The predictive value and inter-observer variability of donor chest radiograph interpretation in lung transplantation

J. Scott Bolton\textsuperscript{a}, Siddharth A. Padia\textsuperscript{a}, Marvin C. Borja\textsuperscript{a}, Patrice Becker\textsuperscript{b}, Jonathan B. Orens\textsuperscript{b}, Charles Wiener\textsuperscript{b}, Stephen C. Yang\textsuperscript{a}, John V. Conte\textsuperscript{a,*}

\textsuperscript{a}Department of Surgery, The Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA
\textsuperscript{b}Department of Medicine, The Johns Hopkins Medical Institutions, Baltimore, MD, USA

Received 4 July 2002; received in revised form 16 December 2002; accepted 7 January 2003

Abstract

Objective: Currently the relationship between the evaluation of the donor chest radiograph and the final disposition of potential donor lungs is unknown, yet potential lung donors receive frequent x-rays. We sought to clarify the role donor chest radiographs and donor lung acceptability.

Methods: We conducted a retrospective review of 84 potential organ donors. Radiographs were reviewed separately by three thoracic surgeons and three pulmonologists and either accepted or rejected with no other information. Data was analyzed by Kappa statistic to judge inter-observer variance and it was compared to actual outcome to determine predictive value.

Results: The Kappa statistics for observer agreement was 0.149 among the surgeons, 0.510 among the pulmonologists, and 0.336 overall, representing slight, moderate and fair agreement respectively. The reviewers’ decisions to accept or reject a lung concurred with the actual clinical outcome 64.2% of the time. The positive predictive value of an accept decision was found to be 78.3% and the negative predictive value of a reject decision was 36.3%.

Conclusions: This study suggests that evaluation of the donor chest x-ray is a highly subjective process and demonstrated the limited role the radiograph presently holds in the determination of organ suitability.

\( \text{q} \) 2003 Elsevier Science B.V. All rights reserved.

Keywords: Lung transplant; Donor chest radiograph

1. Introduction

Information used in the evaluation of brain-dead organ donors for lung donation has typically come from multiple clinical and laboratory sources. These include arterial blood gases, viral serologies, ABO blood type, sputum gram stain and cultures, flexible bronchoscopy, and chest radiographs (CXR). Additionally, the donor’s medical history, and mechanism of injury are taken into account, and a final visual and manual evaluation of the lungs is performed at the time of organ harvest \cite{1,2}.

In the current era of aggressive acceptance of potential organ donors, the importance of chest radiography is unclear. This is in part due to the fact that the relationship between the chest radiograph interpretation and the final disposition of donor lungs is unknown. Despite this, potential organ donors often receive numerous CXRs as part of their evaluation. At many institutions, it is not uncommon for a potential donor to receive multiple CXRs in the 24 h prior to organ harvest. This often delays procurement by adding confusion and tumult to the organ donation process with the requests for multiple CXRs.

Little is known about inter-observer variability of chest X-ray interpretation or the importance of a CXR in the decision to accept or reject an organ. We, therefore, performed a retrospective review of donors evaluated for lung transplantation in an effort to determine the relationship between donor chest X-rays and the acceptability of lungs for transplantation. Additionally, we sought to determine the level of inter-observer variability in the interpretation of donor CXRs within our group.
2. Methods

Over a 5-year period from 1995 to 2000, CXRs from all brain-dead patients evaluated for possible lung donation were routinely retrieved and stored for later evaluation by one of the transplant surgeons within our group. The preoperative CXRs of five unrelated patients who had undergone coronary artery bypass grafting were included with the donor radiographs as controls. The films were organized by patient and the patients were then sorted into a random order. All distinguishing patient and hospital information on the films were covered. The films were shown to six reviewers (three cardiothoracic surgeons and three pulmonologists) all of whom were actively involved in our lung transplantation program. The reviewers were informed that all patients were candidates for lung donation based on standard criteria and were asked to accept or reject each lung based solely on the appearance of the chest X-rays. Standard lung transplant donor criteria include the following: age < 55 years; ABO blood group compatibility; PaO\(_2\) ≥ 300 mmHg on fractional inspired oxygen of 1.0 and positive end-expiratory pressure > 5 cm H\(_2\)O; ≤ 20 pack-year smoking history; absence of chest trauma; no aspiration or sepsis; and gram stain shows sputum sample free of bacteria, fungus and significant number of WBCs. Each review occurred independently and in an isolated setting. Each was completed in a similar length of time using only an illuminated light box.

Data from the review of all radiographs, including those from patients who were not actual organ donors, were analyzed for inter-observer variability by the kappa statistic method for multiple reviewers, as outlined by Fleiss [3]. In summary, kappa is defined as the agreement beyond chance divided by the amount of agreement possible beyond chance: \(\kappa = (O - C)/(1 - C)\), where \(O\) is the observed agreement and \(C\) is the chance agreement. When \(\kappa\) is zero, agreement is only at the level expected by chance. When \(\kappa\) is 1, agreement has no element of chance.

For patients who were actual organ donors, we searched the records of the local organ procurement organization for follow-up information on organ disposition. Patients on whom disposition information was available were sorted into a sub-group for separate analysis. The reviewer’s evaluations were compared with the actual outcome to determine the accuracy of their prediction and to compute a negative predictive value and positive predictive value, per the common statistical method [4]. Each lung was evaluated by the reviewers independently and was analyzed as a distinct statistical event.

3. Results

A total of 89 patients (178 lungs) were reviewed. Of these, 84 were candidates for organ donation and five were not potential donors (controls). There was an average of two films per patient, and they covered periods ranging from 12 to 72 h following admission. A total of ten of the potential organ donor patients were eliminated from the final sub-group analysis because organ disposition information was either unclear or the organs had not been utilized secondary to recipient difficulties.

