Scand j work environ health 9 (1983) 162—168

Pathology standards for asbestosis

by Francis HY Green, MD, Michael Attfield, BSc [1]

GREEN FHY, ATTFIELD M. Pathology standards for asbestosis. *Scand j work environ health* 9 (1983) 162—168. Pathology standards for asbestos-associated diseases of the lungs and pleural cavities were recently developed by the Pneumoconiosis Committee of the United States College of American Pathologists under contract to the US National Institute for Occupational Safety and Health (NIOSH). The purpose of the contract was to develop standardized criteria for the pathological diagnosis of these diseases and to develop a system for grading the severity and extent of asbestosis. The results of a preliminary reading trial and the NIOSH statistical analysis of the trial are presented. These results indicate that the proposed grading schema has acceptable inter- and intra-observer variability. The variability is similar to that observed for radiologists in radiographic reading trials.

Key terms: pneumoconiosis, pulmonary fibrosis, reading trials.

The University of Vermont and the College of American Pathologists, under contract to the National Institute for Occupational Safety and Health (NIOSH), have recently developed standardized criteria for the pathological diagnosis of asbestos-associated diseases of the lungs and pleural cavities. In addition the contract called for the development of a system for grading the severity and extent of asbestosis. The need to develop these standards grew out of several observations. First, while the pathological features of advanced asbestosis appeared to be generally well recognized, the early lesions of asbestosis are not well documented in the literature, nor are possible differences in the pathological response to different asbestos types or to asbestos combined with confounding minerals such as silica. Most importantly the minimal histological criteria for the diagnosis of asbestosis have not been defined. Second, in order to develop sound epidemiologic studies and to compare findings between different laboratories, it is necessary to have agreed definitions and a reproducible system for measuring the extent and severity of the disease. Third there has been a veritable explosion of compensation claims against the major asbestos manufacturers (2). Many of these involve the possibility of occupational lung cancer in which the presence or absence of histologically confirmed asbestosis may be pivotal.

Previous experience in developing standards for coal worker's pneumoconiosis has shown that standardized criteria are of value in all these areas (5). They also provide a basis for developing more accurate criteria or better systems of grading as more is learned about the natural history of asbestos-associated diseases.

In 1980 a committee of eight US pulmonary pathologists and one pulmonary epidemiologist [2] was assembled under the aegis of the Pneumoconiosis Committee of the College of American Pathologists. The Committee was chaired by Dr JE Craighead of the University of Vermont. The purpose of the Committee was to review case material from asbestos-exposed individuals in both the United States and worldwide and develop a report. To our surprise very few collections of pathological case material were available. In those

[1] Pathology and Statistics Sections, Division of Respiratory Disease Studies, Morgantown, West Virginia, United States.

Reprint requests to: Dr FHY Green, Division of Respiratory Disease Studies, 944 Chestnut Ridge Road, Morgantown, WV 26505, USA.

[2] JE Craighead, JL Abraham, A Churg, FHY Green, J Kleinerman, PC Pratt, TA Seemayer, V Vallyathan, H Weill.

that were located, exposure and clinical information was usually lacking. It was also difficult to obtain specimens from workers exposed to relatively pure forms of the four major commerical forms of asbestos (chrysotile, crocidolite, amosite and anthophyllite). We were grateful therefore to obtain access to material from the Commission de la Sante' et de la Securite' du Travail du Quebec, where exposure was predominantly to chrysotile asbestos, and from Drs LO Meurman and Y Collan in Finland, where exposure was predominantly to anthophyllite. Case material was also obtained from Great Britain and South Africa. The Committee met four times to review the assembled case material. The resulting report reflects our experience of the case material and extensive discussion (1). In addition the opinions of many recognized experts in the field were solicited, and these were incorporated where appropriate.

The Committee's charge was to describe only lesions of the lungs and pleural cavities; therefore neoplasms of the larynx and gastrointestinal tract were not considered. Four lesions are considered in detail in the report, ie, pleural plaques, asbestosis, lung cancer, and mesothelioma. Emphasis has been placed on the probable morphogenetic sequence of events in the evolution of asbestosis and on the minimal histological criteria to establish the diagnosis. In addition the report of the Committee devotes considerable time to the significance of asbestos bodies and fibers in the lung and recommends methods for the pathological study of a suspected case of disease due to asbestos exposure.

The monograph also includes a proposed grading system for asbestosis and the results of a preliminary reading trial. A separate analysis of this trial was conducted by NIOSH, and this analysis is presented in detail in the present communication.

