British Heart Journal, 1974, 36, 280–290.

Observer variation in radiological assessment of pulmonary vasculature

S. Weitzman, W. A. Pocock, D. M. Hawkins,[1] and J. B. Barlow

From the Cardio-Vascular Research Unit, Department of Medicine, University of the Witwatersrand, and the Cardiac Clinic, General Hospital, Johannesburg, South Africa

Observer error in the interpretation of the pulmonary vasculature was studied: 10 experts commented on the presence of pulmonary venous and arterial hypertension, plethora, oligaemia, thromboembolic disease, and the size of the pulmonary artery segment and aorta in 100 pairs (posteroanterior and lateral views) of chest radiographs. Apart from the age of the patient, no clinical data were provided. The observers, comprising 5 radiologists, 4 cardiologists, and 1 thoracic surgeon, viewed each pair of radiographs on 2 separate occasions. The 2000 assessments, each of which included 7 categories, were analysed with the aid of a computer to determine the inter- and intra-individual variation and an attempt was made to assess the reliability of the diagnostic criteria.

The greatest accuracy in diagnosis was shown in the category of thromboembolic disease, followed by that of pulmonary venous hypertension, whereas the assessment of pulmonary arterial hypertension, particularly when associated with the Eisenmenger syndrome, was the poorest. The diagnosis of oligaemia was twice as accurate when a right-to-left shunt was present than when this was absent. The radiographic signs of pulmonary thromboembolic disease are highly diagnostic of the condition but are not always recognized on the plain chest film. The percentage overdiagnosis in all categories ranged from 5·6 to 21·1 per cent (mean 9·6%) and that of underdiagnosis from 8·4 to 17·6 per cent (mean 12·9%). With the exception of 2 of them, all observers were biased towards underdiagnosis. There was some correlation between the accuracy and consistency of each observer. The 5 clinicians were more accurate in all categories than the radiologists, though the difference was not always statistically significant.

Accuracy of diagnosis did not correlate well with the age of the patient, the diagnosis, or the severity of the disease.

In 1949, Garland, discussing the scientific evaluation of diagnostic procedures, stated that in every area of activity there was a significant observer variation among experts in that area and that this could be tested. He concluded, 'the first reaction of every new observer is "This does not happen to me". To these we suggest "Try it systematically".'

Experts in various disciplines have since accepted Garland's challenge; the observer variation in the interpretation of electrocardiograms (Davies, 1958; Seymour and Conway, 1969), the recording of peripheral pulses (Meade *et al.*, 1968) and of arterial blood pressure (Anderson and Cowan, 1961), the clinical detection of cyanosis (Comroe and Botelho, 1947), and the radiographic assessment of pulmonary tuberculosis (Yerushalmy *et al.*, 1951; Cochrane and Garland, 1952; Whyte, 1955) and of pneumoconiosis (Shiels and Thomas, 1954) have all been evaluated. Almost all studies have shown significant 'interindividual' and 'intraindividual' variation among the experts in the different disciplines.

To our knowledge, the radiological assessment of pulmonary vasculature, in terms of pulmonary venous and arterial hypertension, pulmonary oligaemia and plethora, and thromboembolic disease, has not been subjected to the test of observer variation. In this cardiac clinic, differences of opinion among experienced observers, including both cardiologists and radiologists, are not infrequent and the clinical data have usually influenced the final radiographic

Received 18 June 1973.

[1] Department of Statistics, University of the Witwatersrand, Johannesburg, South Africa.

assessment. This study was, therefore, devised in order to determine the inter- and intraindividual variation among experts in this field and, in addition, to assess the reliability of the diagnostic criteria.

Subjects and methods

Ten experts, comprising 5 radiologists (Observers 1–5), 4 cardiologists (one of whom was J.B.B.), and 1 thoracic surgeon (Observers 6–10), were asked to interpret a posteroanterior and lateral chest *x*-ray of 100 subjects at 2 separate viewings. Each pair of films was numbered and the numbers and orders were changed for the second viewing. The interval between the first and second viewing ranged from 5 to 14 days so that the study was completed within 4 weeks. The number of films viewed per session and the time spent on each film were determined by the respective individual, thus obviating observer fatigue. Each observer sat alone during the sessions and recorded his assessment of each pair of films on the form provided (Table 1). No clinical data, apart from the age of the patient, were provided. A minor part of the study was the evaluation of the size of the pulmonary artery and aorta. The 'small aorta' category was excluded as it was considered that this could not be adequately assessed on the 2 films provided, and the observers were requested to record the aorta as 'normal' if it was not visible or appeared small. The observers were asked to comment on the quality of the radiographs. The data, comprising 2,000 individual assessments, were recorded on punched cards. The cards were then analysed by means of a FORTRAN programme on an IBM system 360/50 computer.

The one hundred pairs of radiographs were selected from patients attending this Cardiac Clinic. The ages of the subjects ranged from 1 month to 66 years. Fifty-seven were below 20 years of age, 13 of whom were younger than 5 years. The diagnosis was confirmed by cardiac catheterization in 78 patients. The time intervals between the catheterization and the radiographs varied from 1 day to 13 years. Thirty-three films were taken within 1 month and a further 30 within 6 months of the procedure. In all patients in whom the catheter data were used, we (S.W. and W.A.P.) were confident, on clinical and electrocardiographic assessment, that no significant change had occurred between the time of the radiographs and the cardiac investigation. The 22 patients whose films were included without catheter confirmation were considered to have unequivocal evidence clinically, on electrocardiogram and at operation (where this was undertaken), of the accepted radiological diagnosis.

The subjects were divided into categories as follows.

Pulmonary plethora

In addition to 5 patients with an atrial septal defect and associated thromboembolic disease, there were 23 others with a left-to-right shunt of varying magnitude. Five had a persistent ductus arteriosus, 8 a ventricular septal defect, 7 an atrial septal defect, 2 had both a persistent ductus arteriosus and a ventricular septal defect, and 1 had left-to-right shunts at all 3 levels. Catheter data were available in all except 2 children, both of whom had a small shunt through a persistent ductus arteriosus, diagnosed clinically and confirmed at operation. Eight in this group were below 5 years of age.

