Maximal Inspiratory Pressure*  
Learning Effect and Test-Retest Reliability in Patients With Chronic Obstructive Pulmonary Disease

Janet L. Larson, Ph.D., R.N.; Margaret K. Covey, M.S., R.N.; Candice A. Vitalo, B.S., R.N.; Charles G. Alex, M.D., F.C.C.P.; Minu Patel, M.S.; and Mi Ja Kim, Ph.D., R.N.†

Reliability of the maximal inspiratory pressure (Pimax) was examined by measuring Pimax once a week for 4 weeks in 91 patients with chronic obstructive pulmonary disease using an aneroid pressure gauge. Five Pimax trials were conducted at each test. From the first to the fourth test, the Pimax increased by a mean of 9 cm H2O (SD = 10). From the third to the fourth test, Pimax increased by a mean of 2 cm H2O and performance appeared to be plateauing. The test-retest reliability coefficient was r = 0.97 for Pimax measured at the third and fourth test session. The 95 percent confidence interval for the absolute difference in Pimax at the third and fourth test was 3 to 5 cm H2O. We conclude that performance of Pimax improves with practice in naive COPD patients and Pimax is reliable when measured with an aneroid gauge by experienced data collectors if patients are given sufficient practice.

(Chest 1993; 104:448-53)

Pimax = maximal inspiratory pressure; RV = residual volume

METHOD

Data for this study were acquired from COPD patients during a 4-week period prior to participating in a pulmonary rehabilitation program. The study was approved by the appropriate institutional committees on human research and informed consent was obtained from all patients. Patients qualified for the study if they (1) had clinically stable COPD with moderate to severe airflow obstruction (FEV1 < 65 percent predicted), (2) had no clinical evidence of asthma and <30 percent improvement in FEV1, with bronchodilators, and (3) had no exacerbations for at least 2 months prior to the study and remained free of respiratory tract infections during the 4 weeks of data collection. Patients were dropped from the study if they reported a respiratory tract infection during the 4-week period of data collection. Potential subjects were excluded if they had major health problems that might interfere with testing. All patients were naive in regard to respiratory muscle testing and had never performed tests of Pimax. One hundred twenty-nine patients enrolled in the pulmonary rehabilitation program during data collection and 91 of them provided usable data with Pimax measured at four separate sessions. All data reported herein were taken before initiating pulmonary rehabilitation.

Maximal inspiratory pressures were measured with an aneroid pressure gauge (No. 2000-2000 cm mannegelic pressure gauge, Dwyer Instruments, Michigan City, Ind) that measured pressures from 0 to -200 cm H2O in increments of 5 cm H2O. Its calibration was verified at regular intervals against a column of water and found to be accurate within 3 cm of water throughout its full range. Sixty centimeters of pressure tubing connected the aneroid pressure gauge to a flange-type mouthpiece. An air leak was established with a small hole (diameter = 1.16 mm) in the adapter between the mouthpiece and rubber tubing.

The Pimax was defined as the largest negative pressure generated at the mouth and maintained for at least 1 full second. A minimum of five technically satisfactory trials were conducted for each test and data were discarded if there was an air leak around the mouthpiece or if the pressure was held for less than 1 full second as estimated by the data collector. The initial length of the inspiratory muscles was controlled by initiating each effort from residual volume (RV). This procedure was adopted because in the clinical situation RV is more reproducible than functional residual capacity (FRC). Patients were instructed to take their time and slowly empty their...
Table 1—Sample Characteristics (N=91)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>65 (7)</td>
</tr>
<tr>
<td>FEV₁, % predicted</td>
<td>40 (13)</td>
</tr>
<tr>
<td>FVC, % predicted</td>
<td>68 (16)</td>
</tr>
<tr>
<td>FEV₁/FVC</td>
<td>41 (9)</td>
</tr>
<tr>
<td>MVV* % predicted</td>
<td>40 (22)</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>72.45 (16.2)</td>
</tr>
<tr>
<td>Height, cm</td>
<td>170 (10)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>25 (4)</td>
</tr>
<tr>
<td>Hollingshead Index</td>
<td>40 (12)</td>
</tr>
</tbody>
</table>

*MVV = maximal voluntary ventilation.

Lungs to RV, thereby avoiding problems associated with variability in lung volumes caused by dynamic hyperinflation. Patients were not allowed to view the gauge during testing.

Data were collected by two researchers who were experienced in conducting these tests. Tests were conducted in a quiet room with no distractions and the same instructions were given to patients at the beginning of each testing session. Patients were coached in a consistent manner during each maximal inspiratory effort and brief rest periods of slightly less than 1 min were taken between repeated Pimax trials. Patients were instructed to take their time and indicate when they were ready to perform each maneuver. Test sessions were conducted at the same time of day and by the same researcher for each patient. Patients were instructed to take their usual medications as scheduled on each day of testing to control for any potential drug effects on respiratory muscle function.

