New method for evaluating maximal respiratory pressures: Concurrent validity, test-retest, and inter-rater reliability


Objective: To evaluate concurrent validity, test-retest, and inter-rater reliability of MRP performed with a new instrument in healthy individuals, and to compare values obtained at different volumes in healthy individuals and individuals with COPD.

Methods: MRP of 100 healthy individuals were obtained using the TrueForce and the MicroRPM at residual volume (RV) and total lung capacity (TLC) to evaluate concurrent validity. MRP were obtained at functional residual capacity (FRC) using the TrueForce to evaluate reliability. Comparisons of inspiratory pressure values (FRC compared to RV) and expiratory pressure values (FRC compared to TLC) were performed with 100 healthy individuals and 15 individuals with COPD.

Results: The intraclass correlation coefficient (ICC) was 0.77 and 0.86 for concurrent validity for inspiratory and expiratory pressures, respectively. Test-retest reliability showed an ICC of 0.87 for inspiratory pressure, and 0.78 for expiratory pressure; inter-rater reliability showed an ICC of 0.91 for inspiratory pressure, and 0.84 for expiratory pressure. Measurements performed at RV and TLC were higher when compared to FRC [mean difference (95%CI)= -8.30 (-11.82, -4.78) cmH₂O; -37.29 (-42.63, -31.96) cmH₂O] in healthy individuals, and -11.09 (-15.83, -6.35) cmH₂O; -57.14 (-71.05, -43.05) cmH₂O in COPD, for inspiratory and expiratory pressures, respectively.

Conclusion: MRP performed with the TrueForce presented good concurrent validity, good test-retest reliability, excellent inter-rater reliability for inspiratory pressure and good inter-rater reliability.
Introduction

Maximal respiratory pressure (MRP) measurements are commonly used for evaluating respiratory muscle strength. For these measurements, maximal inspiratory (PImax) and expiratory (PEmax) efforts are performed at the mouth against an occluded piece connected to a manometer. This is a simple, non-invasive, and well-tolerated method, which can be useful for assessing and monitoring patients with respiratory muscle weakness due to respiratory, cardiac, or neuromuscular conditions.

However, the values obtained with these measurements are influenced by different factors, including the lung volume at which the test is performed. PImax is usually performed at residual volume (RV) and PEmax at total lung capacity (TLC), which reflects the pressure of the respiratory muscles in addition to the elastic recoil pressure of the respiratory system, overestimating the actual muscle respiratory strength. The American Thoracic Society/European Respiratory Society statement (ATS/ERS) reports an increase of 30 cmH2O if PImax is performed at RV and an increase of 40 cmH2O if PEmax is performed at TLC in healthy individuals. Studies have shown that higher PImax values were obtained if the test was performed at RV in healthy individuals and in individuals with chronic obstructive pulmonary disease (COPD), and higher PEmax values were observed if tests were performed at TLC also in healthy individuals.

The elastic recoil pressure of the respiratory system at functional residual capacity (FRC) under physiological conditions is close to zero. MRP measurements performed at FRC would reflect respiratory muscle pressure, minimizing the influence of elastic recoil. Studies have used this method to evaluate MRP but they identified FRC through a qualitative method (visual inspection) using spirometric data, or through a very complex and not clinically applicable method (body plethysmography). A new digital manometer named TrueForce was recently developed. It monitors volume and flow in real time, enabling identification of FRC and MRP measurements at this lung volume, which was not possible in the current available manometers. In addition, this instrument is simple to operate and it communicates with a dedicated software, which enables graphically visualizing the curves: pressure versus time and volume versus time in real time, and it enables evaluating variables such as maximum average pressure, peak pressure, and plateau pressure.

Before a measurement instrument becomes available for research and clinical use, its measurement properties must be established. The primary aim of this study was to evaluate concurrent validity of the TrueForce for MRP performed at RV and TLC, and test-retest and inter-rater reliability of MRP performed at FRC with the TrueForce in healthy individuals. The secondary aim was to compare measurements performed at different lung volumes (FRC and RV for PImax, and FRC and TLC for PEmax) with the new instrument within two groups: healthy individuals and individuals with COPD.

