

---

# Interobserver Variability of Interpretation of Chest Roentgenograms

Peri ARBAK\*, Öznur AKKOCA\*, Füsün ÜLGER\*, Özgür KARACAN\*, Akın KAYA\*, Uğur GÖNÜLLÜ\*

\* Department of Pulmonary Diseases Faculty of Medicine Ankara University, ANKARA

## ABSTRACT

Interobserver variability of pulmonary radiographic findings has been evaluated in many studies. Aim of this study is to evaluate the effect of education and experience on radiographic observation. Study population consisted of 100 ambulatory patients and their postero-anterior chest roentgenograms were evaluated by three reader (assistant, specialist, professor). Radiographic assesment of reader III was golden standart. Agreement rates were calculated. Kappa statistics was used. Reader II had higher agreement rate and kappa values than reader I with the expectation of position and hilus evaluation. This study showed the importance of education and experience.

**Key Words:** Radiographic interpretation, agreement rate.

## ÖZET

### Radyografi Yorumunda Gözlemciler Arası Farklılık

Radyografi yorumunda okuyucular arasındaki farklılık birçok çalışmada vurgulanmıştır. Bu çalışmanın amacı radyografi yorumunda eğitim ve deneyimin etkilerini değerlendirmektir. Kliniğimizde ayaktan izlenen 100 hastanın postero-anterior (PA) akciğer grafipleri 3 farklı okuyucu tarafından değerlendirildi (asistan, uzman, profesör). Profesör olan üçüncü okuyucu altın standart olarak kabul edildi. Görüş birliği hızları değerlendirildi ve kappa istatistik kullanıldı. Uzman olan ikinci okuyucu pozisyon ve hilusun değerlendirilmesi dışında asistan okuyucuya göre daha yüksek görüş birliği hızına sahip bulundu. Bu çalışma radyografi yorumunda eğitim ve deneyimin önemini vurgulamış oldu.

**Anahtar Kelimeler:** Radyografi yorumu, görüş birliği hızı.

Interobserver variability of interpretation of pulmonary radiographic findings has been evaluated in many studies. Earlier studies have used interobserver agreement rate of the chest radiograph in the diagnosis of certain pulmonary diseases, such as tuberculosis and pneumoconiosis or pneumonia in pediatric patients (1-3). In recent studies, agreement rate was evaluated between different clinical settings or different medical stages in some medical subject, such as student, resident or professor (4).

Aim of this study is to evaluate the interobserver variability of radiographic interpretation and to investigate the effect of education and experience on radiographic diagnosis.

## MATERIALS and METHODS

The study population consisted of 100 ambulatory patients who were admitted to the outpatient clinic in our hospital between October and December in 1996.

Chest radiographs of these 100 patients were evaluated by reader I, II, III who were assistant, specialist and professor on pulmonary diseases, respectively. Diagnosis was not known by observer at the beginning and performed by clinical, radiological findings, laboratory determinants and invasive investigations, if necessary. Radiographic assesment of reader III was golden standart. Agreement rate of reader I and II according to reader III was calculated and kappa statistics was used to interpretate the significance of interobserver agreement rate. A kappa greater than. 0.75 indicates excellent agreement, while a kappa of 0.40-0.75 and less than 0.40 indicates fair to good and poor agreement, respectively.

### RESULTS

The mean age of 100 patients (58 male, 42 female) was  $56 \pm 16$  years. Disease of patients are shown in Table 1. Chronic obstructive pulmonary disease (COPD) was the most common disease (43%). Other common diseases were pneumonia (14%) and upper airway infection (13%). Agreement rate between reader I and III for interpretation of technical and positional status, pleura, diaphragm, mediastinum, hilum and osseous parts is given in Table 2 (kappa statistics also shown). Poor agreement rate was identified between reader II and III for identification quality of graphy, positional status, hilum; while fair to good agreement was observed for evaluation of pleura, mediastinum and osseous parts (Table 3). Reader II had higher agreement rate and kappa values than reader 1 with the exception of position and hilum evaluation. Results of detection of paranchymal lesions by reader I and III are shown in Table 4. There was excellent agreement for determination of mass lesions while fair to good agreement was determined for detection of consolidation, interstitial patterns and cavitation. Agreement rates for interpretation of paranchymal lesions by readers II and III are shown in Table 5. Seven chest radiographs were determined as normal by reader I, reader II and III diagnosed 10 and 24 chest radiographs as normal, respectively.

