Heterogeneity of Lung Volume Reduction Surgery Outcomes in Patients Selected by Use of Evidence-Based Criteria

Matthew R. Lammi, MD, Nathaniel Marchetti, DO, Shari Barnett, MD, and Gerard J. Criner, MD

Department of Internal Medicine, Section of Pulmonary and Critical Care Medicine, Louisiana State University Health Sciences Center, New Orleans, Louisiana; and Department of Internal Medicine, Section of Pulmonary and Critical Care Medicine, Temple University School of Medicine, Philadelphia, Pennsylvania

Background. Despite its benefit, lung volume reduction surgery (LVRS) is underused, partially because of the heterogeneous responses and lack of recent outcomes data.

Methods. Data from 59 consecutive emphysema patients who underwent LVRS were analyzed. The proportion of patients responding based on 6-minute walk distance (6-MWD), exercise capacity (watts), and forced expiratory volume in 1 second (FEV1) were calculated. Baseline variables were correlated with improvements in 6-MWD, maximal watts, and FEV1.

Results. Eighty-eight percent of patients responded to LVRS, with a higher proportion of FEV1 and 6-MWD responders in our cohort compared with similar patients from the National Emphysema Treatment Trial. Significant associations existed between lower baseline 6-MWD and increased 6-MWD after operation ($r = 0.423$), more extensive emphysema and increased FEV1 ($r = 0.491$), and hyperinflation and increased maximal watts ($r = 0.438$). The probability of survival was 0.93 at 90 days, 0.90 at 1 year, and 0.80 (3 years). The lowest exercise group (<20 watts on baseline testing) had an increased risk for death (RR 13.3, $p = 0.001$).

Conclusions. There were durable improvements in FEV1 and exercise capacity in patients meeting the National Emphysema Treatment Trial criteria. Survival was comparable to that in similar patients from the National Emphysema Treatment Trial; response rates were higher in our cohort for FEV1 and 6-MWD. Those with lower 6-MWD, more emphysema, and more hyperinflation at baseline were most likely to respond to LVRS. Those with lowest exercise capacity at baseline may have a higher risk of death after LVRS.


Lung volume reduction surgery (LVRS) has been shown to improve exercise capacity, quality of life, and mortality in selected patients with severe emphysema [1]. Despite these favorable outcomes, LVRS continues to be underused in the United States, with recent data showing declining numbers of operations per year since 2004 [2]. Outcomes data of patients undergoing LVRS after the National Emphysema Treatment Trial (NETT) [1] are needed, with the only such published report showing excellent results [3]. Also, when data are presented as mean values, information may be difficult to interpret because of the large heterogeneity in response to LVRS. If some patients do well with an intervention and some do not, aggregated data may miss a group of patients with high benefit or differentially poorer outcomes.

One factor that may limit the widespread adoption of LVRS is the large variation in response to the procedure. Besides descriptions from NETT of the original four subgroups and the high-risk-for-death group [4], attempts to predict response with baseline variables have been inconclusive [5–7]. To explore the heterogeneity in response that exists even in patients selected to have high benefit on the basis of the NETT criteria, we conducted a retrospective analysis of patients undergoing LVRS at a single institution after Centers for Medicare and Medicaid Services (CMS) approval. Response rates were measured, and associated variables were analyzed.

Patients and Methods

Patient Selection

Fifty-nine consecutive patients with severe emphysema who underwent LVRS at Temple University Hospital after CMS approval (from 2004 to 2011) were retrospectively analyzed. All patients met the NETT inclusion and exclusion criteria [1] and had upper lobe–predominant emphysema based on visual assessment of the computed tomographic (CT) scan. Before undergoing LVRS, all patients received 8 weeks of outpatient pulmonary rehabilitation and were using maximal medical therapy. Patients were stratified into high or low exercise capacity based on the results of their post-rehabilitation
maximal cardiopulmonary exercise test (low exercise = less than 25 watts of maximal exercise capacity for women and less than 40 watts for men) [1]. This work was approved by the Internal Review Board of Temple University (protocol #20046).

