Video-assisted thoracic surgery versus open thoracotomy for non-small cell lung cancer: a meta-analysis of propensity score-matched patients

Christopher Cao a,b,c, Con Manganas b, Su C. Ang a,b, Sheen Peeceeyen b and Tristan D. Yan a,c,d,*

Abstract

OBJECTIVES: This meta-analysis aimed to compare the perioperative outcomes of video-assisted thoracic surgery (VATS) with open thoracotomy for propensity score-matched patients with early-stage non-small cell lung cancer (NSCLC).

METHODS: Four relevant studies with propensity score-matched patients were identified from six electronic databases. Endpoints included perioperative mortality and morbidity, individual postoperative complications and duration of hospitalization.

RESULTS: Results indicate that all-cause perioperative mortality was similar between VATS and open thoracotomy. However, patients who underwent VATS were found to have significantly fewer overall complications, and significantly lower rates of prolonged air leak, pneumonia, atrial arrhythmias and renal failure. In addition, patients who underwent VATS had a significantly shorter length of hospitalization compared with those who underwent open thoracotomy.

CONCLUSIONS: In view of a paucity of high-level clinical evidence in the form of large, well-designed randomized controlled trials, propensity score matching may provide the highest level of evidence to compare VATS with open thoracotomy for patients with NSCLC. The present meta-analysis demonstrated superior perioperative outcomes for patients who underwent VATS, including overall complication rates and duration of hospitalization.

Keywords: Video-assisted thoracic surgery • Thoracotomy • Non-small cell lung cancer • Propensity score analysis • Meta-analysis

INTRODUCTION

The introduction of minimally invasive surgery has revolutionized many disciplines of surgical practice in recent decades. The Cancer and Leukemia Group B (CALGB) 39 802 prospective, multi-institutional study examined the feasibility of standardized video-assisted thoracic surgery (VATS) lobectomy for early-stage non-small cell lung cancer (NSCLC) and demonstrated that this technique was associated with a low complication rate and short intercostal drainage duration [1]. Although the safety of performing VATS lobectomy for NSCLC is generally considered to be acceptable, its utilization has spread slowly over the past decade [2].

Currently, there is limited robust clinical data to suggest superior outcomes for VATS compared with open thoracotomy for patients with NSCLC. Two early randomized trials comparing VATS lobectomy with open surgery had a number of limitations and did not provide data on long-term outcomes [3, 4]. In addition, the use of rib-spreading in these studies would exclude them from being ‘true’ VATS lobectomies according to the currently accepted definition [1]. In a recent meta-analysis, VATS lobectomy demonstrated a reduced systemic recurrence rate and an improved 5-year survival mortality rate when compared with open lobectomy [5]. However, most comparative studies included in this review were not well-matched cohorts. Significant differences in baseline patient characteristics have hindered direct comparisons between VATS and open thoracotomy. In view of the lack of randomized controlled trials comparing VATS lobectomy under the strict CALGB definitions with open lobectomy [2], we performed a meta-analysis on all studies involving propensity score-matched patients with NSCLC who underwent complete VATS lobectomy vs open lobectomy. Propensity score matching is considered to strengthen outcome measures in a number of ways. Most importantly, it enables investigators to retrospectively assemble a study cohort in which patients are balanced in all observed significant covariates. This makes the assessment of the intervention more accurate by minimizing the potential bias between the comparative groups [6–8]. The primary assessed outcomes were overall perioperative mortality and morbidity, with individual complications and duration of hospitalization as secondary outcomes.
PATIENTS AND METHODS

Literature search strategy

Electronic searches were performed using Ovid Medline, PubMed, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), ACP Journal Club and Database of Abstracts of Review of Effectiveness (DARE) from their date of inception to April 2012. To achieve the maximum sensitivity of the search strategy and to identify all studies, we combined the terms ‘video-assisted thoracic surgery’ or ‘VATS’ or ‘thoracoscopic surgery’ with ‘propensity’ or ‘propensity score’ or ‘propensity match’ as key words or MeSH terms. The reference lists of all retrieved articles were reviewed for further identification of potentially relevant studies. Eligible comparative studies for the present meta-analysis included those in which perioperative data were available for 1:1 propensity score-matched patients with NSCLC who underwent VATS or open thoracotomy. Abstracts, case reports, conference presentations, editorials and expert opinions were excluded. Studies on robotic video-assisted thoracic surgery were excluded. Review articles are omitted due to potential publication bias and possible duplication of results.

