Observer variation in computed tomography of pleural lesions in subjects exposed to indoor asbestos


ABSTRACT: To assess the reliability of computed tomography (CT) in detecting discrete pleural lesions, the interobserver and intra-observer variability in reading the conventional and high-resolution CT (HRCT) scans of 100 volunteers, who had worked for ≥10 yrs in a building with known asbestos contamination, was evaluated.

In the first session, pleural abnormalities were detected by a single radiologist (A1) in 13 subjects. In the second session, the scans were read again independently by the same radiologist (A2) and two other experienced radiologists (B, C). The final decision for the presence of pleural lesions was made in a final consensus reading.

This gave a diagnosis of pleural abnormalities in 18 subjects, of whom eight (44%) had been detected by all three readers, five (28%) by two readers and four (22%) by only one reader; one scan, rated normal by all readers during the second session, was reconsidered because pleural abnormalities had been noted at the first reading (A1). The intra-observer agreement for reader A was good (kappa (κ) 0.68) but the interobserver agreement between the readers was only fair to moderate (weighted κ: A2-B = 0.43, A2-C = 0.45, B-C = 0.26) in the second reading session.

In conclusion, when looking for the prevalence of pleural lesions in indoor asbestos exposed subjects, the potential lack of consistency in reporting the presence of small pleural abnormalities must be borne in mind and strict precautions must be taken.

a health screening examination to detect possible health effects of past asbestos exposure. The present study group includes the first 100 consecutive volunteers who were examined at the University Hospitals Gasthuisberg in Leuven. This group consisted of 45 males and 55 females with a mean ± SD age of 54 ± 4 yrs. The subjects had worked in the Berlaymont building for 16 ± 5 yrs with the time since first employment ranging 13–32 yrs. The volunteers underwent a standardized questionnaire, clinical examination, and lung function testing including measurement of lung volumes, maximal expiratory flow rates and diffusing capacity for carbon monoxide. These results are not described in the present article. The study was carried out with the subjects’ informed consent and the approval of the ethical committee of the University Hospitals Gasthuisberg, Leuven.

**Imaging**

Each subject underwent a CT scan, without injection of contrast material, which consisted of a combination of spiral and high-resolution CT (HRCT). All CT studies were performed with a GE Hispeed CT (General Electric, Milwaukee, Wisconsin, USA). Spiral scans were obtained in prone body position with a slice thickness of 5 mm and a table increment of 9 mm, using 120 kV and 200 mA. HRCT was performed immediately after spiral CT and images of 1 mm slice thickness were obtained at 10 mm intervals. In addition, HRCT scans were performed in prone body position at 30 mm intervals. All HRCT images were acquired using 120 kV, 150 mA and a 1 s acquisition time. Both spiral and HRCT images were photographed at two window settings, appropriate for viewing the lung parenchyma and the pleura and the mediastinum [9].

**Reading of the images**

All scans were read by three experienced board-certified radiologists: reader A and B were two chest radiologists, skilled in the interpretation of thoracic CT scans, reader C was a general radiologist. The CT-scans were read twice by reader A and once by readers B and C. The first reading by reader A (A1) took place when the CT scans were performed during the health screening examination (over a period of ~1 yr). Six months after the end of the study, the scans were read again independently by reader A (A2) and two additional readers (B and C). The readers were unaware of any clinical or other information but they knew that the subjects had worked in the Berlaymont building and that abnormal pleural images had already been found in some cases by reader A1.

Since not every pleural thickening is an asbestos-related pleural plaque, the combined CT-scans were rated on a five-point scale on the basis of presence, extent, distribution and appearance of the pleural changes, taking into account the reader’s confidence in the diagnosis of a pleural plaque. The scoring system was as follows: 0 = normal (including the presence of subcostal fat, intercostal muscles and vessels); 1 = abnormal, but no pleural plaque (e.g. sequelae of tuberculosis or pleurisy: subpleural nodule); 2 = abnormal with a low probability to correspond to a pleural plaque; 3 = abnormal with moderate probability to correspond to a pleural plaque; 4 = abnormal with high probability to correspond to pleural plaque. The readers recorded their interpretation on a form with a precise identification of the CT images showing abnormalities. The CT scans of the subjects in whom a pleural abnormality had been identified at least once were reconsidered for a consensus reading session by the three readers using the same scoring system as described.

