Revisiting the accuracy of peak flow meters: a double-blind study using formal methods of agreement

Z. Nazira, S. Razaqa, S. Mira, M. Anwara, G. Al Mawlawia, M. Sajada, A. Shehabb, R.S. Taylorc,*

aSchool of Medicine, University of Birmingham, UK
bDivision of Clinical Investigation, Queen Elizabeth Hospital, Birmingham, UK
cDepartment of Epidemiology & Public Health, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received 31 March 2004; accepted 14 October 2004

Summary

Background: There is widespread use of peak flow meters in both hospitals and general practice. Previous studies to assess peak flow meter accuracy have shown significant differences in the values obtained from different meters. However, many of these studies did not use human subjects for peak flow measurements and did not compare meters of varying usage. In this study human subjects have been used with meters of varying usage.

Methods: Participants were tested using two new (meters A and C) and one old peak flow meter (meter B) in random order. The study was double-blinded. Participants were recruited from the university campus.

Results: Four hundred and nine individuals participated. The difference between peak flow means of A and B was -9.93 l/min (95% CI: -12.37 to -7.48, P<0.0001). The difference between peak flow means of B and C was 20.08 l/min (95% CI: 17.85–22.29, P<0.0001). The difference between peak flow means of A and C was 10.15 l/min (95% CI: 7.68–12.61, P<0.0001).

Conclusion: There was a significant difference between the values obtained from the new and old peak flow meters and also between the two new peak flow meters. We conclude that there is need for caution in interchangeably using flow meters in clinical practice.

© 2004 Elsevier Ltd. All rights reserved.

KEYWORDS

Peak flow meters; Methods of agreement; Double-blind
Introduction

Portable devices for measuring peak expiratory flow (PEF) were pioneered by Martin Wright in 1959 and PEF is now widely accepted as an independent measure of lung function. Patients can record PEF easily and reliably, without the help of technically skilled personnel. This has proved to aid both physicians and patients in the management of asthma. By helping to monitor disease progression and response to treatment, peak flow meters allow for early intervention during exacerbation of asthma thereby avoiding the need for unnecessary hospitalisation. It is therefore essential for peak flow meters to have a linear response in order to obtain an unbiased measurement of PEF variability.

Several studies have aimed to test accuracy and reliability of peak flow meters. In these studies meters were compared to devices made by other manufacturers, devices of the same brand and a standard (pneumotacograph). These studies showed significant inter and intra-meter variation with increasing usage.

Unlike the majority of previous studies, our study used human subjects rather than machines to test the meters. This is more reflective of clinical practice. Furthermore, the sample size of this study (409 subjects) exceeds that of previous studies, the largest being 212 subjects.

The aim of this study was to compare the agreement of three peak flow meters using a double-blinded design and formal statistical methods of agreement.

Methods/design

The study was conducted using a cross-sectional design. Three Mini Wright peak flow meters were used. Meters 'A' and 'C' were new and had not been previously used and Meter 'B' had been used in the hospital setting for the past year. Volunteers were asked to force expire into each meter three times, the highest value was recorded. The mean value was not recorded as one poor effort could significantly lower the mean, therefore giving a false reading. Participants’ age, sex and asthma status was recorded (by self-report and medication history).

Randomisation and blinding

A convenience sample was drawn from The University of Birmingham consisting of students and staff. The study design ensured both observers and subjects were blinded to the identity of the peak flow meter being used. A single group member provided standard instructions to participants for correct meter usage and correct technique was demonstrated where necessary. This removed any bias due to variation in subject technique. A spreadsheet package was used to generate random sequences for the order in which subjects were tested with each meter.

Analysis

The level of agreement between devices was analysed using three statistical approaches:

1. Association: Pearson’s product moment correlation coefficients were calculated to assess any association between the devices.
2. Comparison of means: an ANOVA (analysis of variance) was used to show any difference between devices and if present, a paired t-test was used to show where the significant difference(s) existed.
3. The Bland and Altman method was used to illustrate the degree of variation in agreement between devices.