For purposes of this study, pulmonary arterial hypertension, in association with plethora, was judged to be present if the pulmonary systolic arterial pressure was greater than 45 mmHg. Of the 23 patients, 11 fell into this category and, apart from 2 with peak pressures of 45 mmHg, the pressures ranged from 60 to 94 mmHg. The shunt ratio in this group ranged from 2 to 8:1.

The maximum pulmonary arterial systolic pressure in the 12 patients with plethora and without hypertension was 38 mmHg. The shunt ratio ranged from 1·4 to

TABLE 1 *Form on which radiographic assessment was recorded*

Date of viewing	*Viewing No.*	*X-Ray No.*		*Observer No.*	
	Normal	*Possibly prominent*	*Prominent*		
Aorta					
1) Ascending	1	2	3		
2) Arch	1	2	3		
3) *Pulmonary artery segment*	*Small*	*Possibly small*	*Normal*	*Possibly prominent*	*Prominent*
	1	2	3	4	5
4) *Lung fields*	*Oligaemia*	*Doubtful oligaemia*	*Normal*	*Doubtful plethora*	*Plethora*
	1	2	3	4	5
5) *Pulmonary venous hypertension*	*Absent*	*Doubtful*	*Present*		
	1	2	3		
6) *Pulmonary arterial hypertension*	*Absent*	*Doubtful*	*Present*		
	1	2	3		
7) *Thromboembolic disease*	*Absent*	*Doubtful*	*Present*		
	1	2	3		
Additional comments					
X-ray quality	*Good*	*Fair*	*Poor*		

7·8:1 and in only 3 patients was it less than 2:1. Six of these patients had an atrial septal defect.

Pulmonary oligaemia

Fourteen patients with lesions likely to produce pulmonary oligaemia were included. Their ages ranged from 1 month to 34 years. Eight patients had isolated pulmonary stenosis with gradients across the pulmonary valve ranging from 40 to 200 mmHg. Of these, 2 were shunting right to left through a patent foramen ovale. There were 2 patients with Fallot's tetralogy, 3 with Ebstein's anomaly, 1 of whom had a right-to-left shunt at atrial level, and 1 infant had tricuspid atresia.

Pulmonary venous hypertension

There were 24 patients with pulmonary venous hypertension, including one with associated thromboembolic disease. With one exception, a patient with severe mitral regurgitation, all had mitral stenosis as the dominant lesion. Fourteen of these had associated pulmonary arterial hypertension, as judged by evidence of right ventricular hypertrophy clinically and electrocardiographically and confirmed at cardiac catheterization in 13 (pulmonary arterial systolic pressures ranged from 50 to 145 mmHg). The remaining uncatheterized patient, a 10-year-old girl, had electrocardiographic evidence of pronounced right ventricular hypertrophy and 'strain', and at operation the pulmonary artery was large and tense. The mean left atrial pressure in the 13 patients subjected to catheterization ranged from 15 to 45 mmHg.

Only 4 of the 10 patients with tight mitral stenosis and no clinical or electrocardiographic evidence of right ventricular hypertrophy were subjected to cardiac catheterization, and in all 4 the pulmonary arterial pressure was below 40 mmHg. The 10 patients were classed as having pulmonary venous hypertension only.

Tight mitral stenosis was confirmed at valvotomy in 14 of the 19 patients subjected to operation, in that the orifice was narrowed to 1·5 cm or less. In the remaining 5 with orifice sizes ranging from 1·8 to 3 cm, the leaflets were calcified and rigid with resultant functional tight stenosis.

Pulmonary arterial hypertension

This category comprised 9 patients with the Eisenmenger syndrome (Wood, 1956), 5 of whom had a persistent ductus arteriosus, 2 a ventricular septal defect, and 1 had both lesions. The remaining uncatheterized patient had unequivocal clinical and electrocardiographic evidence of the syndrome and probably had a persistent ductus arteriosus.

Pulmonary thromboembolic disease

Twelve patients had chronic pulmonary thromboembolic disease which was confirmed on pulmonary arteriography in 11 and by lung scan in the remaining patient. The pulmonary arterial pressure was above 40 mmHg in all except one patient, who was none the less included in the pulmonary arterial hypertension group since her pulmonary arterial systolic pressure rose from 40 to 52 mmHg after angiography. Five of the 12 patients had an associated secundum atrial septal defect and all were shunting left to right, with ratios from 1·5 to 3·8:1. One patient had associated tight mitral stenosis.

Normal vasculature

Eleven patients had normal pulmonary vasculature. Two had minimal mitral regurgitation, 1 had mild infundibular pulmonary stenosis (gradient 20 mmHg), 3 had idiopathic dilatation of the pulmonary artery, and 1 had the 'straight back' syndrome. The remaining 4 patients had innocent pulmonary ejection or vibratory systolic murmurs.

Prominent aorta and pulmonary artery segment

Eight patients with prominence of the ascending aorta were included in the study. All had aortic stenosis, aortic incompetence, or a mixed lesion, and prominence of the aorta was confirmed by aortography or at operation. In 3 the aortic arch was also considered to be prominent. Two patients in the pulmonary venous hypertension category had prominence of the ascending aorta caused by associated aortic valve lesions. Among the remaining radiographs there were 15 which were thought by ourselves (S.W. and W.A.P.) to show either definite or probable prominence of the aortic arch (patients with atherosclerosis, hypertension, and persistent ductus arteriosus). Since there were no objective criteria for this group the results in this category were not analysed in detail.

Similarly, the assessment of the size of the pulmonary artery segment by ourselves (S.W. and W.A.P.) was somewhat arbitrary and was influenced by the diagnosis, age of the patient, and the catheter data. This category was therefore also not analysed in detail.