Several grading systems were considered by the Committee, some of which were relatively complicated and designed with specific research goals in mind. An important consideration in developing a grading system is that it should be relatively simple to perform and have acceptable reproducibility. It is also important that the grading system measure the most significant parameters of the disease and be comparable to systems used by other disciplines, for example, the radiographic classification of pneumoconiosis developed by the UICC (Internationale Union Contre Cancer) (7).

The system adopted fulfills many of these conditions. It is semiquantitative and measures the extent and severity of the pulmonary fibrosis of asbestosis. Certain aspects of a grading system proposed by Hinson et al (4) were incorporated into this schema. The proposed grading system is not intended to be a substitute for diagnosis, which should be made according to the criteria set out in the monograph (1). It is based on histological material only. This procedure has several advantages. First there is a permanent record. Thus the grading can be repeated by different observers should such repetition be necessary for medicolegal or scientific purposes. In addition retrospective studies can be performed. Second the system is applicable to small specimens such as biopsy samples. Third fibrosis is more easily recognized histologically than grossly. It could be argued that the small area sampled in a tissue section could seriously bias the reading in view of the irregular distribution of the disease. However, if multiple representative sections are taken from each lobe, then this bias can be minimized, and an average grade can be obtained for the whole lung or for individual lobes (1). Although inflation fixation is recommended, the grading system is applicable to uninflated lung sections also.

As mentioned earlier, both severity and extent are measured. The criteria for severity are based on a likely pathogenic sequence of events in the evolution of asbestosis. The criteria for severity and extent are presented in table 1. Typical examples are shown in fig 1—4.

Materials and methods

The material used for testing the grading system was obtained from the Archives of the Armed Forces Institute of Pathology, courtesy of Dr L Hochholzer, Dr J Rust and Col E Cowart. The material was screened from several hundred cases by two pathologists and selected from among

Table 1. Asbestosis grading schema (1).

Grade	Definition
Severity [a]	
0	No fibrosis associated with bronchioles.
1	Fibrosis involving the wall of at least one respiratory bronchiole with or without extension into the septa of the immediately adjacent layer of alveoli; no fibrosis in more distant alveoli
2	Fibrosis as in grade 1 plus involvement of alveolar ducts and/or 2 or more layers of adjacent alveoli; there still must be a zone of nonfibrotic alveolar septa between adjacent bronchioles
3	Fibrosis as in grade 2 but with coalescence of fibrotic change such that all alveoli between at least two adjacent bronchioles have thickened, fibrotic septa; some alveoli may be obliterated completely
4	Fibrosis as in grade 3 but with formation of new spaces of a size larger than alveoli ranging up to as much as 1 cm; this lesion has been termed "honeycombing;" the spaces may or may not be lined by epithelium
Extent [b]	
A = 1	Only occasional bronchioles are involved — most show no lesion
B = 2	More than "occasional" but less than half of all bronchioles involved
C = 3	More than half of all bronchioles involved

[a] Lesions associated with individual respiratory bronchioles are evaluated. The grade is based on *the most severe* lesion in the slide, not a visual average of the changes found in the various individual respiratory units.

[b] The proportion of the respiratory bronchioles involved by the disease process is assessed. The grades pertain to the relative numbers of bronchioles, in the slide, involved by *any degree of fibrosis,* not just the numbers involved to the maximum degree as recorded under severity.

Fig 1. Grade I asbestosis. The fibrosis is confined to the walls of a respiratory bronchiole (RB). (Hematoxylin and Eosin, 15 ×)

Fig 2. Grade II asbestosis. The fibrosis is centered on a respiratory bronchiole but is extending out into adjacent alveolar septa. An asbestos body is seen in the fibrosis (arrow). (Hematoxylin and Eosin, 15 ×)

others because they were believed to be representative of different degrees of asbestosis. Two sections, A and B, were assessed in each of 20 cases. Slide A of each pair was read twice (once in the morning and once in the afternoon) by each of nine pathologists, while slide B was read only once by each. Each slide was read independently by every pathologist with approximately 5 min allotted to each slide. It should be noted that all nine readers were specialists in pulmonary pathology.

Results

Interpathologist variation

Analysis was undertaken separately for severity and extent, as it was felt that a combined score was inappropriate. For instance, it was felt that a product score of 4, obtained from a severity of four times an extent of 1, had a different interpretation from a severity of 2 and an extent of 2. In addition, since both grading schemes are new, it was felt desirable to evaluate both as it might be found that one was reliable and the other was not. Modifications could then be made with greater confidence about the source of the unreliability.