All analysis was performed using STATA v.6 and Microsoft Excel 97.

Results

Four hundred and nine subjects were tested, 261 males (64%) and 148 females (36%). Demographic details of our sample are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Table of subject characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Number of non-asthmatics (%)</td>
<td>230 (56.23)</td>
</tr>
<tr>
<td>Number of asthmatics (%)</td>
<td>31 (7.58)</td>
</tr>
<tr>
<td>Total</td>
<td>261</td>
</tr>
<tr>
<td>Mean age (SD)</td>
<td>20.89 (3.93)</td>
</tr>
</tbody>
</table>
1. **Association using Pearson’s product moment correlation coefficient**: A strong positive correlation was seen in all three comparisons. A versus B, correlation coefficient ($r$) = 0.962. B versus C, $r = 0.969$ and A versus C, $r = 0.961$.

2. **Comparison of means**: The absolute means for the readings taken using meters A, B and C were 553.52, 563.45 and 543.37 l/min, respectively. The mean differences were $-9.93 \text{l/min}$, $P<0.0001$ between devices A and B, $20.08 \text{l/min}$ between devices B and C, and $10.15 \text{l/min}$ between devices A and C.

**Table 2** Comparison of devices—peak flow (l/min) reported.

<table>
<thead>
<tr>
<th>Device comparison</th>
<th>A versus B</th>
<th>B versus C</th>
<th>A versus C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute mean (sd)</td>
<td>553.52 (90.46)</td>
<td>563.45 (91.08)</td>
<td>543.37 (91.51)</td>
</tr>
<tr>
<td>Mean difference (95% CI)</td>
<td>$-9.93 (-12.37, -7.48)$</td>
<td>$20.08 (17.85, 22.29)$</td>
<td>$10.15 (7.68, 12.61)$</td>
</tr>
</tbody>
</table>

**Figure 1** (a) Bland and Altman plot, A v B; (b) Bland and Altmann plot, B v C; (c) Bland and Altmann plot, A v C.
min, \( P < 0.0001 \) between B and C and 10.15 l/min, \( P < 0.0001 \) between A and C (see Table 2).

3. Bland and Altman plots: Bland and Altman plots show the mean difference between meters under investigation and limits of agreement (mean difference ± 2 standard deviations). For A versus B the limits of agreement were -60.12 and 40.26 l/min. For B versus C, the limits of agreement were -25.62 and 65.76 l/min and for A versus C the limits of agreement were -40.63 and 60.93 l/min (Figs. 1a–c).

Discussion

Despite an apparent association between the three peak flow meters, formal methods and agreement analysis show that there were, in fact, important differences between the devices. Each Bland and Altman plot (Fig. 1) shows a substantial lack of agreement between the meters. The limits of agreement further confirm these discrepancies. When comparing devices A and B, the difference between PEF values obtained could have ranged anywhere from 40.26 to -60.12 l/min. Between meters B and C, the difference could have been from 65.76 to -25.62 l/min and for A and C, the true difference could have been as high as 60.93 l/min or low as -40.63 l/min.

Our study confirms the conclusions drawn from previous studies. However, the results obtained from previous studies are based largely on: (i) non-human trials; (ii) non-blinded designs; (iii) basic methods of analysis; and (iv) small sample sizes. Furthermore, unlike previous work, the results of this study are less likely to be limited by methodological limitations or biases.

Nevertheless we recognise that the findings of this study need confirmation in other settings with a clinical population and subjects of varying age. Values obtained from peak flow meters must be accurate and reliable to prevent unnecessary diagnosis and clinical mismanagement. The differences of up to 65 l/min observed in this study, suggest that peak flow meters are not interchangeable in clinical practice. Therefore, the same meter should be used in the management of a particular patient. However, even this may cause problems as peak flow meters may become inaccurate with time. In light of our findings, peak flow meters should be calibrated regularly and readings obtained from them should not be viewed in isolation but should be interpreted cautiously and in conjunction with other spirometric values and clinical symptoms.

References