Results

The results were analysed by computer according to the errors in diagnosis rather than to the correct diagnoses. Errors were either of underdiagnosis, that is failure to diagnose the presence of plethora, pulmonary venous or arterial hypertension and thromboembolic disease, or of overdiagnosis in which these features were scored as present when they were, in fact, absent. Each assessment of a radiograph was marked by the computer in relation to a 'master card'. Any scoring by the observer to the left of the correct assessment was marked by the computer as underdiagnosis, whereas scoring to the right of the correct answer was scored as overdiagnosis. Since oligaemia and plethora were assessed in one category under 'lung fields' (Table 1), failure to recognize oligaemia (i.e. radiographs called normal or plethoric) was scored as overdiagnosis by the computer, whereas incorrect diagnosis of oligaemia when it was absent was scored as underdiagnosis.

TABLE 2 *Number and distribution of errors in each category*

Category	*Underdiagnosis*								*Correct*		*Overdiagnosis*							
	−4		*−3*		*−2*		*−1*		*0*		*+1*		*+2*		*+3*		*+4*	
	No.	%	*No.*	%	*No.*	%	*No.*	%	*No.*	%	*No.*	%	*No.*	%	*No.*	%	*No.*	%
1) Ascending aorta					81	4·1	36	1·8	1631	81·6	188	9·4	64	3·2				
2) Aortic arch					85	4·3	138	6·9	1521	76·1	183	9·2	73	3·7				
3) Pulmonary artery	6	0·3	2	0·1	86	4·3	276	13·8	1471	73·6	121	6·1	37	1·9	1	0·1		
4) Oligaemia plethora	2	0·1	9	0·5	171	8·6	115	5·8	1361	68·1	134	6·7	190	9·5	15	0·8	3	0·2
5) Pulmonary venous hypertension					69	3·5	59	3·0	1764	88·2	64	3·2	44	2·2				
6) Pulmonary arterial hypertension					350	17·5	175	8·8	1301	65·1	98	4·9	76	3·8				
7) Thromboembolic disease					84	4·2	58	2·9	1808	90·4	35	1·8	15	0·8				

Percentages corrected to the first decimal point.

Error frequency

Table 2 shows the total number and distribution of errors in each category. The total number of possible correct diagnoses in each category is 2,000 (10 observers, 100 radiographs, 2 viewings). The number and percentage of correct diagnoses are shown in the centre of Table 2 with the negative errors (errors of underdiagnosis) on the left and the errors of overdiagnosis on the right. It must be remembered (Tables 1 and 2) that categories 3) and 4) have 5 possible diagnoses, and observers may therefore be as much as 4 classes out in either direction. For example, plethora was diagnosed instead of oligaemia a total of 3 times (0·2%) and oligaemia instead of plethora twice (0·1%), while a prominent pulmonary artery segment was called 'small' 6 times (0·3%). The highest percentage of correct diagnoses is in the thromboembolic category (90·4%). The assessment of pulmonary venous hypertension is also highly accurate (88·2%), while the least accuracy is in the category of pulmonary arterial hypertension (only 65·1% correct), with the majority of errors being those of underdiagnosis and as many as 17·5% being 2 classes under.

Observer accuracy and bias

The total number of errors of each observer of a possible 1400 (7 categories, 100 radiographs, 2 viewings) was determined and is expressed as a percentage of over and underdiagnosis (Fig. 1). With the exception of Observer 1, whose incidence of overdiagnosis was much greater than that of his colleagues, the overall accuracy is similar with the greatest accuracy in diagnosis shown by Observer 10, closely followed by Observer 2. The percentage overreading of the 10 observers ranges from 5·6 to 21·1 per cent (mean 9·6%) and the percentage underreading from 8·4 to 17·6 per cent (mean 12·9%). If the results of Observer 1 are excluded, the incidence of overdiagnosis of the remaining 9 ranges from 5·6 to 11·9 per cent (mean 8·3%) and that of underdiagnosis from 8·4 to 17·6 per cent (mean 12·7%).

The percentage over- and underdiagnosis of the observers in each of the 7 categories is shown in Fig. 2. Thus, Observer 2 shows the greatest accuracy in the diagnosis of pulmonary venous hypertension, with a total error of only 2 per cent, while Observer 10 has the fewest errors in the assessment of pulmonary arterial hypertension and thromboembolic disease. The overall accuracy of the 10 observers in each category is shown in Fig. 2 and confirms (Table 2) that the greatest number of errors occurs in the assessment of pulmonary arterial hypertension.

Observer bias reflects the tendency of each person to over- or underdiagnose. It can be seen (Fig. 2) that Observer 1 tends to overdiagnosis, particularly in the assessment of aortic prominence, while Observer 9 has a negative bias in all categories. The overall bias for each category is also represented in

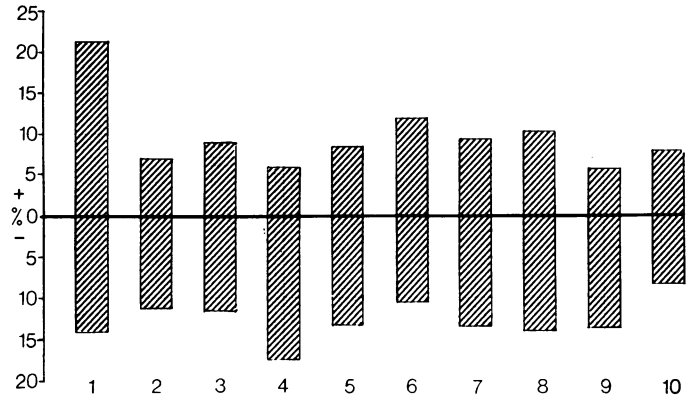

FIG. 1 *Diagnostic errors of each observer of over (+) and under (−) diagnosis, expressed as a percentage. The majority of observers have a bias towards underdiagnosis. Observers 1–5 are radiologists.*