**Statistical analysis**

Intra-observer agreement for reader A and inter-observer agreement between the three readers were assessed by means of the $\kappa$ statistics [10]. To assess the intra-observer agreement for the detection of pleural abnormalities, the scores 1, 2, 3, and 4 were grouped from the second reading session. To evaluate inter-reader agreement, taking into account the subjective scores of the second reading session, weighted $\kappa$ statistics were used. The weighted $\kappa$-values were compared with a test based on a standard normal statistic using the estimation of the asymptotic standard error on a weighted $\kappa$ given by Cohen [11]. The 95% confidence intervals for the $\kappa$ statistics were calculated. According to the literature [11, 12]: a $\kappa$-value of < 0.20 indicates poor agreement; a $\kappa$-value of 0.21–0.40, fair agreement; a $\kappa$-value of 0.41–0.60, moderate agreement; a $\kappa$-value of 0.61–0.80, good agreement; and a $\kappa$-value of 0.81–1.00, excellent agreement. Statistical analyses were performed on a personal computer using the Statistical Package for the Social Sciences (SPSS) software (SPSS, Chicago, IL, USA).

**Results**

**Reading of the images**

No overt cases of pulmonary fibrosis, diffuse pleural thickening, bronchial cancer or mesothelioma were detected by HRCT. However, in 13 subjects, a pleural abnormality was initially detected by radiologist A. In four of these subjects other possibilities than asbestos plaques were evoked such as sequelae of (tuberculous) infections, which was confirmed by the questionnaire. In total, 12 subjects had mentioned a history of asbestos with a low probability to correspond to a pleural plaque; 4 = abnormal with high probability to correspond to pleural plaque. The readers recorded their interpretation on a form with a precise identification of the CT images showing abnormalities. The CT scans of the subjects in whom a pleural abnormality had been identified at least once were reconsidered for a consensus reading session by the three readers using the same scoring system as described.

Intra-observer agreement for reader A and inter-observer agreement between the three readers were assessed by means of the $\kappa$ statistics [10]. To assess the intra-observer agreement for the detection of pleural abnormalities, the scores 1, 2, 3, and 4 were grouped from the second reading session. To evaluate inter-reader agreement, taking into account the subjective scores of the second reading session, weighted $\kappa$ statistics were used. The weighted $\kappa$-values were compared with a test based on a standard normal statistic using the estimation of the asymptotic standard error on a weighted $\kappa$ given by Cohen [11]. The 95% confidence intervals for the $\kappa$ statistics were calculated. According to the literature [11, 12]: a $\kappa$-value of < 0.20 indicates poor agreement; a $\kappa$-value of 0.21–0.40, fair agreement; a $\kappa$-value of 0.41–0.60, moderate agreement; a $\kappa$-value of 0.61–0.80, good agreement; and a $\kappa$-value of 0.81–1.00, excellent agreement. Statistical analyses were performed on a personal computer using the Statistical Package for the Social Sciences (SPSS) software (SPSS, Chicago, IL, USA).

Intra-observer agreement for reader A and inter-observer agreement between the three readers were assessed by means of the $\kappa$ statistics [10]. To assess the intra-observer agreement for the detection of pleural abnormalities, the scores 1, 2, 3, and 4 were grouped from the second reading session. To evaluate inter-reader agreement, taking into account the subjective scores of the second reading session, weighted $\kappa$ statistics were used. The weighted $\kappa$-values were compared with a test based on a standard normal statistic using the estimation of the asymptotic standard error on a weighted $\kappa$ given by Cohen [11]. The 95% confidence intervals for the $\kappa$ statistics were calculated. According to the literature [11, 12]: a $\kappa$-value of < 0.20 indicates poor agreement; a $\kappa$-value of 0.21–0.40, fair agreement; a $\kappa$-value of 0.41–0.60, moderate agreement; a $\kappa$-value of 0.61–0.80, good agreement; and a $\kappa$-value of 0.81–1.00, excellent agreement. Statistical analyses were performed on a personal computer using the Statistical Package for the Social Sciences (SPSS) software (SPSS, Chicago, IL, USA).

No overt cases of pulmonary fibrosis, diffuse pleural thickening, bronchial cancer or mesothelioma were detected by HRCT. However, in 13 subjects, a pleural abnormality was initially detected by radiologist A. In four of these subjects other possibilities than asbestos plaques were evoked such as sequelae of (tuberculous) infections, which was confirmed by the questionnaire. In total, 12 subjects had mentioned a history of tuberculosis or pleurisy before starting employment in the Berlaymont building.

The second reading session differed from the first readings by reader A in that a grading system was incorporated for scoring the probability that the pleural irregularity corresponded to a pleural plaque. Table 1 summarizes the scores for each reader separately as well as the score obtained by consensus.

By consensus, 82 scans were eventually considered as normal based on 64 scans (78%) rated normal by all readers, 16 scans (20%) rated abnormal by just one
reader and two scans (2%) where two readers had detected pleural abnormalities. Pleural abnormalities were retained by consensus in 18 subjects of whom eight (44%) had been detected by all three readers, five (28%) by two readers and four (22%) by only one reader. One scan, rated normal by all readers during the second reading session, was reconsidered because pleural abnormalities had been noted at the first reading (A1). Thus, the first (A1) and consensus reading agreed in 95% of the cases, with the consensus reading detecting five additional cases of pleural abnormalities. These five additional cases were graded, by consensus, as follows: two grade 1, one grade 2, and two grade 3. To illustrate the type of pleural lesions found by HRCT, one example of each category score is shown in figure 1.