Methods

Study design and sample

This was a methodological study. The inclusion criteria for healthy individuals were: age between 20 and 40 years old, body mass index between 18.5 and 29.99 kg/m², normal lung function according to predicted values, self-reported absence of cardiac and neuromuscular diseases and absence of contraindications for the performance of MRP tests. The inclusion criteria for individuals with COPD were: diagnosis of COPD confirmed by lung function test, age between 40 and 85 years, and clinically stable (no exacerbations and/or hospitalization in the past four weeks). The exclusion criteria for healthy individuals were: inability to understand or perform any of the procedures, fever and/or cold in the two weeks prior to the tests, or exhaustive exercise in the last 48 hours before tests; the individual would also be excluded for precautionary reasons if peripheral oxygen saturation before or during the tests was less than 90%, heart rate between 60 and 100 bpm, and/or blood pressure before the test was greater than or equal to 160/100mmHg. The exclusion criteria for individuals with COPD were: other pulmonary diseases or inability to understand and/or perform the procedures. The study was approved by the Institution Ethics Committee from the Universidade Federal de Minas Gerais (CAAE: 80257617.0.0000.5149), Belo Horizonte, MG, Brazil, and written informed consent was obtained from all participants.

Main measurement instrument

MRP at FRC were measured using the new TrueForce manometer (Fig. 1). This instrument obtains information through two sensors, pressure and flow. The pressure sensor is designed to measure differential pressure into a range of ±5psi, with a frequency response up to 2 kHz and...
resolution = 0.04 cmH₂O. The flow sensor presents a flow range of ±200 slm, typical accuracy of 1.5%, sample rate up to 2 kHz, and enables volume to be calculated in real time, which permits maneuvers at FRC. The interface used was a flanged silicone mouthpiece with a 2 mm air leak orifice present in the instrument. In addition, the TrueForce is portable and its communication with the software is via Bluetooth. The PImax and PEmax variables, which are calculated by the area of one second around the peak pressure value, were analyzed through the graphic interface of the Manovac-FRC software program.

Complementary measures

A Koko® PFT spirometer (nSpireHealth Inc., USA) was used to assess lung function, and the test was performed according to the ATS/ERS recommendations. Obtained values were compared with those predicted by Pereira et al.12

Body weight and height were assessed with a calibrated scale with stadiometer (Filizola ind. Ltda, Brazil). A MicroRPM® manometer (Micro Medical, UK) was used to evaluate the concurrent validity of the new instrument. This reliable instrument is well-established in the literature for measuring PImax at RV and PEmax at TLC.20,21 The PUMA PC software program (Micro Medical, UK) was used to operationalize MRP.

The GOLD classification of airflow limitation severity was used to characterize the individuals with COPD as follows: mild (forced expiratory volume in first second · FEV₁ ≥ 80% predicted), moderate (50% ≤ FEV₁ < 80% predicted), severe (30% ≤ FEV₁ < 50% predicted), and very severe (FEV₁ < 30% predicted).13

Procedures

For healthy individuals, data were collected on two days within a 7 to 15-day interval by trained evaluators (both with a background in physical therapy and trained by a senior investigator).22 Participants were screened for previous physical activity, health condition, and vital signs in all testing sessions. The sessions occurred at the same time of the day for each participant, and instructions about the protocol were provided in a standardized manner performed by trained evaluators, minimizing variations in the environment and measurement setting. The flowchart of measurements is presented in Fig. 2.

The participants underwent a lung function test on the first day after an initial interview for demographic and clinical data. Next, the participants underwent MRP at FRC, RV, and TLC by Evaluator 1. The participants remained in a sitting position with their trunk and lower limbs supported, head in a neutral position and used a nose clip for the PImax and PEmax measurements performed at FRC with the TrueForce. They were instructed to breathe several times following a pattern similar to a sigh: inhaling just above the tidal volume and exhaling as if relaxing the chest. After performing the required number of respiratory cycles, a yellow light indicated that a breathing pattern was found, and the FRC range was identified. In the next respiratory cycle, a green light alerted the evaluator to occlude the occlusion orifice and request a maximal inspiration or expiration according to the test (PImax or PEmax). The evaluator also needed to press the cheeks of the participants in the case of PEmax. Participants performed at least five maneuvers with a 1-minute interval between them. At least three of these five efforts should be acceptable (without air leakage, pressure maintained for at least 1.5 seconds, and volume of the maneuver within limits of FRC), and reproducible (variation < 10%). The last maneuver could not be the greatest, which would suggest a learning effect. The highest value was selected. PImax at RV and PEmax at TLC were then performed after a 10-minute interval and vital signs returned to baseline, according to the ATS/ERS recommendations.