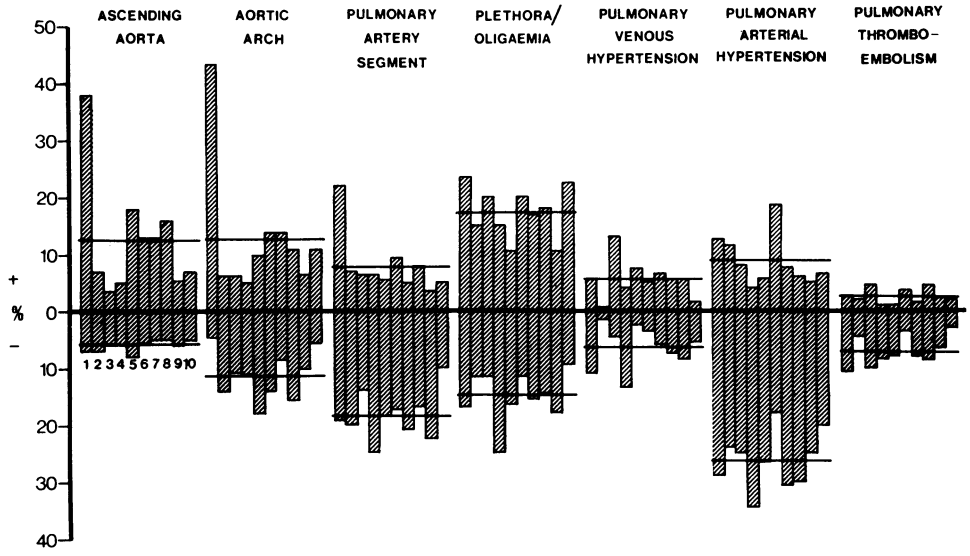

FIG. 2 *Percentage of over (+) and under (−) diagnosis of each observer in the different categories. The mean of all observers for each category is indicated by the horizontal lines.*

Fig. 2. As might be expected, pulmonary arterial hypertension shows the greatest negative bias, while thromboembolic disease, pulmonary venous hypertension, and prominence of the pulmonary artery segment are less frequently underdiagnosed, and aortic prominence and plethora are overdiagnosed.

Consistency

The consistency, or intraindividual variation, of each observer is indicated by the mean absolute deviation (Table 3). This is the average over the 100 radiographs of the number of classes by which the assessments in the 2 viewings differed. If the 2 diagnoses are identical (whether correct or not) this difference is zero. Thus a high degree of consistency between the pairs of diagnoses is reflected in a low value for the mean absolute deviation. It may be seen (Table 3) that the observer consistency varies conspicuously in each category but that Observers 1 and 8 vary their diagnosis by an average of one class in every 3 viewings, Observers 3, 6, 7, and 10 by one class in every 4 viewings, and Observers 2, 4, 5, and 9 by one class in every 5 viewings.

More important than the individual observer

TABLE 3 *Mean absolute deviation*

Observer	*Ascending aorta*	*Aortic arch*	*Pulmonary artery segment*	*Oligaemia plethora*	*Pulmonary venous hypertension*	*Pulmonary arterial hypertension*	*Thrombo-embolic disease*	*Average each observer*
1	0·380	0·370	0·330	0·390	0·220	0·400	0·120	0·316
2	0·130	0·180	0·280	0·210	0·060	0·370	0·070	0·186
3	0·030	0·190	0·240	0·360	0·250	0·380	0·170	0·231
4	0·090	0·150	0·230	0·270	0·220	0·190	0·100	0·179
5	0·160	0·150	0·250	0·180	0·160	0·300	0·060	0·180
6	0·230	0·260	0·170	0·280	0·140	0·400	0·070	0·221
7	0·170	0·280	0·160	0·380	0·160	0·420	0·110	0·240
8	0·280	0·370	0·310	0·400	0·190	0·400	0·160	0·301
9	0·080	0·110	0·250	0·170	0·160	0·240	0·120	0·161
10	0·210	0·190	0·240	0·250	0·190	0·450	0·070	0·230
Average all observers	0·176	0·225	0·246	0·289	0·175	0·355	0·105	0·224

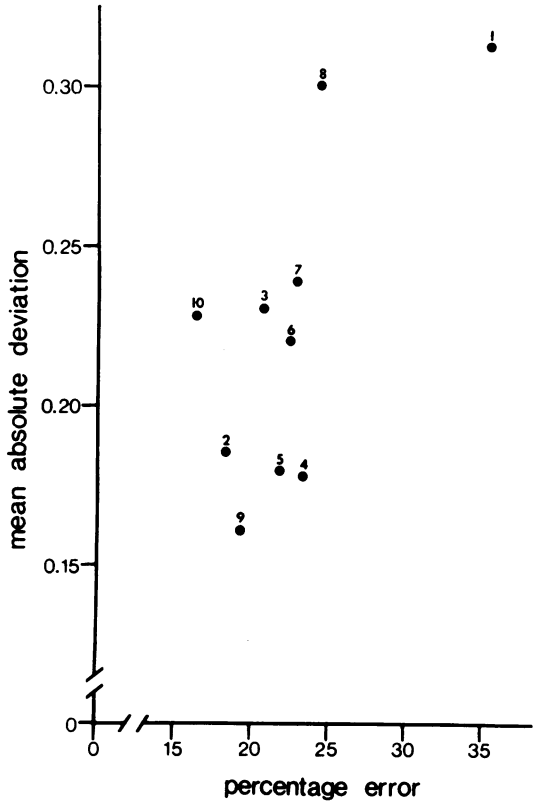

FIG. 3 *Comparison of the overall consistency (mean absolute deviation) and accuracy (percentage error) of the 10 observers.*

consistency is the overall consistency for each category, which is greatest in thromboembolic disease (0·105, or one class in every 10 viewings) and least in pulmonary arterial hypertension (0·355, or one class in every 3 viewings). Comparison of the consistency and accuracy of each observer is shown in Fig. 3. Observer 9, though the most consistent, is not the most accurate, whereas Observer 10 exhibits the greatest accuracy but is less consistent.