The intra-observer agreement of reader A between the first and the second reading was good (κ = 0.68) concerning the detection of pleural abnormalities. The inter-reader agreement for each pair of observers in the second reading session is summarized in table 2. These weighted κ-values, evaluating the inter-reader agreement for the subjective scores of the second reading session, were fair to moderate.

Table 1. – Pleural irregularity determined by three readers (A, B, C) separately and the final score obtained by consensus in 100 computed tomography scans

<table>
<thead>
<tr>
<th>Pleural irregularity</th>
<th>First reading</th>
<th>Second reading</th>
<th>Consensus reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
<td>A2 B C</td>
<td>A + B + C</td>
</tr>
<tr>
<td>0 = Normal</td>
<td>87</td>
<td>80 88 75</td>
<td>82</td>
</tr>
<tr>
<td>1 = Abnormal, no pl.pl.</td>
<td>4*</td>
<td>8 7 19</td>
<td>7</td>
</tr>
<tr>
<td>2 = Abnormal, low probability pl.pl.</td>
<td>1 3 6 2</td>
<td>5 0 0 4</td>
<td>4</td>
</tr>
<tr>
<td>3 = Abnormal, moderate probability pl.pl.</td>
<td>9*</td>
<td>6 2 0 5</td>
<td>5</td>
</tr>
<tr>
<td>4 = Abnormal, high probability pl.pl.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

pl.pl.: pleural plaque; *: in the first reading (A1), four cases were graded 1 since the radiologist clearly stated in the protocol that these lesions were not asbestos-related; #: for the other nine cases, the pleural lesions were diagnosed as possibly asbestos-related but no distinction was made between grade 2, 3 and 4. The second reading differed from the first reading in that a grading system was incorporated for scoring the probability that the pleural irregularity corresponded to a pleural plaque. For comparison of the first and consensus reading, nine cases from A1 need to be considered as a pool of grades 2, 3 and 4.

Discussion

The use of HRCT has played an increasing role in the detection of pleural lesions due to occupational and environmental exposure to asbestos [7, 13–20]. However, besides the pilot study [8], epidemiological studies using HRCT for assessing the prevalence of pleural plaques due to indoor asbestos exposure have not been published. Nevertheless, such epidemiological studies can be important since the prevalence of these pleural changes can be interpreted as a biomarker and early health effect of indoor asbestos exposure [21].

In this epidemiological study, the authors opted to combine the readings of the spiral CT (covering the entire chest wall at the expense of resolution) and HRCT (increasing resolution, but at the expense of the amount of tissue scanned), in order to achieve an optimal detection rate for pleural lesions. Nevertheless, this approach is not recommended for pure screening purposes, since conventional CT is recognized as a relatively high-dose diagnostic procedure. HRCT (consisting of 1 mm thick slices performed every 1 cm) has, conversely, an effective radiation dose, which is much lower than that of conventional CT, even with high-dose techniques [22, 23]. The participants of the present study, however, were mature adults who were fully informed about the radiation risk of the procedure and they all volunteered to have these tests performed.

A major problem when undertaking epidemiological studies involving radiology is that trained observers have different thresholds for perceiving the presence or absence of abnormalities. In addition, the technical factors of CT-scanning are very critical for the detection rate of pleural lesions, especially when the extent of the pleural lesions is less pronounced, which will increasingly be the case as a result of strict asbestos legislation [24]. This problem of interobserver accuracy can be tested directly when there is a "gold standard", but in the case of minor pleural plaques, accuracy cannot be tested due to the absence of a standard such as that afforded by surgical, autopsy, or other definite pathological evidence [3]. In the absence of such a "gold standard", intra- and interobserver comparisons of variability can be used.

The observer variability was measured by using the κ statistic, which allows the removal of the component of agreement attributable by chance [25]. The intra-reader agreement for reader A, between the first and second reading, was good, indicating that the reader was consistent in detecting pleural abnormalities. Conversely, there was only a fair to moderate inter-reader agreement for the subjective scores in the second reading session. This can be explained because the main requirement for inter-reader agreement is a commonly accepted language describing the pleural plaques. Pleural plaques are focal areas of pleural thickening (calcified or not) which occur along the lateral chest walls, primarily in the lower portions of the chest. They are seen on the parietal pleura; they are sharply demarcated on all sides and, in particular, from the lung. Pleural lesions with these characteristics are practically pathognomonic for asbestos exposure [26].