All data were extracted from article texts, tables and figures. Two investigators (C.C. and S.C.A.) independently reviewed each retrieved article. Discrepancies between the two reviewers were resolved by discussion and consensus. The final results were reviewed by senior investigators (C.M. and T.D.Y.).

Statistical analysis

The propensity score is the conditional probability of receiving an intervention given the individual’s measured covariates. Propensity score matching is considered to significantly strengthen observational studies [6, 8]. However, it should be acknowledged that propensity score matching has its own limitations and does not adequately substitute for well-designed, randomized controlled trials. Patients included in the present study were individually matched according to their propensity scores in selected studies comparing VATS with open thoracotomy for patients with NSCLC. Meta-analysis was performed by combining the results of reported incidences of postoperative morbidity, postoperative mortality, individual postoperative complications and duration of hospitalization when comparable outcomes were available for analysis. The relative risk (RR) was used as a summary statistic. In the present study, the random effect models were tested, where it was assumed that there were variations between studies and the calculated ratios, and thus a more conservative value was obtained. \( \chi^2 \) tests were used to study heterogeneity between trials. The \( I^2 \) statistic was used to estimate the percentage of total variation across studies, due to heterogeneity rather than chance. \( I^2 \) can be calculated as:

\[
I^2 = 100\% \times \frac{(Q - df)}{Q},
\]

with \( Q \) defined as Cochrane’s heterogeneity statistics and \( df \) defined as degree of freedom [9]. An \( I^2 \) value >50\% was considered substantial heterogeneity. If there was substantial heterogeneity, the possible clinical and methodological reasons for this were qualitatively explored. In the present meta-analysis, the results using the random effects model were presented to take into account the possible clinical diversity and methodological variation among studies. Specific analyses considering confounding factors were not possible because raw data were not available. All \( P \)-values were two-sided. All statistical analysis was conducted with Review Manager Version 5.1.2 (Cochrane Collaboration, Software Update, Oxford, UK).

RESULTS

Quantity and quality of trials

A total of 19 references were identified through the six electronic database searches. After exclusion of duplicate or irrelevant references, seven potentially relevant articles were retrieved for more detailed evaluation [10–16]. After applying the selection criteria, four comparative studies remained for assessment [10–13]. A manual search of the reference lists did not identify any additional relevant studies. All four studies included for final analysis in the present meta-analysis were from retrospective observational studies, as summarized in Table 1. In these four studies, 3634 patients with NSCLC were compared, including 1817 who underwent VATS and 1817 propensity score-matched patients who underwent open thoracotomy.

Assessment of overall perioperative mortality and morbidity

The overall perioperative mortality rate was not significantly different between patients who underwent VATS when compared with propensity score-matched patients who underwent open thoracotomy (1.4 vs 2.0\%; RR, 0.75; 95\% confidence interval [CI] 0.44–1.27; \( P = 0.28 \); \( I^2 = 0\% \)). These results are summarized in Fig. 1. The overall perioperative morbidity rate was significantly lower in patients who underwent VATS when compared with propensity score-matched patients who underwent open thoracotomy (24.9 vs 20.2\%; RR, 0.67; 95\% CI 0.56–0.82; \( P < 0.0001 \); \( I^2 = 48\% \)). These results are summarized in Fig. 2.