Comparison of radiologists and cardiologists

Comparison of the accuracy and bias in each category between the radiologists and cardiologists (including the thoracic surgeon) is shown in Fig. 4. The cardiologists are more accurate in every category, though the difference is not always statistically significant. The degree of bias between the 2 groups varies from category to category and both tend to underdiagnose or overdiagnose the same categories.

Analysis of results in selected categories

i) Pulmonary arterial hypertension Having shown in the study that the radiographic diagnosis of pulmonary hypertension was both inaccurate and inconsistent, it was decided to analyse the results in order to determine in which group the greatest difficulty was experienced. Table 4 shows the errors, expressed as a percentage, of over- and underdiagnosis of pulmonary arterial hypertension in the different groups. Pulmonary arterial hypertension was overdiagnosed to a significant degree in the presence of pulmonary venous hypertension (41%) and plethora (22·5%), whereas it was conspicuously underdiagnosed (85·6%) in the Eisenmenger syndrome and to a lesser extent when associated with plethora and pulmonary venous hypertension (63·2 and 56·2%, respectively). The least underdiagnosis, not unexpectedly, was in the thromboembolic group (35·8%).

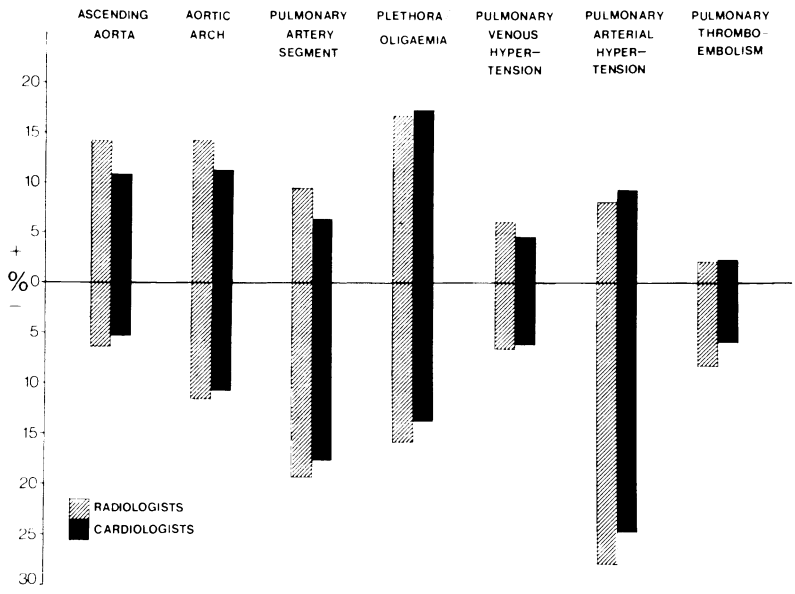

FIG. 4 *Comparison of diagnostic accuracy and bias (over and underdiagnosis expressed as a percentage) in each category between the 5 radiologists and the 5 cardiologists.*

TABLE 4 *Pulmonary arterial hypertension: errors of over- and underdiagnosis*

Category	*No. of radiographs*	*Total possible errors*	*Total errors*	*Percentage error*
Overdiagnosis				
Prominent aorta	8	160	3	1·9
Oligaemia (excluding PS)	6	120	6	5·0
Pulmonary stenosis (PS)	8	160	12	7·5
Idiopathic dilatation of pulmonary artery	4	80	6	7·5
Normal	7	140	11	7·9
Plethora	12	240	54	22·5
Pulmonary venous hypertension	10	200	82	41·0
Underdiagnosis				
Thromboembolic disease	12	240	86	35·8
Plethora and pulmonary arterial hypertension	11	220	139	63·2
Pulmonary venous and arterial hypertension	13	260	146	56·2
Eisenmenger syndrome	9	180	154	85·6

ii) Oligaemia The incidence of overdiagnosis (radiographs that were assessed as normal or plethoric) in the oligaemia group was 64·3 per cent (Table 5). The results in the 8 cases with pulmonary stenosis and intact ventricular septum were then analysed separately from the remainder (Ebstein's anomaly, Fallot's tetralogy, and tricuspid atresia). In the first group the percentage error was significantly higher (73·1% as compared to 52·5%). The results in the 14 cases were then analysed according to the presence or absence of a right-to-left shunt, irrespective of the diagnosis, and it was found that oligaemia was correctly diagnosed significantly more often when this was present (47·5% error) than when there was no shunt (76·9% error). Furthermore, assessment of oligaemia in patients with pulmonary stenosis and a right-to-left shunt was almost twice as accurate (47·5% error) as in the patients with pulmonary stenosis and no shunt (81·7% error).

TABLE 5 *Oligaemia category: errors of overdiagnosis*

Category	*No. of radiographs*	*Total possible errors*	*Total errors*	*Percentage error*
Total oligaemia group	14	280	180	64·3
Pulmonary stenosis (PS)	8	160	117	73·1
Other	6	120	63	52·5
All with right-to-left shunt	6	120	57	47·5
All without right-to-left shunt	8	160	123	76·9
PS with right-to-left shunt	2	40	19	47·5
PS without right-to-left shunt	6	120	98	81·7

TABLE 6 *Thromboembolic disease: errors of over- and underdiagnosis of 3 observers*

Overdiagnosis (88 radiographs)

Observer	*Thromboembolic disease absent (0)*		*Possible thromboembolic disease (+1)*		*Thromboembolic disease (+2)*	
	No.	%	*No.*	%	*No.*	%
2	172	97·7	4	2·3	0	0
6	169	96·0	5	2·8	2	1·1
10	172	97·7	1	0·56	3	1·7
Mean		97·1		1·9		0·93

Underdiagnosis (12 radiographs)

Observer	*Thromboembolic disease (0)*		*Possible thromboembolic disease (−1)*		*Radiographs called normal (−2)*	
	No.	%	*No.*	%	*No.*	%
2	15	62·5	5	21	4	17
6	17	71·0	4	17	3	12
10	18	75·0	0	0	6	25
Mean		69·5		12·7		18

iii) Thromboembolic disease The overall diagnostic accuracy was good (90·4% correct), with only 2·6 per cent overdiagnosis and 7·1 per cent underdiagnosis (Table 2). However, if only the 12 radiographs with thromboembolic disease were considered, it was found that in the 240 viewings the percentage underdiagnosis was, in fact, 59 per cent. This includes diagnoses that were only one class under (possible thromboembolic disease) as well as those that were 2 classes under (radiographs called 'normal'), but none the less remains a significant error of underdiagnosis.