**Table 1. Distribution of patients according to underlying diseases.**

	Number of patients (n= 100)
COPD	43
Pneumonia	14
Malignancy*	1
Interstitial lung disease**	1
Upper respiratory tract disease	13
Tuberculosis	1
Cor pulmonale	4
Bronchiectasis	3
Asthma	4
Tuberculous pleurisy	1
Normal	5
Others***	10

\* Primary, secondary lung malignancy

\*\* Sarcoidosis, connective tissue disease

\*\*\* Pulmonary thromboembolism, sleep apnea, aspergilloma, congestive heart failure, lung abscess.

### DISCUSSION

Radiographic interpretation variability has been investigated in many studies. While earlier studies evaluated interobserver reliability by radiologist, recent studies compared the radiographic diagnosis of medical students, specialists and practitioners (4). Radiographies of tuberculosis and pneumoconiosis were evaluated about interpretation differences, afterwards pneumonia was in interest (5,6). Present study didn't focus on any disease, chest radiographs of 100 patients who were admitted to our outpatient clinic were evaluated. Many studies about interobserver variability has considered the judgement of radiologist as golden standart, but we have chosen interpretation of reader III (university staff on pulmonology) as golden standart, since he has educational responsibility. Other two readers were an assistant and specialist, so we have aimed to outline the effect of education and experience on radiographic interpretation.

In our study, reader II was found to have higher kappa values than reader I in observation of technical, positional status, pleura, mediastinum, hilum and osseous parts. Herman et al re-

**Table 2. Agreement rates and kappa values of reader I and III (Interpretation of chest radiograph).**

	Reader I	Reader III	Agreement rate (%)	Kappa values
Film quality				
Normal	51	76	59	0.17
Abnormal	49	24		
Position of radiograph				
Normal	64	78	74	0.38
Abnormal	36	22		
Pleura-Diaphg.				
Normal	50	63	73	0.46
Abnormal	50	47		
Mediastinum				
Normal	52	78	68	0.34
Abnormal	48	22		
Hilum				
Normal	59	75	74	0.42
Abnormal	41	25		
Osseous Parts				
Normal	98	99	99	0.66
Abnormal	2	1		

**Table 3. Agreement rates and kappa values of readers II and III (Interpretation of chest radiograph).**

	Reader II	Reader III	Agreement rate (%)	Kappa values
Film quality				
Normal	75	76	75	0.32
Abnormal	25	24		
Position of radiograph				
Normal	78	78	76	0.30
Abnormal	22	22		
Pleura-Diaphg.				
Normal	61	63	76	0.49
Abnormal	39	37		
Mediastinum				
Normal	60	78	76	0.45
Abnormal	40	22		
Hilum				
Normal	61	75	70	0.32
Abnormal	39	25		
Osseous Parts				
Normal	98	99	99	0.66
Abnormal	2	1		

**Table 4. Agreement rates and kappa values of reader I and III on interpretation of paranchyma.**

	Reader I	Reader III	Agreement rate (%)	Kappa values
Consolidation				
Yes	26	16	76	0.28
No	74	84		
Interstitial pattern				
Yes	29	26	75	0.37
No	71	74		
Mass				
Yes	0	0	100	-
No	100	100		
Cyst-Cavity				
Yes	8	1	93	0.20
No	92	99		
Hyperlucency				
Yes	18	11	89	0.56
No	82	89		
Fibrosis				
Yes	5	6	93	0.32
No	95	94		