**Surgical Approach**
All patients underwent bilateral lung volume reduction through a median sternotomy. The same experienced thoracic surgeon performed all of the surgical procedures, and the goal was to resect 20% to 30% of the most diseased portion of both lungs.

**Study Objective**
The main objective of the present study was to assess the outcomes of LVRS in patients selected to be of highest benefit based on the NETT criteria.

**Variables Collected**
All patients had postrehabilitation (preoperative) testing of pulmonary function, 6-minute walk distance (6-MWD), and cardiopulmonary exercise capacity. These same variables were routinely repeated at 6 months and then yearly after the procedure.

Pulmonary function testing was performed according to American Thoracic Society/European Respiratory Society guidelines [8] using the reference values of the National Health and Nutrition Examination Survey [9]. Postbronchodilator values for forced vital capacity (FVC), forced expiratory volume in 1 second (FEV₁), total lung capacity (TLC), and residual volume (RV) were used in the subsequent analyses. Lung volumes were measured by body plethysmography [10], and diffusion capacity for carbon monoxide (DLCO) [11] was measured with standard techniques. The 6-MWD test was conducted without encouragement in an L-shaped hallway [12]. The distance walked in 6 minutes was measured in meters. Cardiopulmonary exercise testing (CPET) was performed on a braked cycle ergometer (ViaSprint 150P; ViaSys Healthcare; Hoechberg, Germany) according to American Thoracic Society/European Respiratory Society guidelines [13] according to the protocol used in NETT [14]. After a period of baseline rest and unloaded warmup pedaling, a symptom-limited maximal exercise phase was performed, which consisted of increasing levels of tension on the bike at a rate of 5 or 10 watts per minute until exhaustion.

Emphysema was quantified on each patient’s preoperative CT scan with Slicer software [15]. Areas of emphysema were characterized as having low attenuation, defined as ≤500 Hounsfield units [16]. The percentage of emphysema in each patient was calculated. Values were also stratified on the basis of lung field, which was split into upper, middle, and lower lung fields.

**Data Analysis**
Comparisons were made between the cohort described here and the upper lobe–predominant surgical arm from NETT [1, 17]. Continuous variables were compared with unpaired t tests. Changes in variables over time were compared by one-way analysis of variance with repeated measures in a random effects model. Post-hoc Tukey’s test was used to compare pairs of time points.

To explore the heterogeneity of response, a responder analysis was performed. Responders were defined a priori on the basis of three distinct criteria 6 months after LVRS: an increase in 6-MWD ≥54 meters [18], FEV₁ improvement of ≥12% and 200 mL [8], or an increase of ≥10 watts on maximal CPET [1]. The proportion of patients labeled as responders for each criterion was

<table>
<thead>
<tr>
<th>Variable</th>
<th>Temple cohort (n = 59)</th>
<th>NETT-LVRS Low Exercise (n = 139)</th>
<th>NETT-LVRS High Exercise (n = 206)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (% male)</td>
<td>44.1%</td>
<td>55%</td>
<td>57%</td>
</tr>
<tr>
<td>Age (years)</td>
<td>66 ± 6.5</td>
<td>67.2 ± 5.2</td>
<td>66.6 ± 6.4</td>
</tr>
<tr>
<td>Smoking history (pack-years)</td>
<td>63 ± 32.2</td>
<td>64.8 ± 29.2a</td>
<td></td>
</tr>
<tr>
<td>Oxygen use (% yes)</td>
<td>52.5%</td>
<td>52%a</td>
<td></td>
</tr>
<tr>
<td>FEV₁ (% predicted)</td>
<td>29.5 ± 8.3</td>
<td>25.4 ± 6.8</td>
<td>29.5 ± 6.5</td>
</tr>
<tr>
<td>TLC (% predicted)</td>
<td>125 ± 17</td>
<td>128 ± 14</td>
<td>125 ± 14</td>
</tr>
<tr>
<td>RV (% predicted)</td>
<td>199 ± 41</td>
<td>224 ± 43</td>
<td>207 ± 42</td>
</tr>
<tr>
<td>DLCO (% predicted)</td>
<td>32 ± 8</td>
<td>26 ± 8</td>
<td>31 ± 8</td>
</tr>
<tr>
<td>% Emphysema</td>
<td>33.9 ± 9.1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>% Emphysema, upper lung field</td>
<td>50.5 ± 13.1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low exercise group (%)</td>
<td>37%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* Values for NETT LVRS group combined (high and low exercise).