Table 2 shows the distribution of the severity gradings for each pathologist. Although the distributions do not differ markedly, there is some variation between pathologists. For example C clearly believed that the majority of the slides were grade 3 severity, while G thought that most were grade 2. Pathologist B read one of the slides as being category 0. The mean level was 2.6 with a range of 2.3—2.9.

Table 3 shows the information for extent in a similar manner. Again the distributions do not differ markedly among the pathologists. In particular, all except one categorized most slides into categories B and C. If we assign the scores A = 1, B = 2, C = 3, we can obtain mean values of extent which give a rough indication of the average level. These show a minimum of 2.0 for G and a maximum of 2.6 for A.

Fig 3. Grade III asbestosis. The fibrosis extends between adjacent respiratory units. (Hematoxylin and Eosin, 15 ×)

Fig 4. Grade IV asbestosis. The normal pulmonary architecture has been replaced by fibrous septa enclosing cystic spaces (honeycombing). (Hematoxylin and Eosin, 15 ×)

Table 2. Distribution of the pathologists' severity scores.

Pathologist	Severity score					Missing	Number	Mean
	1	2	3	4	0			
A	7	14	18	19	—	—	58	2.8
B	11	8	16	22	1	—	58	2.8
C	2	13	29	13	—	1	57	2.9
D	4	18	18	18	—	—	58	2.9
E	14	15	13	16	—	—	58	2.5
F	18	14	18	8	—	—	58	2.3
G	8	25	11	14	—	—	58	2.5
H	10	18	14	16	—	—	58	2.6
I	9	23	14	12	—	—	58	2.5
Total number	83	148	151	138	1	1	521	•
Percentage	16	28	29	26	0	0	99	•

Table 3. Distribution of the pathologists' extent scores.

Pathologist	Extent score			Missing	Number	Mean [a]
	A	B	C			
A	3	18	37	—	58	2.6
B	6	27	25	—	58	2.3
C	3	23	31	1	57	2.5
D	8	28	22	—	58	2.2
E	10	20	28	—	58	2.3
F	19	15	24	—	58	2.1
G	17	24	17	—	58	2.0
H	12	24	22	—	58	2.2
I	5	40	13	—	58	2.1
Total number	83	219	219	1	521	•
Percentage	16	42	42	0	100	•

[a] Calculated assuming A = 1, B = 2, C = 3.

Table 4. Interpathologist reproducibility for severity and extent.

	Severity		Extent	
	Mean	Range	Mean	Range
Percentage of exact agreement	57	38— 71	53	31— 76
Percentage of agreement to ± 1 category	94	81— 100	92	76— 100
Average difference without regard to sign	0.5	0.31—0.78	0.6	0.24—0.90

Table 5. Intrapathologist reproducibility for severity and extent.

	Severity		Extent	
	Mean	Range	Mean	Range
Percentage of exact agreement	71	55— 85	69	60— 80
Percentage of agreement to ± 1 category	98	90— 100	98	90— 100
Average difference without regard to sign	0.3	0.15—0.45	0.3	0.2—0.45

Although two pathologists may obtain identical distributions of grades of severity or extent, this is no guarantee that they agree exactly on the classification of every slide. Indeed film readers often obtain very similar prevalences among groups of X-ray films but disagree on individual classifications. Thus, in addition to studying the distributions of severity and extent among pathologists, we must also examine agreement of assessments of individual slides.

Interobserver reproducibility is often conveniently studied through the examination of pairwise agreement and disagreement. The following three statistics have been chosen to measure reproducibility:

(a) Percentage of exact agreement, defined as *number of slides on which the pair of pathologists agreed exactly* $\times$ *100*
Total number of slides examined jointly (58 in most cases)

(b) Percentage of agreement to within $\pm$ one category, defined as *number of slides for which pathologists agreed to within* $\pm$ *one category* $\times$ *100*
Number of slides examined jointly (58 in most cases)

(c) Degree of difference, defined as *sum of differences between two readings (less sign)*
Total number of slides examined jointly (58 in most cases)

The first statistic has an obvious interpretation. As its name implies, it measures the degree to which pairs of pathologists agreed exactly on the assessment of slides. The second has a similar meaning except that the definition of agreement has been widened to include a minor disagreement of plus or minus one category. (Under this definition, extent scores of A and B and of B and C would count as agreement, while scores of A and C would not.) The third statistic reflects the degree, in terms of categories, to which two pathologists might differ in their interpretation of the same slide. These three statistics have been calculated for each possible pair of pathologists; there were 36 such pairs.

Table 4 shows the mean and range of the percentage of exact agreement for severity over the 36 pairs. The mean percentage of exact agreement of 57 implies that random pairs of pathologists from this group would be expected to agree exactly on the assessment of one slide in two. The lowest pairwise agreement was 38 %, and the highest 71 %. Table 4 also shows the results for extent. The mean is 53 %, and the values ranged from 31 to 76 %.