Since the radiographic diagnosis of thromboembolic disease is a controversial one, it was decided to analyse the results of the 3 observers (2, 6, and 10) who were most accurate in this category. Table 6 shows the number and distribution of errors of these observers in the thromboembolic group. The total number of possible correct diagnoses for each observer is 24 (12 radiographs, 2 viewings). As can be seen (Table 6), thromboembolic disease was correctly diagnosed in nearly 70 per cent of cases, while possible thromboembolic disease was suggested in another 12·7 per cent and the diagnosis missed completely in 18 per cent. The errors of overdiagnosis are also shown in Table 6; thromboembolism was incorrectly diagnosed in less than 1 per cent of the 528 viewings (88 radiographs, 2 viewings, 3 observers), while possible thromboembolic disease was suggested in a further 1·9 per cent.

Possible variables

Finally, in order to determine whether there were any trends in diagnosis, the categories were analysed in relation to possible variables; these included the age of the patient, severity of the haemodynamic disturbance, and the diagnosis. For example, the plethora group was analysed according to age, magnitude and level of the left-to-right shunt and

height of the pulmonary arterial pressure. Similarly, the results in pulmonary venous hypertension were analysed according to the mean left atrial pressure, associated pulmonary arterial hypertension, and age of the patient. No correlation could be found between any of these variables and the accuracy of diagnosis.

Quality of radiographs

The quality of the films was assessed as good in 60 per cent, fair in 34 per cent, and poor in 6 per cent. It was found that the 6 observers attached to this cardiac clinic, and therefore familiar with the particular make of film used in the unit (which comprised 82% of the total), marked a significantly higher number of radiographs as good (about 80%) than did the outside observers (34% good and 56% fair).

Discussion

Numerous surveys (Birkelo *et al.*, 1947; Comroe and Botelho, 1947; Garland, 1949; Yerushalmy *et al.*, 1951; Cochrane and Garland, 1952; Grøth-Petersen, Lovgreen, and Thillemann, 1952; Shiels and Thomas, 1954; Whyte, 1955; Davies, 1958; Anderson and Cowan, 1961; Tuddenham, 1962; Fouché, Beck, and Schrire, 1963; Scheff, 1963; Schreiber, 1963; Smith, 1967; Meade *et al.*, 1968; Seymour and Conway, 1969) have shown a significant observer error in the many areas studied. Most of the studies have been done in the medical sphere, many of which were in radiology and often related to the diagnosis of tuberculosis. In 1949, Garland found an interindividual variation of 9 to 24 per cent and an intraindividual variation of 3 to 31 per cent in the radiographic diagnosis of pulmonary tuberculosis, whereas in a later study he and Cochrane (1952) found a 30 per cent underreading of positive chest films and an interindividual variation of 20 per cent. Whyte (1955) reported that the average extent to which any one of three radiologists and three experienced physicians agreed with any one of his colleagues varied from 51 to 60 per cent. Of interest is the fact that in Whyte's study the radiologists were more accurate and consistent than the physicians, in contrast to the present series in which the cardiologists were more accurate in every category. Tuddenham (1962), in his Memorial Fund lecture, quoted an earlier experiment in which he found that 20 to 30 per cent of 'reportable' findings were overlooked by three experienced radiologists. He also referred to other studies (Birkelo *et al.*, 1947; Yerushalmy *et al.*, 1951; Cochrane and Garland, 1952; Grøth-Petersen *et al.*, 1952) in which experts missed from 25 to 30 per cent of positive chest films. Smith (1967), in an extensive review of studies of observer error in various areas, summarized the results by stating that underreading has been reported in 6 to 54 per cent of cases and overreading in 0·9 to 18 per cent. However, the use of per cent calculations is deceptive and if the total number is considered then *overreading* occurs much more frequently (Scheff, 1963; Smith, 1967). Scheff (1963) has attributed this to the fact that underreading (missing a lesion) is considered unacceptable in current medical practice. In contrast, the findings in the present study show a tendency towards underdiagnosis (Fig. 1) by all except 2 observers (1 and 6) in the majority of categories, particularly that of pulmonary arterial hypertension.

Evaluation of accuracy of diagnosis (Table 2) shows that this is most accurate in the categories of thromboembolism (90·4%), pulmonary venous hypertension (88·2%) and aortic prominence (81·6%), while the least accuracy is shown in the diagnosis of pulmonary arterial hypertension (65·1%) and oligaemia (68·1%). The consistency in the various categories (Table 3) correlates fairly well with the accuracy of each observer in those categories (Fig. 2) suggesting that the radiological diagnosis of pulmonary thromboembolism, venous hypertension, and aortic prominence is clear cut, whereas the radiographic features of pulmonary arterial hypertension and oligaemia, or the criteria used in their assessment, are less accurate and reliable.

Attempts to correlate the accuracy of diagnosis with the severity of disease, as assessed clinically and at cardiac catheterization, were unsuccessful. This unexpected finding is in accordance with the experience of other investigators (Garland, 1949; Grøth-Petersen *et al.*, 1952; Tuddenham, 1962). One of Tuddenham's conclusions, which has been confirmed by Garland (1949) and by Grøth-Petersen *et al.* (1952), was that errors of underreading were not due to the fact that only small or difficult lesions were analysed, since among them were included cases of extensive disease. He attributed this to incomplete 'search' rather than to poor quality of the films, since 80 per cent of the abnormalities missed were obvious in retrospect.