Assessment of postoperative complications

A number of individual perioperative morbidities were comparable between different studies involving propensity score-matched patients. These included prolonged air leak, pneumonia, pulmonary embolism, atrial arrhythmia, myocardial infarction, significant bleeding, empyema, sepsis and acute renal failure. Meta-analysis identified significantly lower incidences of prolonged air leak (8.1 vs 10.4\%; RR, 0.78; 95\% CI 0.63–0.96; \( P = 0.02 \); \( I^2 = 0\% \)), pneumonia (3.2 vs 5.0\%; RR, 0.65; 95\% CI 0.47–0.89; \( P = 0.008 \); \( I^2 = 0\% \)), atrial arrhythmia (7.3 vs 11.7\%; RR, 0.62; 95\% CI 0.51–0.77; \( P < 0.0001 \); \( I^2 = 0\% \)) and renal failure (0.9 vs 3.0\%; RR, 0.32; 95\% CI 0.12–0.88; \( P = 0.03 \); \( I^2 = 0\% \)) for patients who underwent VATS when compared with propensity score-matched patients who underwent open thoracotomy. The incidences of pulmonary embolism (0.3 vs 0.4\%; RR, 0.85; 95\% CI 0.26–2.86; \( P = 0.80 \); \( I^2 = 0\% \)), myocardial infarction (0.2 vs 0.1\%; RR, 1.35; 95\% CI 0.25–7.15; \( P = 0.73 \); \( I^2 = 0\% \)), significant bleeding (1.0 vs 0.8\%; RR, 1.08; 95\% CI 0.36–3.21; \( P = 0.89 \); \( I^2 = 33\% \)), empyema (0.3 vs 0.6\%; RR, 0.70; 95\% CI 0.10–5.01; \( P = 0.72 \); \( I^2 = 54\% \)) and sepsis (0.5 vs 1.0\%; RR, 0.55; 95\% CI 0.23–1.32; \( P = 0.18 \); \( I^2 = 0\% \)) were not significantly different.
between the two treatment groups. A summary of the significant findings are presented in Fig. 3a–d.

Assessment of the length of stay

The duration of hospitalization was calculated from the date of surgery, and statistically comparable data were available from three studies [10, 12, 13]. Meta-analysis identified a significantly shorter length of stay for patients who underwent VATS compared with open thoracotomy (standard mean difference 95% CI −0.37; 95% CI −0.51 to −0.22; \( P < 0.00001; \) \( I^2 = 48\% \)). These results are summarized in Fig. 4.

<table>
<thead>
<tr>
<th>Institutions</th>
<th>Author</th>
<th>Reference number</th>
<th>Publication year</th>
<th>Study period</th>
<th>Study type</th>
<th>n VATS</th>
<th>n Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helsinki University Central Hospital, Finland</td>
<td>Ilonen*</td>
<td>[17]</td>
<td>2011</td>
<td>2000–2010</td>
<td>ROS</td>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>MSKCC, USA</td>
<td>Flores</td>
<td>[18]</td>
<td>2009</td>
<td>2002–2007</td>
<td>ROS</td>
<td>313</td>
<td>313</td>
</tr>
</tbody>
</table>

ROS: retrospective observational study.
*Study included for detailed statistical analysis.

Figure 1: Forest plot of the RR of postoperative mortality after VATS vs open thoracotomy for NSCLC. The estimate of the RR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. On each line, the numbers of events as a fraction of the total number allocated are shown for both treatment groups. For each subgroup, the sum of the statistics, along with the summary RR, is represented by the middle of the solid diamonds. A test of heterogeneity between the trials within a subgroup is given below the summary statistics.

Figure 2: Forest plot of the RR of postoperative morbidity after VATS vs open thoracotomy for NSCLC. The estimate of the RR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. On each line, the numbers of events as a fraction of the total number allocated are shown for both treatment groups. For each subgroup, the sum of the statistics, along with the summary RR, is represented by the middle of the solid diamonds. A test of heterogeneity between the trials within a subgroup is given below the summary statistics.

DISCUSSION

Since the technique of VATS lobectomy was first described in the early 1990s, a number of studies have demonstrated superior perioperative outcomes for this procedure when compared with conventional thoracotomy, including reduced incidence of arrhythmias, pneumonia, pain and lower levels of inflammatory markers [17–19]. With increasing experience and technological innovation, a number of surgical approaches have been developed in numerous specialized centres [20, 21]. Despite encouraging results for patients with NSCLC who underwent VATS, the acceptance of this procedure remains controversial, and conflicting evidence exists in the form of large retrospective studies that
failed to show any significant benefit from VATS [22]. Ideally, well-designed, large-scale, multi-institutional randomized controlled trials utilizing the current CALGB definition of VATS lobectomy should be conducted to reach a consensus on this controversial issue. However, recruitment challenges and ‘lack of equipoise’ have been cited in recent studies as reasons limiting such trials [2, 15, 23]. Skepticism within the surgical community is reflected by the relatively low proportion of VATS lobectomy being performed [2, 23].