Much of the disagreement between pairs of pathologists consisted of only one category since the percentage of agreement to within $\pm$ one category statistics were all large (table 4). For severity the lowest and highest values were 81 and 100 %, respectively, while for extent they were 76 and 100 %, respectively. Table 4 shows that the degree of difference between any two pathologists in the group averaged about half a category. The worst situation was seen for extent, for which the average disagreement for one pair of pathologists amounted to nearly one category. It is of interest to note that complete agreement on the severity scores was obtained for 9 out of the 58 slides (16 %), and for extent 7 of the 58 (12 %) (the corresponding statistics for $\pm$ one category were 57 and 43 %, respectively).

Intrapathologist variation

The same three statistics as for interpathologist variation were used to explore intrapathologist variation. Twenty slides were each read twice by each pathologist, a procedure giving rise to nine values of each statistic for severity and nine also for extent. The results are expressed in terms of means and ranges in table 5. The intrapathologist reproducibility on severity was much better than the interpathologist reproducibility, as might be expected. For example, the mean percentage of exact agreement was 71 compared to the 57 % for the interpathologist evaluation. The least reproducible pathologist agreed with his own assessments only 55 % of the time, while the best agreed 85 % of the time. However all of the pathologists rarely differed by more than one category in their repeated readings. The average amount of disagreement was about a third of a category. The results for extent almost duplicated those for severity.

Comments

The interpathologist agreement in this trial was close to 55 % for exact pairwise agreement. This value is comparable to that obtained for another established important medical tool, the ILO U/C classification of radiographs (6). For example, Felson et al (3) obtained exact pairwise agreement of 59 %, and agreement to ± one category of 98 % in their classification of coal miners' films for profusion of small rounded opacities.[3] (Since the radiologists in that report were only using a three-point scale, we might conclude that the pathologists' agreement was possibly slightly superior.)

Experience with film reading has shown that, while readers frequently agree on the presence or absence of disease, there is frequent disagreement on level of abnormality for those films which are said to be abnormal by at least one reader (3). Thus the agreement seen between pairs of readers tends to reduce as the percentage of abnormal films being examined increases. In this respect then, if this tendency can be applied to the pathologists, this trial was a severe test of their agreement as it contained only abnormal samples. In other situations, when the material being studied includes samples of normal tissue, one would expect the agreement to be better.

In conclusion the results of this preliminary trial are encouraging. However it would be invalid to generalize these results as the panel of pathologists that took part was not a random selection of all pathologists. Additional, better-controlled trials are indicated with sections carefully chosen to be equally representative of the different grades and to include normal tissue. The time interval between readings should be extended to measure intrapathologist variation better, and readings should be made by both expert and nonexpert readers.

[3] Values obtained from tables I—III of that report eliminating category O.

Acknowledgments

We would like to thank K Kennedy and S Shores for their secretarial assistance.

References

1. Craighead JE, Abraham JL Churg A, Green FHY, Kleinerman J, Pratt PC, Seemayer TA, Vallyathan V, Weill H. The pathology of asbestos-associated diseases of the lungs and pleural cavities: Diagnostic criteria and proposed grading schema. Arch pathol lab med 106 (1982) 543—596.
2. Craighead JE, Mossman BT. The pathogenesis of asbestos-associated diseases. New engl j med 306 (1982) 1446—1455.
3. Felson B, Morgan WKC, Bristol LJ, Pendergrass EP, Dressen EL, Linton OW, Reger RB. Observations on the results of multiple readings of chest films in "Coal Miners' Pneumoconiosis." Radiology 109 (1973) 19—23.
4. Hinson KFW, Otto W, Webster I, Rossiter CE. Criteria for the diagnosis and grading of asbestosis. In: Bogovski P, Timbrell V, Gilson JC, Wagner JC, ed. Biological effects of asbestos. International Agency for Research on Cancer, Lyon 1973. (IARC scientific publication no 8, 54).
5. Kleinerman J, Green F, Laqueur W, Taylor G, Harley R, Pratt P, Wyatt J, Naeye R. Pathology standards for coal workers' pneumoconiosis. Arch pathol lab med 103 (1979) 375—432.
6. Medical Radiography and Photography Staff. ILO U/C 1971 international classification of radiographs of the pneumoconioses. Med radiogr photogr 48 (1972) 67—110.
7. ———. UICC-Cincinnati classification of the radiographic appearance of the pneumoconioses. Chest 58 (1970) 57—67.