All values are mean ± standard deviation unless indicated.

Values obtained after rehabilitation.

CT = computed tomography; DLCO = diffusion capacity for carbon monoxide; FEV₁ = forced expiratory volume in 1 second; LVRS = lung volume reduction surgery; NETT = National Emphysema Treatment Trial; RV = residual volume; TLC = total lung capacity.
quantified for all time points of follow-up. This was compared with the response rate from the NETT upper lobe–predominant surgical patients by use of Fisher’s exact test. To explore the associations between baseline variables (FVC, FEV₁, RV, TLC, 6-MWD, maximal watts, and percentage of emphysema on quantitative CT) and changes in 6-MWD, FEV₁, and maximal watts, Spearman correlations were performed.

A Kaplan-Meier survival analysis was performed. Comparisons in survival between the high and low exercise groups were conducted with the log rank (Mantel-Cox) test. Characteristics of the low exercise group who died were compared with those of the low exercise group who survived beyond 3 years by use of unpaired t tests in a post hoc fashion.

## Results

### Baseline Characteristics

Fifty-six percent of patients in our cohort were female, with an overall mean age of 66 ± 6.5 years (Table 1). There was no difference in the gender distribution between our cohort and that of NETT (p = 0.09). Because our patients were selected by use of the NETT criteria, they had similar pulmonary function compared with the NETT upper lobe–predominant surgical group (Table 1). Quantitative CT analysis confirmed upper lobe predominance of emphysema for all patients. Thirty-seven percent of patients in our cohort were in the low exercise group.

### Response Over Time

In comparison with the postrehabilitation baseline, there was a significant increase in mean 6-MWD at 6 and 12 months (Fig 1A). Maximal watts on cardiopulmonary exercising testing were significantly increased up to 24 months (Fig 1B). The FVC and FEV₁ significantly improved from baseline up to 36 months (Fig 1C).

### Responder Analysis

In comparison with the upper lobe surgical patients from NETT, our cohort had a higher proportion of 6-MWD responders at 6 and 12 months (Table 2). There was a higher proportion of FEV₁ responders in our cohort than in the NETT group at 6, 12, 24, and 36 months. There was no difference in the proportion of patients responding in terms of maximal watts at any time point.

The majority of our patients responded at 6 months either based on FEV₁ alone or based on all three variables (FEV₁, 6-MWD, and maximal watts) (Table 3). Only

---

**Fig 1.** (A) Values before and after surgical procedure for 6-minute walk distance, (B) maximal watts on cardiopulmonary exercise testing, and (C) FEV₁ for our cohort (“Temple”) and the upper lobe–predominant surgical group from the National Emphysema Treatment Trial (NETT). Data are presented as mean ± SEM. *A statistically significant improvement from the postrehabilitation to postsurgical variables in our cohort.*
5 (12%) patients in our cohort did not respond on the basis of any of the three criteria. When the association between baseline variables and changes in variables after LVRS was assessed, there was a significant correlation between a lower baseline 6-MWD and a greater increase in 6-MWD 6 months after LVRS ($r = -0.423$, $p = 0.006$). Those with a higher percentage of emphysema at baseline (both global and upper lung field emphysema) tended to have a greater increase in FEV$_1$ after LVRS ($r = 0.491$, $p = 0.002$ for upper lung field percentage of emphysema). A higher baseline TLC correlated with a larger increase in maximal watts at 6 months ($r = 0.438$, $p = 0.008$). No other variables tested showed a significant association with change in 6-MWD, FEV$_1$, or maximal watts.