One of the critical challenges in examining the current literature on VATS for NSCLC is the heterogeneity of patient selection and surgical technique between institutions. To address the first issue, attempts have been made to crudely match patients according to risk factors such as age, gender, comorbidities and clinical staging [24]. Recently, more advanced statistical methods have been utilized in the form of propensity score matching to compare short- and long-term outcomes for VATS with open thoracotomy [10–14]. In addition, there was a lack of

---

**Figure 3:**

(a) Forest plot of the RR of postoperative prolonged air leak after VATS vs open thoracotomy for NSCLC. The estimate of the RR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. On each line, the numbers of events as a fraction of the total number allocated are shown for both treatment groups. For each subgroup, the sum of the statistics, along with the summary RR, is represented by the middle of the solid diamonds. A test of heterogeneity between the trials within a subgroup is given below the summary statistics.

(b) Forest plot of the RR of postoperative pneumonia after VATS vs open thoracotomy for NSCLC. The estimate of the RR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. On each line, the numbers of events as a fraction of the total number allocated are shown for both treatment groups. For each subgroup, the sum of the statistics, along with the summary RR, is represented by the middle of the solid diamonds. A test of heterogeneity between the trials within a subgroup is given below the summary statistics.

(c) Forest plot of the RR of postoperative atrial arrhythmias after VATS vs open thoracotomy for NSCLC. The estimate of the RR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. On each line, the numbers of events as a fraction of the total number allocated are shown for both treatment groups. For each subgroup, the sum of the statistics, along with the summary RR, is represented by the middle of the solid diamonds. A test of heterogeneity between the trials within a subgroup is given below the summary statistics.

(d) Forest plot of the RR of postoperative renal failure after VATS vs open thoracotomy for NSCLC. The estimate of the RR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. On each line, the numbers of events as a fraction of the total number allocated are shown for both treatment groups. For each subgroup, the sum of the statistics, along with the summary RR, is represented by the middle of the solid diamonds. A test of heterogeneity between the trials within a subgroup is given below the summary statistics.
standardization in defining the VATS lobectomy procedure among thoracic surgeons. Variations in surgical techniques included differences in the use of a rib-spreader, the anatomic isolation vs mass stapling of lobar hilum, the size of the incision, the use of endoscopic instruments vs conventional instruments and visualization through the thorax or only via a monitor [5]. To clarify the definition of ‘true’ VATS, Swanson et al. [1] completed a prospective, multi-institutional study with a predefined VATS lobectomy procedure that mandated the use of videooscopic guidance with anatomical hilar dissection without the use of the rib spreader through a 4- to 8-cm incision site and two port incisions. This is the currently accepted definition of VATS lobectomy and carries the key points emphasized by the pioneers of VATS lobectomy to reduce surgical access trauma, filtering out ‘pseudo-VATS’ techniques that reported compromised results in the past [25].

In the present study, 3634 patients with NSCLC, including 1817 patients who underwent VATS and 1817 propensity score-matched patients who underwent open thoracotomy, were assessed for their perioperative outcomes. Meta-analysis found significantly fewer overall complications after VATS compared with open thoracotomy. Specifically, patients who underwent VATS were significantly less likely to develop prolonged air leak, pneumonia, atrial arrhythmias and renal failure compared with matched patients who underwent open thoracotomy. The duration of hospitalization was also significantly shorter after VATS, and there was a trend towards a lower perioperative mortality rate for patients in this treatment group, but this finding was not statistically significant.

Some limitations in the present study need to be acknowledged. First, the non-randomized nature of the treatment allocation process still necessitates cautious interpretation of the results. Although propensity score matching strengthened this study, it does not replace randomization. Second, patient selection based on preoperative investigations and patients’ baseline characteristics differed between institutions, and there was inevitably some variability in the surgical techniques and skills of surgeons. However, this can also be considered advantageous as the combined experience of a collective group of surgeons may be more representative of VATS procedures in the ‘real world’ than the skills and management of an individual surgeon [15].

This meta-analysis involved propensity score-matched patients with NSCLC to compare VATS with open thoracotomy. VATS was found to have significantly fewer overall complications compared with open thoracotomy, but perioperative mortality was not significantly different between the two groups. In the absence of randomized controlled trials comparing standardized VATS lobectomy with conventional open lobectomy, our findings represent the highest level of clinical evidence in the current literature on this issue.

**Conflict of interest:** none declared.

**REFERENCES**