An important point which has been made in previous studies (Garland, 1949; Schreiber, 1963) is that correlation of radiological features with the clinical findings and laboratory data makes the results of radiographic examination much more reliable. Schreiber (1963) found a statistically significant improvement in performance, due for the most part to a decrease in the number of false negatives. This conclusion is certainly borne out by

this study in that the number of false negatives in the pulmonary arterial hypertension category, particularly in the Eisenmenger group, would have been significantly decreased by a knowledge of the clinical findings, as would the errors of overdiagnosis of plethora and pulmonary arterial hypertension in the normal group.

The use of dual readings has been emphasized by a number of observers (Birkelo *et al.*, 1947; Yerushalmy *et al.*, 1950; Shiels and Thomas, 1954; Seymour and Conway, 1969). Seymour and Conway (1969), in a study of 1,222 electrocardiograms, found that they were in disagreement in 8 per cent whereas agreement could be reached in all except 3 per cent when they reviewed the tracings together. Shiels and Thomas (1954), Birkelo *et al.* (1947), and Yerushalmy *et al.* (1950) reported that dual readings of chest radiographs were of value in decreasing false negatives but increased the number of false positive readings, whereas Birkelo *et al.* (1947) found that use of 3 or more readers decreased underreading without increasing overreading, a procedure which, incidentally, is routine with every new patient seen at our cardiac clinic.

In this study we were particularly interested in the detectability on chest radiographs of plethora, oligaemia, pulmonary arterial hypertension, and pulmonary thromboembolic disease. Lester (1968) found that left-to-right shunts were not recognizable on plain chest films if the pulmonary to systemic flow ratio was less than 1·8:1, whereas Fouché and co-workers (1963) put the dividing line at 2:1. However, Simon (1968) found that with shunts of less than 2:1 some dilatation of upper zone vessels and a decrease in the normal difference between upper and lower zones could be detected. The changes became obvious as the shunt approached 4:1 and gross with shunts greater than 8:1. The results in this study correlate to an extent with Simon's findings (1968), in that with shunt ratios of 5:1 or more none of the observers missed the plethora. At lower shunt magnitudes, however, the correlation was not good, and we found that there was less underreading of shunts of 1·4:1 and 1·7:1 than occurred with bigger shunts (from 2 to 3·3:1). In fact, apart from the very large shunts (greater than 5:1), no correlation could be found between correct assessment of plethora on the chest radiographs and the size of the shunt.

Lester (1968) and Desilets *et al.* (1968) believe that, in the absence of a right-to-left shunt and of cardiac decompensation, patients with Ebstein's anomaly and pulmonary stenosis display normal vasculature. Simon (1968) and Cooley and Schreiber (1967), while agreeing that the pulmonary vasculature is normal in mild or moderate pulmonary stenosis, state that in severe cases the intrapulmonary vessels are reduced in calibre. In view of this controversy, we analysed separately our cases of oligaemia with and without a right-to-left shunt, and also compared the patients with pulmonary stenosis and right-to-left shunt through a patent foramen ovale to those without a shunt (Table 5). We found that oligaemia was missed significantly less often in patients with a right-to-left shunt than when this was absent, and that oligaemia was underdiagnosed by an average of 1·5 classes in every viewing in those patients without a right-to-left shunt. Furthermore, when all the cases of pulmonary stenosis were analysed according to the gradient across the pulmonary valve, no correlation with diagnostic accuracy could be shown, suggesting that even in severe pulmonary stenosis reduction in pulmonary vasculature may be absent or not detectable.

Pulmonary arterial hypertension is defined (Simon, Sasahara, and Cannilla, 1967) as a pulmonary artery pressure greater than 30/15 mmHg. However, when Davies and associates (1953) compared plain chest films with angiograms, they observed that radiographic changes were minimal or absent when the pulmonary arterial systolic pressure was below 40 mmHg, between 40 and 70 mmHg arterial changes were always present but confined to the bases, while with a pulmonary arterial systolic pressure exceeding 70 mmHg, the arterial changes were widespread. Simon (1968) found an obvious exaggeration of calibre ratios between central and peripheral vessels if the pulmonary arterial pressure was greater than 40 mmHg, and noted that the central arterial dilatation was most obvious in patients with reversal of previous pulmonary-systemic shunts. For purposes of this study, only patients with a pulmonary arterial systolic pressure of not less than 45 mmHg were accepted into the pulmonary arterial hypertension category, and the majority had pressures above 60 mmHg. Nevertheless, the percentage underdiagnosis was disappointingly high, even in the presence of severe hypertension. It must be emphasized, however, that in the studies quoted (Davies *et al.*, 1953; Simon *et al.*, 1967), the radiographs were compared to angiograms and the diagnoses known. It is almost certain that knowledge of the clinical findings, particularly in the Eisenmenger group, would have significantly decreased the incidence of underreading by our observers.

The diagnosis of pulmonary thromboembolic disease presents a major problem in medical practice today. Harris (1959) described 60 cases of fatal pulmonary embolism, both acute and chronic, and found that in 50 per cent the correct diagnosis had

not even been considered. The findings on plain chest films in chronic thromboembolic disease, first described by Westermark (1938) and since confirmed by others (Weidner, Swanson, and Wilson, 1967; Greenspan and Steiner, 1969; Siber, 1969), include increased size and abrupt termination of major vessels and increased radiolucency with decreased or absent vascularization in areas supplied by the occluded vessels. The overall percentage underreading of the 12 radiographs showing thromboembolic disease was 59 per cent. When the results of the three most accurate observers were analysed, it was found that the percentage underdiagnosis (both classes) was 30·7 per cent (Table 6) while the percentage overdiagnosis was only 2·8 per cent. In other words, there was a 30 per cent incidence of false negatives but only 3 per cent false positives. It appears, therefore, that the radiographic signs of thromboembolic disease, though highly diagnostic of the condition, are not always recognized on plain chest films, even by those observers who are very aware of the condition.