PImax and PEmax at FRC were repeated by Evaluator 1 on the second day, and then by Evaluator 2 (blinded) after a 10-minute interval and vital signs having returned to baseline with the aim to evaluate test-retest and inter-rater reliability, respectively. The tests using the digital manometer MicroRPM® at RV and TLC were subsequently performed after resting 10 more minutes and vital signs having returned to baseline again to evaluate the concurrent validity, also by Evaluator 1 and following standardized recommendations. The test order was randomized: PImax at RV and PEmax at TLC using TrueForce and MicroRPM® were randomized within days; and the tests performed on the same day were all randomized, including the PImax and PEmax order.

For individuals with COPD, the measurements of MRP at FRC and at RV and TLC were obtained on the same day in a

Fig. 2 Flowchart of the study protocol for healthy individuals. Abbreviations: PImax, maximal inspiratory pressure; PEmax, maximal expiratory pressure; RV, residual volume; TLC, total lung capacity; FRC: functional residual capacity.
randomized order by one evaluator following the same protocol described for healthy individuals.

**Sample size**

The sample size calculation for the measurement properties assessment followed the methodology proposed by Mokkink et al. They recommend a sample of at least 100 individuals for a very good reliability assessment and a sample of at least 50 individuals for a very good validity assessment of an instrument. The same sample of healthy individuals were evaluated for the comparison between MRP obtained at different lung volumes, in addition to 15 individuals with COPD. The sample size for this comparison was based on all patients referred to our outpatient rehabilitation program, who met inclusion criteria, until its interruption on the beginning of the COVID-19 pandemic.

**Statistical analysis**

The data distribution was verified by the Shapiro-Wilk test. Data were presented as mean and standard deviation. Concurrent validity, test-retest, and inter-rater reliability were evaluated by intraclass correlation coefficient (ICC) two-way mixed model (ICC_{2,1}), with a 95% confidence interval (CI). ICC values less than 0.5 were considered poor, between 0.5 and 0.75 moderate, between 0.75 and 0.9 good, and greater than 0.9 excellent reliability. Agreement between measurements was also evaluated by graphical analyses using Bland-Altman plots. Differences between MRP performed at FRC and at RV and TLC were compared using the paired t-test analyses within groups separately for healthy individuals and individuals with COPD. The data were analyzed in the SPSS version 23.0 and the significance level was set at 5%.

**Results**

A total of 151 individuals were initially recruited. However, 33 did not meet inclusion criteria, while another 18 were excluded for not attending the second visit. Therefore, 100 healthy participants completed the protocol for the reliability evaluation and the first 50 participants were assessed for concurrent validity. Fifteen individuals with mild to severe COPD were evaluated.

Table 1 shows the demographic, anthropometric, and clinical data of the participants.

**Concurrent validity, test-retest, and inter-rater reliability**

Table 2 presents data regarding the concurrent validity, test-retest, and inter-rater reliability. PImax at RV and PEmax at TLC performed with the new instrument (TrueForce) and with the instrument considered gold standard (MicroRPM®) presented ICC values of 0.77 and 0.86, respectively. The test-retest reliability of PImax and PEmax obtained at FRC presented ICC values of 0.87 and 0.78, respectively. Inter-rater reliability showed an ICC value of 0.91 for PImax and 0.84 for PEmax.

**Discussion**

This study showed that PImax at RV and PEmax at TLC performed with TrueForce and MicroRPM® have similar values. MRP performed with TrueForce presented good test-retest reliability for PImax and for PEmax. The inter-rater reliability was excellent for inspiratory pressure and good for expiratory pressure. PImax and PEmax performed at RV and TLC, respectively, presented significantly higher values compared to measures performed at FRC for healthy individuals and for those with COPD.