**Survival**

In our cohort, the probability of survival was 0.93 (95% confidence interval [CI] 0.83 to 0.97) at 90 days, 0.90 (95% CI 0.79 to 0.95) at 1 year, 0.80 (95% CI 0.66 to 0.88) at 3 years, and 0.69 (95% CI 0.55 to 0.80) at 5 years. Survival curves for the upper lobe–predominant, low exercise group (Fig 2A) and the upper lobe–predominant, high exercise group (Fig 2B) appeared similar to those obtained from the long-term follow-up of NETT [17].

**High vs Low Exercise Survival**

In our cohort, there was no difference in 5-year survival in the high versus low exercise groups ($p = 0.91$ by log rank test) (Fig 3). When the baseline characteristics of the 6 patients in the low exercise group who died within 3 years were compared with the 16 low exercise patients who survived beyond 3 years, there was a difference in baseline maximal exercise capacity. The mean maximal watts in the group who died was 17.3 ± 9.5 compared with 25.3 ± 8.8 in those who survived ($p = 0.04$) (Fig 4). The relative risk for death in the below 20 watts group was 13.3 (95% CI 1.9 to 92.0, $p = 0.001$). All 5 patients who died in this subgroup were female. When extended to the entire cohort (both high and low exercise groups), the RR for death in the below 20 watts group versus the greater than or equal to 20 watts group was 19.6 (95% CI 2.5 to 152.4, $p = 0.0008$). There were no differences in baseline pulmonary function tests, 6-MWD, percentage of emphysema, echo parameters (left ventricular ejection fraction, diastolic dysfunction, right ventricular dysfunction), albumin, or use of oxygen or prednisone in those in the low exercise group who died compared with those who survived more than 3 years (data not shown).

**Comment**

The major finding of the current study is that there is a heterogeneous response to LVRS, even in a group of emphysema patients carefully selected to be highly benefitted on the basis of the NETT classification. However, 88% of patients in our cohort responded at 6 months when response was defined in terms of 6-MWD, FEV$_1$, or maximal exercise capacity. Although survival seemed comparable with that of similar patients from NETT, those with the lowest exercise capacity were at a higher risk for death after LVRS.

In the United States, LVRS is vastly underused. A recent study reported that up to 15% of patients with advanced emphysema from a single center would qualify for LVRS [19]. However, data from the Centers for Medicare and Medicaid Services reveal that 216 LVRS procedures were performed in 2004 and that the number has declined each subsequent year until 2010, when 92 LVRS procedures were performed in the United States [2].

The reasons for this underuse are multifactorial, but one limitation to wider use may be the large heterogeneity in response. Even with the subgroup analysis from NETT outlining which patients should be most highly benefitted, it is difficult at times to predict who will respond to LVRS. Prior research focusing on physiologic variables, radiologic parameters, and phenotypic characterization of patients has yielded inconclusive results.

---

**Table 2. Proportion of Responders**

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Temple</th>
<th>NETT</th>
<th>$p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-MWD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 mo</td>
<td>24/52</td>
<td>77/276</td>
<td>0.01</td>
</tr>
<tr>
<td>12 mo</td>
<td>22/43</td>
<td>65/235</td>
<td>0.004</td>
</tr>
<tr>
<td>24 mo</td>
<td>14/42</td>
<td>44/203</td>
<td>0.11</td>
</tr>
<tr>
<td>FEV$_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 mo</td>
<td>29/43</td>
<td>111/291</td>
<td>0.0004</td>
</tr>
<tr>
<td>12 mo</td>
<td>29/42</td>
<td>84/244</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>24 mo</td>
<td>21/31</td>
<td>56/206</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>36 mo</td>
<td>12/21</td>
<td>36/126</td>
<td>0.01</td>
</tr>
<tr>
<td>Watts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 mo</td>
<td>18/36</td>
<td>105/280</td>
<td>0.15</td>
</tr>
<tr>
<td>12 mo</td>
<td>19/35</td>
<td>93/237</td>
<td>0.1</td>
</tr>
<tr>
<td>24 mo</td>
<td>10/24</td>
<td>68/194</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Response rates in the Temple cohort vs the NETT upper lobe surgical group. See Methods section for definitions of response.