Pulmonary function tests were conducted prior to the first measurement of Pimax using a spirometer that met American Thoracic Society (ATS) standards. Data were reported from the best of three forced vital capacity (FVC) maneuvers. Body mass index was calculated as the ratio of weight (kilogram) to height squared (meter squared).

In a preliminary study, interrater reliability was established for the two data collectors and reliability of the method was established by comparing measurements taken with the aneroid gauge to measurements recorded with a transducer and strip chart recorder. For this purpose, Pimax was measured in a separate group of 15 subjects. Measurements of Pimax were taken simultaneously by both data collectors from the aneroid gauge and at the same time pressures were recorded on a strip chart recorder from a pressure transducer (Viggo Spectromed, Oxnard, Calif). Throughout all reliability tests, both data collectors were kept blind to the results. Data were compared for differences between the two data collectors and for differences between the two methods, aneroid pressure gauge, and transducer with strip chart recorder. Mean net difference was calculated to reflect systematic measurement error such as one data collector or one technique producing higher or lower values than the other. The mean of absolute differences ([Pimax1 – Pimax2]) was calculated to reflect magnitude of fluctuation in Pimax regardless of the direction. Absolute differences reflect the magnitude of combined systematic and nonsystematic fluctuations in the data. With mean net difference nonsystematic positive and negative differences tend to cancel each other out. When the results from the two data collectors were compared, the mean net difference in Pimax was zero (SD = 4) and mean absolute difference was 3 cm H₂O (SD = 2). The interrater reliability coefficient was r = 0.99 (df = 13). When data from the aneroid gauge were compared with data recorded with the transducer, the mean net difference in Pimax was 0 cm H₂O (SD = 7) and 0 cm H₂O (SD = 8) for each of the data collectors, respectively. The mean absolute difference in Pimax for the two methods was 5 cm H₂O (SD = 5) and 6 cm H₂O (SD = 5) cm H₂O for each of the data collectors. Reliability coefficients for the two methods were r = 0.97 (df = 13) and r = 0.96 (df = 13) for each of the data collectors.

Data Analysis

Descriptive statistics were calculated for differences and absolute differences. Analysis of variance for repeated measures was calculated to identify significant differences between test sessions and to determine if there was a linear effect or curvilinear effect over time. Net differences and absolute differences were calculated to describe systematic and nonsystematic fluctuations in repeated measures of Pimax. Test-retest reliability was calculated with Pearson's correlations.

RESULTS

The sample was 69 men and 22 women with moderate to severe COPD (Table 1). Most patients were taking standard pharmacologic therapies, including inhaled β-agonists, oral β-agonists, and methylxanthines. Nine were using home oxygen. Ten were currently smoking, one had never smoked, and the remainder had quit smoking. The sample was middle class as indicated by the Hollingshead Index.

Performance gradually improved as demonstrated by the mean (SD) Pimax for each test session: first −53 (24) cm H₂O, second −57 (25) cm H₂O, third −60 (25) cm H₂O, and fourth −62 (25) cm H₂O. Analysis of variance for repeated measures identified a significant increase in Pimax with a significant linear effect and no significant curvilinear effect over time (p < 0.001). Statistically, performance did not plateau;
however, between the third and fourth tests, the net increase in Pimax was so small that it was not clinically significant and performance appeared to be plateauing (Table 2). The frequency distribution for net differences in Pimax is presented in Figure 1. A weak negative correlation between Pimax and age was strengthened from the first (r = -0.19; p>0.05) to the fourth test session (r = -0.30; p<0.01).

Sample characteristics were examined to determine if selected characteristics influenced the magnitude of improvement observed with four test sessions. Age correlated negatively (r = -0.33; df=89) with net differences in the first and fourth measures of Pimax (Pimax4 minus Pimax1). However, the correlation with age decreased when net difference scores were calculated either as a percent of the initial Pimax (r = -0.19; df = 89) or as a percent of the fourth Pimax (r = -0.16; df = 89). No significant correlations were observed between improvement in Pimax and spirometry or body mass index. Mean net differences and mean absolute differences among repeated measures of Pimax were similar for men and women.

This finding was further demonstrated by grouping patients based on the magnitude of improvement from the first to fourth measure of Pimax. From the first to the fourth measure, 61 patients increased Pimax by ≤10 cm H2O and 30 patients increased by >10 cm H2O. At the first test session, Pimax was similar for both groups with means of -53 (SD = 25) cm H2O for the group that improved by ≤10 cm H2O and -54 (SD = 22) cm H2O for the group that improved by >10 cm H2O. At the fourth test session, mean Pimax was -56 (SD = 25) and -74 (SD = 22) cm H2O, respectively, for the same groups. Sample characteristics for these groups were compared using independent t tests. The groups were significantly different with respect to age (p<0.01) and not significantly different for any other variables (Table 3).