The use of portable and digital mouth pressure meters has been studied since 1994, when Hamnegaard et al. presented data regarding the concurrent validity, test-retest, and inter-rater reliability. PImax at RV and PEmax at TLC performed with the new instrument (TrueForce) and with the instrument considered gold standard (MicroRPM®) presented ICC values of 0.77 and 0.86, respectively. The test-retest reliability of PImax and PEmax obtained at FRC presented ICC values of 0.87 and 0.78, respectively. Inter-rater reliability showed an ICC value of 0.91 for PImax and 0.84 for PEmax.
Table 2  Data on concurrent validity, test-retest, and inter-rater reliability evaluated in 100 healthy individuals.

<table>
<thead>
<tr>
<th>Measurement properties</th>
<th>TrueForce</th>
<th>MicroRPM®</th>
<th>ICC</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PImax&lt;sub&gt;RV&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>106.5 ± 28.5</td>
<td>113.7 ± 32.7</td>
<td>0.77</td>
<td>0.60, 0.87</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PEmax&lt;sub&gt;TLC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>126.3 ± 33.7</td>
<td>135.4 ± 37.8</td>
<td>0.86</td>
<td>0.73, 0.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PImax&lt;sub&gt;FRC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>96.8 ± 23.0</td>
<td>98.2 ± 23.9</td>
<td>0.87</td>
<td>0.80, 0.90</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PEmax&lt;sub&gt;FRC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>81.7 ± 23.7</td>
<td>89.0 ± 26.4</td>
<td>0.78</td>
<td>0.69, 0.84</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter-rater reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluator 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PImax&lt;sub&gt;FRC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>98.2 ± 23.9</td>
<td>100.6 ± 37.8</td>
<td>0.91</td>
<td>0.86, 0.94</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PEmax&lt;sub&gt;FRC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>89.0 ± 26.4</td>
<td>91.7 ± 30.2</td>
<td>0.84</td>
<td>0.77, 0.89</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Evaluator 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PImax&lt;sub&gt;FRC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEmax&lt;sub&gt;FRC&lt;/sub&gt; cmH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data are expressed as mean ± standard deviation. Abbreviations: PImax, maximal inspiratory pressure; PEmax, maximal expiratory pressure; RV, residual volume; TLC, total lung capacity; FRC, functional residual capacity; ICC, intraclass correlation coefficient; 95% CI, 95% confidence interval.

validated a hand-held device in healthy individuals and individuals with respiratory diseases. Pessoa et al. also evaluated the concurrent validity of a digital manometer in measuring MRP. The correlation between the values observed was high for all variables. Our results agree with these findings, showing good agreement between the instruments for MRP performed at RV and TLC.

There is a lack of recent studies evaluating test-retest and inter-rater reliability of MRP measured at FRC, which may be explained by the absence of a simple instrument available for this evaluation. Larson and Kim studied the test-retest reliability of PImax measured at different lung volumes in 31 healthy individuals, and showed that the two tests of PImax measured at FRC presented excellent correlation (r=0.90) and no significant difference between them. Jardim et al. evaluated the test-retest reliability of PImax and PEmax measured at FRC in healthy individuals using an aneroid manometer in three consecutive days, with no significant differences between them. No clear description on how they identified FRC was provided.

The results of the present study showed excellent inter-rater reliability for PImax and good reliability for PEmax in healthy individuals. Jalan et al. tested the reliability of a pressure device in 40 healthy individuals, and presented an ICC of 0.92 for PImax performed by two evaluators. However, the PImax was obtained at RV. To the best of our knowledge, there are no studies evaluating inter-rater reliability of MRP measured at FRC. Therefore, comparisons were limited.

Our results also showed that PImax obtained at RV and PEmax obtained at TLC presented higher values compared to measures performed at FRC. These findings are probably explained by the additional passive elastic recoil pressure of the respiratory system. Ringqvist evaluated MRP of healthy individuals at different vital capacity percentages, reporting significant increases when performing PImax at RV and PEmax at TLC compared to FRC (11.4 cmH<sub>2</sub>O and 54.5 cmH<sub>2</sub>O, respectively). The ATS/ERS statement in 2002 states that PImax obtained at RV may contribute to an increase of 30 cmH<sub>2</sub>O in the final result and an increase of 40 cmH<sub>2</sub>O if PEmax is obtained at TLC, in healthy individuals. Windisch et al. evaluated 533 healthy individuals and observed a mean increase of ~10 cmH<sub>2</sub>O when PImax was performed at RV compared to FRC. Our results are consistent with their findings regarding the significant difference between MRP at different lung volumes, but there are variations among the mean increases, mostly for PEmax. This may be explained by variations in the clinical, demographic, and anthropometric factors of individuals included in the studies, or by different methods used for evaluating MRP and different lung volumes.