The authors wish to record their appreciation to those who took part in this observer study: Drs. M. E. Berk, E. E. Epstein, G. E. Gale, R. Glyn Thomas, H. Jackson, P. E. Marchand, R. B. K. Tucker, M. Zinober,[1] and M. M. Zion.

We thank Dr. H. M. Salmon, Superintendent of the Johannesburg General Hospital, for permission to publish.

This study was supported in part by the South African Medical Research Council.

[1] Now deceased.

References

Anderson, W. F., and Cowan, N. R. (1961). Observer error in recording arterial blood pressure. *British Heart Journal*, **23**, 169.

Birkelo, C. C., Chamberlain, W. E., Phelps, P. S., Schools, P. E., Zacks, D., and Yerushalmy, J. (1947). Tuberculosis case finding. *Journal of the American Medical Association*, **133**, 359.

Cochrane, A. L., and Garland, L. H. (1952). Observer error in the interpretation of chest films. *Lancet*, **2**, 505.

Comroe, J. H., and Botelho, S. (1947). The unrealiability of cyanosis in the recognition of arterial anoxemia. *American Journal of the Medical Sciences*, **214**, 1.

Cooley, R. N., and Schreiber, M. H. (1967). *Radiology of the Heart and Great Vessels*, 2nd ed., pp. 4, 312, and 323. Williams and Wilkins, Baltimore.

Davies, L. G. (1958). Observer variation in reports on electrocardiograms. *British Heart Journal*, **20**, 153.

Davies, L. G., Goodwin, J. F., Steiner, R. E., and Van Leuven B. D. (1953). Clinical and radiological assessment of the pulmonary arterial pressure in mitral stenosis. *British Heart Journal*, **15**, 393.

Desilets, D. T., Marcano, B. A., Emmanouilides, G. C., and Gyepes, M. T. (1968). Severe pulmonary valve stenosis and atresia. *Radiologic Clinics of North America*, **6**, 367.

Fouché, R. F., Beck, W., and Schrire, V. (1963). The roentgenologic assessment of the degree of left-to-right shunt in secundum type atrial septal defect. *American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine*, **89**, 254.

Garland, L. H. (1949). On the scientific evaluation of diagnostic procedures. *Radiology*, **52**, 309.

Greenspan, R. H., and Steiner, R. E. (1969). The radiologic diagnosis of pulmonary-thromboembolism. In *Frontiers of Pulmonary Radiology*, p. 222. Ed. by M. Simon, E. J. Potchen, and M. Le May. Grune and Stratton, New York.

Grøth-Petersen, E., Lovgreen, A., and Thillemann, J. (1952). On the reliability of reading of photofluorograms and value of dual reading. *Acta Tuberculosea Scandinavica*, **26**, 13.

Harris, H. W. (1959). Pulmonary embolism and infarction. *Medical Clinics of North America*, **43**, 69.

Lester, R. G. (1968). Radiological concepts in the evaluation of heart disease (1). *Modern Concepts of Cardiovascular Disease*, **37**, 113.

Meade, T. W., Gardner, M. J., Cannon, P., and Richardson, P. C. (1968). Observer variability in recording the peripheral pulses. *British Heart Journal*, **30**, 661.

Scheff, T. J. (1963). Decision rules, types of error, and their consequences in medical diagnosis. *Behavioural Science*, **8**, 97.

Schreiber, M. H. (1963). The clinical history as a factor in roentgenogram interpretation. *Journal of the American Medical Association*, **185**, 399.

Seymour, J., and Conway, N. (1969). Value of dual reports on routine electrocardiograms. *British Heart Journal*, **31**, 610.

Shiels, D. O., and Thomas, D. L. G. (1954). Comparison of different types of X-ray films for case finding and diagnosis of silicosis. *Medical Journal of Australia*, **1**, 200.

Siber, F. J. (1969). The radiologic spectrum of pulmonary embolic disease. *Medical Clinics of North America*, **53**, 375.

Simon, M. (1968). The pulmonary vasculature in congenital heart disease. *Radiologic Clinics of North America*, **6**, 303.

Simon, M., Sasahara, A. A., and Cannilla, J. E. (1967). The radiology of pulmonary hypertension. *Seminars in Roentgenology*, **2**, 368.

Smith, M. J. (1967). *Error and Variation in Diagnostic Radiology*, pp. 156 and 168. Charles C. Thomas, Springfield, Illinois.

Tuddenham, W. J. (1962). Visual search, image organization, and reader error in roentgen diagnosis. *Radiology*, **78**, 694.

Weidner, W., Swanson, L., and Wilson, G. (1967). Roentgen techniques in the diagnosis of pulmonary thromboembolism. *American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine*, **100**, 397.

Westermark, N. (1938). On roentgen diagnosis of lung embolism. *Acta Radiologica*, **19**, 357.

Whyte, H. M. (1955). Differences in opinions of radiological changes in pulmonary tuberculosis. *Australasian Annals of Medicine*, **4**, 100.

Wood, P. (1956). *Diseases of the Heart and Circulation*, 2nd ed., p. 392. Eyre and Spottiswoode, London.

Yerushalmy, J., Garland, L. H., Harkness, J. T., Hinshaw, H. C., Miller, E. R., Shipman, S. J., and Ziverling, H. B. (1951). An evaluation of the role of serial chest roentgenograms in estimating the progress of disease in patients with pulmonary tuberculosis. *American Review of Tuberculosis*, **64**, 225.

Yerushalmy, J., Harkness, J. T., Cope, J. H., and Kennedy, B. R. (1950). The role of dual reading in mass radiography. *American Review of Tuberculosis*, **61**, 443.

Requests for Reprints to Professor J. B. Barlow, Department of Medicine, General Hospital, Johannesburg, South Africa.