The test-retest reliability was examined by comparing the third and fourth test of Pimax. Third and fourth tests of Pimax were highly reliable as reflected by a mean absolute difference of less than 5 cm H2O (Table 2). A cumulative frequency distribution of absolute differences in Pimax illustrated that 79 percent of subjects were within 5 cm H2O and 93 percent.

Table 3—Comparison of Patients Grouped According to the Magnitude of the Learning Effect From the First to the Fourth Measure of Pimax*

<table>
<thead>
<tr>
<th></th>
<th>≤10 cm H2O</th>
<th>&gt;10 cm H2O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 61</td>
<td>n = 30</td>
</tr>
<tr>
<td>Age, yr</td>
<td>66 (7)</td>
<td>61 (6)†</td>
</tr>
<tr>
<td>FEV1, % predicted</td>
<td>39 (13)</td>
<td>41 (14)</td>
</tr>
<tr>
<td>FVC, predicted</td>
<td>67 (17)</td>
<td>70 (15)</td>
</tr>
<tr>
<td>FEV1/FVC</td>
<td>41 (9)</td>
<td>42 (10)</td>
</tr>
<tr>
<td>MVV, % predicted</td>
<td>39 (20)</td>
<td>43 (25)</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>71.55 (16.65)</td>
<td>74.25 (14.85)</td>
</tr>
<tr>
<td>Height, cm</td>
<td>170 (10)</td>
<td>170 (7.5)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>24 (4)</td>
<td>25 (4)</td>
</tr>
<tr>
<td>Pimax, first measure, cm H2O - 53</td>
<td>(25)</td>
<td>-54 (22)</td>
</tr>
</tbody>
</table>

*In one group maximal inspiratory pressure (Pimax) increased by ≤10 cm H2O and in the other group Pimax increased by >10 cm H2O. MVV = maximal voluntary ventilation.
†p<0.01.
of subjects were within 10 cm H$_2$O on third and fourth tests (Fig 2). Test-retest reliability coefficient was $r = 0.97$ ($df = 89$) for the third and fourth test of Pimax.

Data were further examined to determine if five trials were sufficient to attain peak performance at each test session. Thirty-four percent of the sample generated their best Pimax at the first trial. Ten percent generated their best Pimax at the fifth trial.

**Discussion**

The mean Pimax was consistent with values reported for similar groups of patients with moderate to severe COPD. The results of this study supported the notion that in naive COPD patients, performance of Pimax improves with practice of the Pimax maneuver. The third and fourth test of Pimax produced reliable data.

**Learning Effect**

Systematic differences between consecutive tests were relatively small but at the fourth test, Pimax was 17 percent higher than at the first test. Other investigators observed similar increases in Pimax during an initial learning period, but for most the observed improvements in Pimax were not statistically significant because they used smaller samples with inadequate statistical power to detect an effect of this magnitude. The study of Ringquist is the one exception. He studied a larger sample and found no statistically significant improvement with the second measure of Pimax. This apparent inconsistency with our results may be attributed to the number of trials employed with each test session and the level of motivation. Ringquist studied healthy military conscripts who were highly motivated and athletic as compared with our COPD patients. Moreover, he conducted approximately ten trials per session, whereas we conducted five trials. It is possible that the high level of motivation and the additional number of trials enabled subjects to learn the Pimax maneuver in a single session.

In other tests of muscle strength, the evidence for a learning effect is inconsistent. A learning effect was observed when testing a variety of muscle groups, including the following: strength of the wrist extensors and elbow extensors as measured by a hand-held dynamometer, strength of the wrist flexors as measured by a tensiometer, and hand grip strength as measured by a hand dynamometer. In contrast, no learning effect was observed in other studies when testing strength of the elbow flexors as measured by a hand-held dynamometer, strength of right elbow flexors and right shoulder abductors, and strength of the right wrist flexors. In these studies and studies of Pimax, inconsistencies may be accounted for by a number of factors, including the individual's familiarity with the required physical task and their prior athletic experience.

In becoming familiar with a strenuous physical task, people learn to ignore extraneous sensations and to perform the task with very little conscious thought. The physical task required to perform a Pimax maneuver from RV is unique and unlike any other physical activities performed in daily life. This could increase the amount of practice required to learn the task. We suspect that in becoming familiar with the Pimax maneuver, patients learned to relax and ignore the sensations of pressure in the head while focusing on sensations of muscle contraction in the chest wall. We observed that many patients appeared to become more relaxed and self-confident after the first test session.

Additionally, past athletic experience may influence patients' mental attitude and their ability to concentrate on the task. Anecdotally, we observed that former athletes approached the task with greater intensity of purpose and with less fear of overexertion. This may account for some of the observed variability in performance.