In this study, the mean increase in PEmax obtained at TLC compared to FRC in individuals with COPD was noteworthy. This may also be explained by the changes in the length of the expiratory muscles with the volume reduction (from TLC to FRC), minimizing their pressure-generating capacity. We highlight that individuals with COPD constantly activate abdominals (mainly the transversus abdominis) due to the flow limitation, which may cause an increased maximal expiratory pressure in greater lung volumes – in which the abdominals present a more advantageous position compared to measurements at FRC, potentially explaining the greater difference between the measurements. For PImax, the difference between measurements obtained at RV and FRC was similar to the values described for healthy young individuals, however, individuals matched on age and sex should be evaluated to determine the influence of lung volume on these measurements for...
individuals with COPD. Langer et al. measured PImax at FRC in individuals with COPD and reported an increase of ~16 cmH₂O when the test was obtained at RV. We hypothesized that lower PImax values identified in individuals with COPD may be associated with real inspiratory muscle weakness but also with lung hyperinflation, which may lead the diaphragm to a less mechanical advantageous position. It is well known that maximal respiratory mouth pressure measurements are the most used tests for evaluating respiratory muscle strength. However, PImax and PEmax are frequently performed at RV and TLC, respectively, which may lead to inaccurate conclusions. The use of the method proposed by this study, which permits more accurate MRP measurements at FRC, enables evaluating muscle strength only, 

![Fig. 3 Bland-Altman plots to visualize the agreement of: (A) Measures performed with TrueForce and MicroRPM - concurrent validity; (B) Measures performed by Evaluator 1 - test-retest reliability; (C) Measures performed by Evaluator 1 and 2 - inter-rater reliability. Data related to the evaluation of 100 healthy individuals. Abbreviations: PImax, maximal inspiratory pressure; PEmax, maximal expiratory pressure; ULA, upper limit of agreement; LLA, lower limit of agreement.](image-url)

---

B.M. Silveira, M.C. Pereira, D.R. Cardoso et al.
minimizing the influence of viscoelastic properties of the respiratory system in the final outcomes. Therefore, the MRP results obtained through this new method may be used to discriminate respiratory muscle weakness due to actual lower respiratory muscle strength or due to impairments in elastic properties related to the different healthy conditions, thereby justifying the clinical application of the new proposed method.

One limitation of this study was the impossibility to test the criterion validity of the instrument for measuring maximal respiratory pressures at FRC due to the lack of feasibility of the gold standard tests for this evaluation. Further studies including patients with different health conditions and with different age ranges are needed with the aim to evaluate the clinical applicability of these measures at FRC.

Conclusion

The TrueForce presented good concurrent validity. Pimax and PEmax performed at functional residual capacity presented good test-retest reliability. Pimax presented excellent inter-rater reliability and PEmax showed good inter-rater reliability. This study also demonstrated that Pimax obtained at RV and PEmax obtained at TLC presented higher values compared to measures performed at functional residual capacity in healthy individuals and in individuals with COPD.

Conflicts of Interest

None to declare.

Acknowledgements

The authors thank the engineer Armando Aguiar de Souza Cruz Neto and co-workers for his contribution in the development of the software’s first version - Manovac. This work was partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – Brazil), Finance Code 001; the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant 309990/2017-3 and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). The sources of financial support had no role in the design, implementation, or reporting of this study.

References

12. Pereira CAC, Sato T, Rodrigues SC. New reference values for maximal respiratory pressures at FRC due to the lack of feasibility of the gold standard tests for this evaluation. Further studies including patients with different health conditions and with different age ranges are needed with the aim to evaluate the clinical applicability of these measures at FRC.