FEV$_1$ = forced expiratory volume in 1 second; 6-MWD = 6-minute walk distance; NETT = National Emphysema Treatment Trial.

---

**Table 3. Type of Response**

<table>
<thead>
<tr>
<th>Responder Variable</th>
<th>No. of Responders</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV$_1$ only</td>
<td>10</td>
<td>24.4</td>
</tr>
<tr>
<td>FEV$_1$, 6-MWD, watts</td>
<td>9</td>
<td>22.0</td>
</tr>
<tr>
<td>FEV$_1$, 6-MWD</td>
<td>5</td>
<td>12.2</td>
</tr>
<tr>
<td>FEV$_1$, watts</td>
<td>4</td>
<td>9.8</td>
</tr>
<tr>
<td>6-MWD only</td>
<td>4</td>
<td>9.8</td>
</tr>
<tr>
<td>6-MWD, watts</td>
<td>2</td>
<td>4.9</td>
</tr>
<tr>
<td>Watts only</td>
<td>2</td>
<td>4.9</td>
</tr>
<tr>
<td>None</td>
<td>5</td>
<td>12.2</td>
</tr>
</tbody>
</table>

Number of patients classified as responding to the three variables in the Temple cohort at 6 months after lung volume reduction surgery (n = 41).

FEV$_1$ = forced expiratory volume in 1 second; 6-MWD = 6-minute walk distance.
In the current study, a lower baseline 6-MWD, higher percentage of emphysema on quantitative CT scan, and higher TLC were significantly associated with a greater improvement in 6-MWD, FEV₁, and maximal exercise capacity 6 months after the procedures, respectively. These routinely obtained baseline variables may be used to assist in patient selection and prognostication before LVRS is performed.

Even with patients selected to be highly benefited on the basis of the NETT experience, there still was a large heterogeneity in response to LVRS in our cohort. As seen in Table 3, there was a wide variation in response rate depending on the definition used. However, if clinically relevant variables such as changes in FEV₁, 6-MWD, and maximal exercise capacity are used to define response, only 12% of patients did not respond to at least one of the criteria. Additionally, almost half of our patients (20/41) responded to either two or all three of the predefined criteria.

Another reason for the underuse of LVRS is a poor perception of outcomes from the operation, which largely derives from publicity surrounding the high-risk-for-death group from NETT [20] and misinterpretation of data from NETT [21]. Additionally, outcomes data from centers performing LVRS after CMS approval have been scarce. As with any intervention, it is important to validate results outside of the context of a clinical trial. The only such reported outcomes data have come from Columbia University, with a report published in 2011 [3], which described 49 patients undergoing LVRS at a single academic center in whom a high probability of survival at 1 and 3 years was observed. There also were improvements in FEV₁, Medical Research Council dyspnea score, 6-MWD, and Body mass, Obstruction,
Dyspnea, Exercise capacity (BODE) index measured at 1 year.

To judge the outcomes of LVRS at our institution outside of a clinical trial, comparisons were made with patients in the NETT with upper lobe–predominant disease who were randomized to the surgical arm. Survival up to 7 years appeared comparable between our cohort and similar patients from NETT (Fig 2). Response rates were higher in our cohort when defined in terms of FEV1 and 6-MWD. This is consistent with the associated variables described above. Specifically, at baseline the 6-MWD was lower at baseline in our cohort than in similar NETT patients, and therefore a larger increase in 6-MWD might be expected. Similarly, FEV1 response rates may have been higher in our cohort because of the higher percentage of emphysema on quantitative CT scan in our patients compared with NETT (33.9% vs 6.9% for our cohort vs 15.9% vs 10.9% in the NETT surgical arm [6]).

