46
3.D: Diffuse Pleural Thickening (yes/no)	0.62	0.53	0.53

Unweighted multi-rater kappa values. Standard errors calculated using bootstrap percentile method; no adjustments for covariates. Different letters in **bold** indicate significant differences ($p < 0.05$).

Pairwise Comparisons of Prevalence by Image Format – Discrete Models

Classification comparison	Film versus Hard Copy*	Film versus Soft Copy	Hard versus Soft Copy
1.A: Film Quality (Cat 1 v 2,3,&4) adjusted	0.65 (0.46-0.91)	1.12 (0.84-1.49)	1.72 (1.43-2.08)
unadjusted	0.67 (0.49 -0.92)	1.11 (0.85-1.45)	1.66 (1.39-1.96)
1.A: Film Quality (Cat 1&2 v 3&4) adjusted	0.42 (0.24-0.71)	0.87 (0.50-1.54)	2.10 (1.63-2.70)
unadjusted	0.47 (0.31 -0.73)	0.89 (0.56-1.41)	1.87 (1.53-2.30)
2.A: Parenchymal Abnrmlts (yes/no)adjusted	0.72 (0.60-0.86)	0.90 (0.78-1.04)	1.26 (1.09-1.46)
unadjusted	0.75 (0.65-0.86)	0.91 (0.80-1.04)	1.22 (1.09-1.35)
2.C: Large Opacities (yes/no) adjusted	0.83 (0.70-0.99)	1.23 (1.04-1.46)	1.48 (1.24-1.76)
unadjusted	0.86 (0.75-0.98)	1.18 (1.03-1.36)	1.38 (1.20-1.58)
2.C: Large Opacities & 'ax' (yes/no) adjusted	0.79 (0.66-0.94)	1.12 (0.99-1.27)	1.43 (1.22-1.67)
unadjusted	0.83 (0.74-0.93)	1.07 (0.98-1.17)	1.29 (1.16-1.44)

Estimates of odds ratios using Generalized Estimating Equations (95% confidence interval); All adjusted estimates are adjusted for age, gender, body mass index, pack-years of smoking, round, and individual B-readers. Models other than film quality are also adjusted for median image quality.

Pairwise Comparisons of Prevalence by Image Format – Discrete Models

Classification comparison	Film versus Hard Copy*	Film versus Soft Copy	Hard versus Soft Copy
3.A: Pleural Abnml (yes/no) adjusted	1.28 (1.08-1.53)	1.59 (1.35-1.88)	1.24 (1.08-1.42)
unadjusted	1.30 (1.10-1.53)	1.53 (1.31-1.78)	1.18 (1.04-1.33)
3.A: Pleura width ($\leq$ 5mm v > 5mm)adjusted	1.49 (1.09-2.03)	2.20 (1.59-3.02)	1.47 (1.14-1.91)
unadjusted	1.25 (0.96-1.63)	1.96 (1.38-2.79)	1.57 (1.22-2.01)
3.C: Costo Angle Oblit (yes/no) adjusted	1.41 (0.99-2.00)	1.39 (0.98-1.97)	0.98 (0.80-1.22)
unadjusted	1.45 (0.99-2.11)	1.36 (0.93-1.99)	0.94 (0.79-1.12)
3.D: Diffuse Pleural Thick (yes/no)adjusted	1.32 (0.97-1.80)	1.43 (1.04-1.98)	1.08 (0.84-1.40)
unadjusted	1.35 (0.94-1.95)	1.45 (0.99-2.12)	1.07 (0.84-1.37)

Estimates of odds ratios using Generalized Estimating Equations (95% confidence interval); All adjusted estimates are adjusted for age, gender, body mass index, pack-years of smoking, round, and individual B-readers. Models other than film quality are also adjusted for median image quality.

Pairwise Comparisons of Prevalence by Image Format – Continuous Models

Classification comparison		Film versus Hard Copy	Film versus Soft Copy	Hard versus Soft Copy
1.A: Film Quality (4-pt scale)	adjusted	-0.166 (0.0002)	0.013 (0.7379)	0.179 (<0.0001)
	unadjusted	-0.166 (0.0002)	0.013 (0.7366)	0.179 (<.0001)
2.B: Small Ops (12-pt scale)	adjusted	-0.419 (<0.0001)	-0.026 (0.6871)	0.393 (<0.0001)
	unadjusted	-0.381 (<.0001)	-0.028 (0.6771)	0.353 (<.0001)
2.B: Small Ops (4-pt scale)	adjusted	-0.148 (<0.0001)	-0.014 (0.5642)	0.134 (<0.0001)
	unadjusted	-0.136 (<0.0001)	-0.015 (0.5596)	0.122 (<.0001)
2.C: Large Ops (4-pt scale)	adjusted	-0.058 (0.0093)	0.041 (0.0078)	0.099 (<0.0001)
	unadjusted	-0.051 (0.0142)	0.041 (0.0098)	0.092 (<0.0001)

Estimates of differences of least square means using Generalized Estimating Equations (p-values); All adjusted estimates are adjusted for age, gender, body mass index, pack-years of smoking, round, and individual B-readers. Models other than film quality are also adjusted for median image quality.

Conclusions: Reliability

- There were few significant differences in the reliability (i.e., inter-rater agreement) of image classifications across formats, and these differences were solely among classifications of image quality.

Conclusions: Prevalence of Findings

- Parameter estimates for image format in adjusted models were similar to results for unadjusted models, which indicates that the covariates (i.e., age, gender, BMI, pack years, round, median image quality, and individual readers) were not acting as confounders of the effect of image format on prevalence of findings.

Conclusions: Prevalence of Findings

- <u>Film Quality</u>: Classifications for FSR and SC images did not differ significantly. HC images had a tendency to be classified significantly worse than FSR and SC images.

- <u>Parenchymal abnormalities & small opacity profusion</u>: Classifications of FSR and SC images did not differ significantly. Classifications of HC images demonstrated significantly greater prevalence of parenchymal abnormalities and small opacity profusion compared to FSR and SC images.

Conclusions: Prevalence of Findings

- <u>Large Opacities</u>: All image formats differed significantly with HC>FSR>SC for both discrete and continuous models

- However, when 'ax' were included as 'large opacities', the significant difference between FSR and SC disappeared in the discrete model

Conclusions: Prevalence of Findings

- <u>Presence of Pleural Abnormalities</u>: All image formats differed significantly with FSR>HC>SC.

  - Consideration of width of in-profile pleural thickening (i.e., $\leq$ 5mm vs. > 5mm) did not eliminate these differences among image formats (in fact, the odds ratios became somewhat larger)

Conclusions: Prevalence of Findings

- <u>Costophrenic Angle Obliteration and Diffuse Pleural Thickening</u>: With one exception (the adjusted model for FSR versus SC), there were no significant differences among the image formats (though the prevalence of these outcomes was low, and therefore lacked power)

Study Strengths

- Large number of subjects and B readers
- Good power for key outcomes – e.g., the study had 90% power to detect a difference between image formats of 0.17 units on the full 12-point ILO scale for profusion of small opacities
- Balanced mix of 'irregular' and 'round' small opacities
- Well-defined methods and up-to-date equipment
- Assessed hard copy and soft copy digital images
- Used digital version of ILO standard images
- Modeled prevalence of findings with adjustment for multiple covariates

Study Limitations

- These analyses did not employ an external 'gold standard', so when, for example, there was a difference in prevalence by image format it was not possible to determine which was closer to the 'truth'.

- Alternate digital image processing parameters were not assessed

Acknowledgments

DRAFT