**Table 5. Agreement rates and kappa values of readers II and III of paranchyma.**

	Reader II	Reader III	Agreement rate (%)	Kappa values
Consolidation				
Yes	21	16	83	0.43
No	79	84		
Interstitial pattern				
Yes	36	26	58	0.02
No	64	74		
Mass				
Yes	0	0	100	-
No	100	100		
Cyst-Cavity				
Yes	2	1	99	0.66
No	98	99		
Hyperlucency				
Yes	22	11	83	0.39
No	78	89		
Fibrosis				
Yes	2	6	94	0.22
No	98	94		

ported that the highest error rate of interpretation was seen on hilum and nodulary lesions (7). In our study also, there was poor agreement between Reader II and III on interpretation of hilum. Poor agreement rate was related to insufficient contrast gradient.

Melbye and Dale found that, kappa agreement rate was 0.50 between radiology assistant and radiology staff, while it was 0.72 between chest physician and radiology jury on interpretation of infiltrations (5). Other studies showed that observation of infiltrations was the most common radiological pattern of disagreement (7,8). In our study, reader II had fair to good agreement rate with reader III on diagnosis of infiltrations, while reader I had poor agreement with reader III. Some studies reported that normal bronchovascular markings were defined as interstitial pattern, erroneously in making diagnosis of pneumoconiosis. It was reported that in the earlier years of medical education, there was lower agreement rates for radiographic evaluation (5,6). In our study, also, reader I had lower agreement rates than reader II. Disagreement between readers is related to some factors, such as technical aspects, education and experience, perceptual and judgemental abilities of readers. Interobserver variability can be reduced by concentrating educational programmes on radiologists. Disagreements have limited detrimental effects on patient management. In medical practision symptomatology, clinical state and laboratory findings, rather than isolated radiological observation, are considered together (9,10). This approach together with radiological education will help to decrease the extent of disagreements. Interobserver consistency is especially important in diagnosis of pneumoconiosis, since

it depends on mainly plain chest roentgenograms. This model of study may be planned for also pneumoconiosis.

## REFERENCES

1. Reger RB, Morgan WKC. On the factors influencing consistency in the radiologic diagnosis of pneumoconiosis. *Am Rev Resp Dis* 1970; 102: 905-15.
2. Cochrane AL, Camb MB, Garland LH. Observer error in the interpretation of chest films. *Lancet* 1952; 2: 505-9.
3. Felton B, Morgan WK, Bristol LJ, et al. Observations on the results of multiple readings of chest films in coal miners pneumoconiosis. *Radiology* 1973; 109: 19-23.
4. Young M, Marrie TJ. Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia. *Arch Intern Med* 1994; 154: 12-26.
5. Melbye H, Dale K. Interobserver variability in the radiographic diagnosis of adult outpatient pneumonia. *Acta Radiol* 1992; 33: 79-81.
6. Albaum MN, Hill LC, Murphy M, et al. Interobserver reliability of the chest radiograph in community-acquired pneumonia. *Chest* 1996; 110: 343-50.
7. Herman PG, Gerson DE, Hessel SJ, et al. Disagreements in chest roentgen interpretation. *Chest* 1975; 68: 278-82.
8. Yerushalmy J. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. *Radiol Clin North Am* 1969; 7: 381-92.
9. Tew J, Calenoff L, Berlin BS. Bacterial or nonbacterial pneumonia accuracy of radiographic diagnosis. *Radiology* 1977; 124: 607-12.
10. Wijnands WJA. Diagnosis and interventions in lower respiratory tract infections. *Am J Med* 1992; 92: 915-75.

## Address for Correspondence:

Peri ARBAK, MD  
Department of Pulmonary Diseases  
Medical Faculty, Ankara University  
Dikimevi, ANKARA