Donor acceptance was looked at from the standpoint of donors and from organs. There was complete agreement in 50/89 donors (56%). Both lungs were accepted by all reviewers in 44/89 donors (49%) (Fig. 1) and no lungs were accepted by all reviewers in 6/89 donors (7%) (Fig. 2). In the remaining 39 donors, there were wide variations in acceptance (Fig. 3). The most aggressive reviewer accepted 83/89 donors (93%) and the least aggressive reviewer accepted 48/89 (54%). In the 39 donors in whom there was no consensus, 19 were accepted by only one reviewer, 15

![Fig. 1. An example of a donor chest radiograph that was accepted by all reviewers.](image-url)
were accepted by two reviewers, three were accepted by four reviewers and two were accepted by five reviewers.

From the viewpoint of donor lungs, there was complete agreement on 50 donors. In 44 donors, both lungs were accepted and six had neither lung accepted. In the remaining 39 donors, there was wide variation in the acceptance of one or both lungs. One reviewer accepted both lungs in 82/89 donors (93%) and one lung in 1/89 donors (1.1%) for a total lung acceptance rate of 165/178 lungs (93%). Another reviewer accepted both lungs in 46/89 (52%) and one lung in two donors for a total lung acceptance rate of 94/178 (53%). Reviewers decided that one lung was acceptable in 13/89 donors (15%). This, however, varied amongst reviewers.

Kappa values for inter-observer variability are displayed in Table 1. Kappa values for the inter-observer variability between the transplant surgeons were 0.195 for all radiographs and 0.149 for the radiographs used in the final analysis of predictive value. Kappa values among the pulmonologists were higher: 0.532 in the large group and 0.510 in the final sub-group. Our transplant group as a whole had a kappa value of 0.377 for the whole data set and 0.336 for the final sub-group. As outlined in Table 2, Landis and Koch propose that kappa values of 0–0.2 represent slight agreement, 0.21–0.4 represent fair agreement, 0.41–0.6 represent moderate agreement, 0.61–0.8 represent substantial agreement, and 0.81–1.0 represent almost perfect agreement between observers [5].

Accuracy of the reviewers’ decision to accept or reject the organ as compared to the actual outcome is outlined in Table 3. The reviewers’ decisions to accept or reject an organ were in agreement with the actual clinical outcome on an average of 64.2% of the time, with a range of 60.2–74.2%. The positive predictive value of an accept decision, i.e. the likelihood that an organ accepted based upon chest X-ray would actually be harvested, ranged from 75.4 to 80.0% among the individual evaluators with an average of 78.3%. The negative predictive value of a reject decision, i.e. the likelihood that an organ rejected based upon chest X-ray would not be used for transplantation, ranged from 27.7 to 50.0% among the different evaluators with an average of 36.3%.

4. Discussion

Any clinical measurement, be it qualitative or quantitative, is subject to some variability, both between different observers and between serial observations made by the same observer. Qualitative assessments, such as the interpretations of radiographs, are particularly susceptible to variability and because of this, many in our group have wondered if the chest radiograph really held any weight in the lung procurement process, or if other, perhaps less subjective information was the really being used to make decisions about lung suitability.

This study confirmed our suspicions about the variability inherent in interpreting donor CXRs. Within our group, inter-observer agreement among the surgeons would be classified as only slight, among the pulmonologists a slightly more respectable moderate, but for our group

<table>
<thead>
<tr>
<th>Kappa value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–0.2</td>
<td>Slight agreement between observers</td>
</tr>
<tr>
<td>0.21–0.4</td>
<td>Fair agreement between observers</td>
</tr>
<tr>
<td>0.41–0.6</td>
<td>Moderate agreement between observers</td>
</tr>
<tr>
<td>0.61–0.8</td>
<td>Substantial agreement between observers</td>
</tr>
<tr>
<td>0.81–1.0</td>
<td>Almost perfect agreement between observers</td>
</tr>
</tbody>
</table>
Table 3

<table>
<thead>
<tr>
<th>Percentage of lungs where evaluator’s decision agreed with outcome</th>
<th>Range (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive predictive value of an evaluator’s accept decision</td>
<td>75.4–80.0</td>
<td>78.3</td>
</tr>
<tr>
<td>Negative predictive value of an evaluator’s reject decision</td>
<td>27.7–50.0</td>
<td>36.3</td>
</tr>
</tbody>
</table>

Overall there was only fair agreement between observers. The comparison of the evaluators’ decisions and the actual outcome (Table 3), also confirms our view that evaluation of the radiograph does not predict to any large degree whether a lung will actually be used – in fact it does so only 64% of the time, and if a chest radiograph is unacceptable the lung will still be used two-thirds of the time based on our experience.

One reason for the lackluster role of the radiograph in lung evaluation may be the number of successful transplants that have been undertaken with lungs that were radiographically flawed. Sundaresan et al. [6] published a series of 44 donors who did not meet classical clinical criteria but were harvested anyway, and of these patients 34 had an ‘unsatisfactory’ chest radiograph. As compared to a control group whose donors did meet classical criteria, there was no difference in graft function or 30-day mortality, with the only negative reported outcome being an increased need for bypass support when conducting bilateral sequential transplants [6].

Even though the importance of some defects on chest X-ray have been proven to be minimal, its interpretation is fraught with variability, and it really does not play a major role in the final lung procurement decision, many transplant centers continue to perseverate over the appearance of the chest radiograph and will obtain frequent and numerous films. While we believe that the radiograph is useful to screen out major pathology and perhaps monitor volume status, we urge those evaluating lungs for procurement to recognize its relatively minor importance.

References