Many patients demonstrated a learning effect of $>10$ cm H$_2$O. Clinically, it would be useful to predict which patients were most likely to demonstrate a large learning effect, but this was not possible given the data we collected. In our study, older patients were less likely to demonstrate a large learning effect by increasing their Pimax more than 10 cm H$_2$O. However, the relationship with age disappeared when the increases in Pimax were calculated as a percentage of actual Pimax, because older patients tended to generate lower values for Pimax. We suspect that the magnitude of the learning effect was influenced by other factors as described above.

Our data supported the need for at least five trials for each measure of Pimax. While most subjects...
generated their best Pimax within the first three trials, as many as 10 percent generated their best Pimax at the fifth trial of each test session. The question of conducting more than five trials with each test session was not addressed by this study. However, these inexperienced subjects required a total of 15 to 20 trials over three test sessions to establish peak performance, suggesting that more than five trials may be required per test session. This would be consistent with the findings of Fiz and coworkers\(^\text{22}\) in which a total of nine trials were needed to establish peak performance in a single test session with inexperienced COPD patients.

**Test-Retest Reliability**

Performance did not technically plateau, but systematic differences between Pimax measured at the third and fourth test sessions were very small; hence, data from these sessions were examined for test-retest reliability. The mean absolute difference, cumulative frequency distribution for the absolute difference, and the 95 percent confidence interval for the absolute difference indicated that Pimax fluctuated by approximately 5 cm H\(_2\)O or less in most of the sample. This was consistent with other measures of strength where day-to-day fluctuations were reported to be ±10 percent.\(^\text{21}\) However, a portion of the sample demonstrated wider fluctuations. Taking a conservative approach, we conclude that when monitoring individual patients at repeated intervals, Pimax must change by approximately 10 cm H\(_2\)O or more before it can be interpreted as a clinical change in strength. Moreover, a repeatable trend must be established to confirm a change in strength.

A pressure transducer and strip chart recorder are frequently used in the research laboratory to reduce methodologic measurement error, but the aneroid gauge is more practical in field testing and in the clinical setting. Our data from 15 subjects suggested that experienced data collectors could acquire reliable data with an aneroid gauge. However, there are two notable limits to this technique. Maximal inspiratory pressure data must be interpreted with caution if patients have difficulty sustaining maximal pressures for 1 full second. Moreover, the aneroid gauge cannot be used reliably when patients generate wildly fluctuating pressures during each effort as this makes it difficult to identify the precise point at which the pressure was maintained for 1 full second.

In this study, the test-retest reliability for Pimax was high in part because we controlled for factors that might influence biologic variability in performance. Repeated measurements were taken at the same time of day to control for the effects of fatigue. We discarded data from all patients who experienced an upper respiratory tract infection because of potential effects on respiratory muscle strength as demonstrated by Mier-Jedrzejowicz et al.\(^\text{23}\) Both data collectors were experienced in conducting tests of Pimax. This was important because experienced data collectors provided clear unambiguous instructions and placed patients at ease allowing them to concentrate on generating a maximal inspiratory effort. Tests were conducted in a quiet environment to improve concentration and the individual Pimax trials were paced in a slow manner to reduce stress. It was not possible to control for subject motivation, but we tried to minimize the effects of different levels of motivation by using the same precise directions for each test.

We conclude it is necessary to control for the learning effect when measuring Pimax in naive COPD patients. To ignore this important source of systematic measurement error increases the chance of reporting an improved strength that does not exist. However, numerous practice sessions are time-consuming and more efficient methods are needed to speed up the learning process. For the group as a whole, test-retest reliability was high and day-to-day fluctuations in performance were consistent with other tests of strength.

**ACKNOWLEDGMENTS:** We gratefully acknowledge the Bioinstrumentation and Biostatistics Facilities of the Research Resources Center, University of Illinois at Chicago, which provided support services necessary to conduct this study.

**REFERENCES**


12. Hollingshead AB. Four factor index of social status. New Haven, Conn: Yale University, 1975

13. Byrd RB, Hyatt RE. Maximal respiratory pressures in chronic

**MIP in Chronic Obstructive Pulmonary Disease (Larson et al)**


22 Fiz JA, Montserrat JM, Picado C, Plaza V, Agusti-Vidal A. How many manoeuvres should be done to measure maximal inspiratory mouth pressure in patients with chronic airflow obstruction? Thorax 1989; 44:419-21


European Society of Computing and Technology in Anesthesia and Intensive Care

The Fourth Annual Meeting of ESCTAIC will be held in Goldeg Castle, Salzburg, Austria, October 6-9. For information, contact Mr. Peter Schwab, c/o University Klinik fur Anasthesie und Allgemeine Intensivmedizin, Anichstrasse 35, A-6020, Innsbruck, Austria.