Survival was similar between the high and low exercise groups in our cohort. To further explore the heterogeneity of response, a post hoc analysis of the low exercise group was performed to investigate whether there were differences between the patients in this group who died compared with those who lived for 3 years after LVRS. In this group, only 1 patient who had a baseline maximal exercise capacity of greater than 20 watts died; conversely, 5 of the 6 patients with a maximal exercise capacity of less than 20 watts died within 3 years of follow-up, giving an odds ratio of 13.3 in this subgroup. Although this observation was based on a small number of subjects, all 5 patients who died in the below 20 watts group were female. Therefore, although having upper lobe–predominant disease and a low exercise capacity was predictive of the greatest relative benefit of LVRS over medical therapy in NETT, it appears that those with the lowest baseline exercise may paradoxically be at a higher risk for death. There is a suggestion from NETT that a similar trend may have occurred [1]. In the figure delineating the exercise capacity cutoffs based on the relative benefit of the surgical procedure compared with medical therapy (Fig 5), there is a slight U-shape relationship with an upward deflection among the patients with the lowest maximal exercise capacity. Interestingly, in our cohort there was no difference between the patients who died and those who survived in terms of pulmonary function, exercise capacity, or echocardiogram parameters. Therefore, it can be speculated that some unidentified factor associated with the low exercise performance, such as more systemic inflammation or unrecognized comorbidities, may be driving their higher risk for death.

![Fig 3. Kaplan-Meier survival curves for the high versus low exercise group in our cohort. (LVRS = lung volume reduction surgery.)](image)

![Fig 4. Baseline maximal watts on postrehabilitation cardiopulmonary exercise testing (CPET) in the low exercise patients who died (“Low-died”) compared with the low exercise patients who survived for more than 3 years (“Low-alive”).](image)

![Fig 5. Figure from the supplement to the National Emphysema Treatment Trial article demonstrating the sensitivity analysis of sex-specific cutoff points for subgroups defined according to maximal workload. For the overall trend, there is a greater relative improvement in outcome with operation in the group with a maximum workload ≤40th percentile. However, there is an upward deflection in those with a maximum workload ≤10th percentile. (Reprinted from Fishman et al, N Engl J Med 2003;348:2059–73 [1] with permission from the New England Journal of Medicine.)](image)
In addition to its relevance to LVRS, the results from the present study may also be informative for bronchoscopic lung volume reduction (BLVR). Studies of BLVR have also been limited by heterogeneity of response and difficulties with proper patient selection. Factors such as preprocedural 6-MWD, hyperinflation, and degree of emphysema on CT scan (found to be associated with differential outcomes in the present study) could be further evaluated in the context of BLVR. Importantly, if our results are replicated and tested in BLVR, patients otherwise fulfilling the LVRS criteria but having a very low exercise capacity may be preferentially considered for BLVR.

This study does have limitations. It is retrospective, and the comparison group (NETT upper lobe surgical patients) was historical. Also, as expected with a complicated intervention performed outside of a clinical trial, follow-up was not complete in all patients, which may have led to survivorship bias. Therefore, outcomes at 24 and 36 months may not be as accurate as those measured at 6 or 12 months. The definition of response rates could be viewed as arbitrary, but they were based on established literature and were defined a priori. In particular, the definition of the minimal clinically important difference for 6-MWD has been controversial, with some authors suggesting an MCID as low as 25 meters in chronic obstructive pulmonary disease [22]. We have chosen a more conservative minimal clinically important difference for 6-MWD, which would lead to a lower rate of response to LVRS in our study than may be truly present. A strength of our study is the responder analysis; this avoided relying on mean data, which can lead to interpretation problems when one is dealing with a heterogeneous response. Additionally, there was excellent baseline characterization of the patient population.

Conclusions

In summary, there was a durable improvement in lung function and exercise capacity in patients selected according to the NETT criteria to have high benefit from LVRS. Despite the heterogeneity in response, survival was comparable with that of similar patients from NETT, and response rates were higher for FEV1 and 6-MWD. Those with a lower 6-MWD, more emphysema, and a higher degree of hyperinflation at baseline were most likely to respond to LVRS. Last, there appears to be a higher-risk-for-death group in those with the lowest postpulmonary rehabilitation exercise capacity (maximal watts <20), which needs further characterization and validation.

References

20. Grady D. Results of costly emphysema operation are mixed, study finds. New York Times May 21, 200321.