However, translating this description of a pleural plaque into radiological diagnostic criteria for HRCT has been a problem. The current International Labour Office classification [27] is not adequate for diagnosing
asbestos-related pleural plaques by CT, leaving us without any good criteria for distinguishing other pleural lesions from plaques [28]. First, the pleural lesions must be differentiated from normal anatomic structures such as subcostal fat, intercostal muscles or veins that may cause densities that are somewhat focal [29]. Secondly, other types of pleural lesions exist which are nonspecific and/or do not have any known connection with exposure to asbestos. This has led to diverging opinions about the radiological definition of plaques diagnosed by CT [15, 30 – 32].

Furthermore, in the present study, the pleural lesions were not those that are commonly described following obvious occupational asbestos exposure. As shown in figure 1, the detected pleural lesions were small, ranging from a few millimetres to at most one centimetre in length. In seven out of the 18 subjects with pleural abnormalities, the lesions were only detected unilaterally. One of the most convincing arguments to make a diagnosis of asbestos plaques is when the lesions are found bilaterally, but this criterion is not absolute and unilateral locations of asbestos-related plaques have been reported [33].

In a number of instances, the possibility that the lesions observed were asbestos-plaques, was only evoked by the radiologist as one of many other possibilities, and with hindsight some of the pleural lesions may have been due to tuberculosis or pleurisy, making it difficult to state that the pleural lesions were asbestos-related. To achieve a higher certainty that the lesions detected on HRCT were pleural plaques, a scoring system was added to the second and consensus reading. Each reader evaluated the pleural abnormalities that were not characteristic of tuberculosis or pleurisy (i.e. score 1), and gave a score of 2, 3, or 4 according to increased probability of being induced by asbestos on the basis of their similarity with frank asbestos plaques. In case of discrepancies between the readers, the decision for the presence of pleural lesions was made in a final consensus reading.

Consensus readings have previously been used in the radiological assessment of asbestos-related abnormalities [7, 34, 35]. In the present study, the final consensus reading resulted in a diagnosis of pleural abnormalities in 18 subjects. Seven of the 18 pleural lesions were considered not to be related to asbestos, but more likely to past tuberculosis or pleurisy. The remaining 11 were considered to be pleural plaques with an increasing probability from low (n ≈ 2), moderate (n ≈ 4) to high (n ≈ 5) for a pleural plaque on the basis of experience and textbook descriptions. The advantage of a consensus reading is that a discussion about the disagreements averages out the differences between the readers of whom A was less "liberal" in seeing abnormalities than reader C, but less "severe" than reader B in ruling out abnormalities. A

Table 2 – Interobserver agreement for the scores of the second reading session

<table>
<thead>
<tr>
<th>Readers</th>
<th>Weighted κ-values (95% confidence intervals)</th>
<th>Grade*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₂ versus B</td>
<td>0.43 (0.19 – 0.66)</td>
<td>Moderate</td>
</tr>
<tr>
<td>B versus C</td>
<td>0.26 (0.07 – 0.44)</td>
<td>Fair</td>
</tr>
<tr>
<td>A₂ versus C</td>
<td>0.45 (0.27 – 0.64)</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

*: a Kappa (κ)-value of ≤0.20 indicated a "poor" agreement; a κ-value of 0.21 – 0.41, "fair" agreement; a κ-value of 0.4 – 0.60, "moderate" agreement; a κ-value of 0.6 – 0.80, "good" agreement; and a κ-value of 0.81 – 1.00, "excellent" agreement [11, 12].
possible disadvantage of consensus readings is that they may operate much like a single-reader interpretation. Although numerous readers may be involved, the information of record is likely to reflect a single, dominant opinion [3], which, however, was not the case in the presented study. The consensus reading detected five additional cases of minor pleural lesions, meaning that a reading by three radiologists increases the detection rate of small pleural lesions compared to a single reading by one radiologist only.

It is important to put so much effort into the detection of subtle pleural lesions, because the diagnostic uncertainties regarding minor asbestos-related pleural changes and the possibly substantial degree of misclassification render it difficult to evaluate the prevalence of these minor pleural changes as an adverse health effect of indoor asbestos exposure [21]. In cases with only "passive exposure" to asbestos, the lesions were found to be extremely small and the consensus reading definitely increased the accuracy of detection. The authors are aware that the detected pleural lesions do not all necessarily correspond to the classical definition of pleural plaques. Therefore, a similar high-resolution computed tomography study has been carried out in 100 control subjects without known or suspected asbestos exposure. The data from this control group is being analysed to verify whether the lesions found here are asbestos-related or not. The present study does not answer this question, but shows the importance of diagnostic accuracy.

Acknowledgements. The authors thank V. De Maertelaer, F. Clottens, G. Nackaerts and W. Desmet for their much-appreciated collaboration.

References

25. Maclure M, Willett WC. Reviews and Commentary.


