Impact of spirometry on GPs’ diagnostic differentiation and decision-making

Niels Chavannes\textsuperscript{a,b,*}, Tjard Schermer\textsuperscript{c}, Reinier Akkermans\textsuperscript{c}, J.E. Jacobs\textsuperscript{a}, Gabrielle van de Graaf\textsuperscript{b}, Ralf Bollen\textsuperscript{b}, Onno van Schayck\textsuperscript{b}, Ben Bottema\textsuperscript{a,c}

\textsuperscript{a}Centre for Quality of Care Research, University Medical Centre St Radboud, Nijmegen, The Netherlands
\textsuperscript{b}Department of General Practice, Caphri Research Institute, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
\textsuperscript{c}Department of General Practice/Family Medicine, University Medical Centre St Radboud, Nijmegen, The Netherlands

Received 11 December 2003; accepted 12 April 2004

Summary Background: Spirometry is increasingly implemented in general practice, while the ability of general practitioners (GPs) to interpret flow–volume curves (F–V curves) has been questioned. Furthermore, the role of spirometry in the GPs decision-making process has barely been studied.

Aim: To compare the achievements of trained GPs in spirometric diagnosis with an expert consensus panel (1) and to assess the influence of spirometry on the GPs decision-making (2).

Method: Twelve cases including a wide range of F–V curves were interpreted by 39 GPs as well as the expert panel. Diagnostic test characteristics were calculated using multi-level analysis and summarised by diagnostic odds ratios (DOR). Differences in decision-making indicators were expressed as odds ratios and 95% confidence intervals.

Results: Normal F–V curves (DOR 65.0) and obstructive F–V curves (DOR 48.9) were reasonably well diagnosed, while rare and mixed pathological patterns achieved considerably lower scores (DOR 3.8). Intermediate scores were obtained in the recognition of incorrect test manoeuvres (DOR 24.4). Spirometry influenced the GPs decision-making in reducing the number of alternative diagnoses (OR 0.266 [0.200, 0.353]), but also increased referral rates (7.26 [4.71, 11.2]) and the use of diagnostic prednisolone courses (4.55 [3.12, 6.64]) substantially.

Conclusion: Trained GPs were able to differentiate between normal and obstructive disease patterns, while F–V curves suggestive of rare and mixed pathology were often missed. Spirometry seems to influence the decision-making process of the GP; whether this represents an initial or a more sustained effect remains to be evaluated in studies of daily primary care practice.

\&2004 Elsevier Ltd. All rights reserved.
Introduction

In general practice, medical history taking and physical examination are the most important instruments to establish diagnosis and initiate treatment. Diagnostic tools originating from secondary care settings such as electrocardiography and spirometry are increasingly used in primary care and the results are being interpreted by general practitioners (GPs). Access to spirometry in primary care has increased rapidly in the past years, surveys ranging from 21% (1998) to 77% (2001) in the UK. By contrast, spirometer utilization is hampered by insufficient training: less than half are used to diagnose COPD. Several national and international guidelines consider formal spirometric testing essential to establish a diagnosis of COPD, while education in its use has been identified as a major goal for primary care physicians. However, the value of spirometry in differentiating between specific respiratory disease patterns still needs to be assessed in general practice. Most authors focus on the quality of spirometry test performance, while studies investigating the interpretative skills of physicians report rather disappointing results, both in primary and secondary care setting.

A number of studies in COPD and asthma suggest that spirometry could reduce both under- and overdiagnosis of obstructive airway disease in general practice, which might influence disease management. Adjustment of treatment after spirometry has been reported in 4–25% of patients with mostly asthmatic complaints. However, the direct influence of spirometry on the decision-making process of GPs has not been assessed. Therefore, the aim of the current study was to determine the achievements of GPs in differentiating between various chronic respiratory diseases when spirometry is provided as a supplementary diagnostic tool. In addition, we investigated the impact of the flow–volume curve on the GPs decision-making process.

Methods

Participants and spirometry training

GPs with an interest in spirometry were recruited from the general practice networks of the Nijmegen and Maastricht Universities in the Netherlands. Most of these GPs already used spirometry in daily practice, had received previous training and were motivated to assess their skills. Additionally, GPs involved in the vocational training in the Nijmegen and Maastricht regions were invited by postal mailing to participate in the study.

Participating GPs received a standardised postgraduate spirometry training course (two three-hours sessions with an interval of one month), and could bring their newly acquired spirometric knowledge and skills into practice for a period of six to nine months before the study started. The spirometry course was based on a format widely used in the Netherlands. During the first session the focus was mainly on the pathogenesis and clinical characteristics of asthma, COPD and other chronic respiratory diseases; theoretical concepts of lung function testing; execution of spirometry tests; and practical guidelines and strategies for spirometry interpretation. The second session was mainly used to discuss actual case descriptions submitted by either the participants or course leaders. Training was provided by a pulmonologist and an experienced lung function technician. Interactive education and feedback on the spot were emphasized throughout the course.

Standardised case descriptions

A set of 12 standardised case descriptions was constructed, based on actual patients from two general practices from our academic networks. The cases were designed in cooperation with a pulmonologist and a GP with ample experience in the field of chronic respiratory diseases. The case set included a range of typical flow–volume curves suggesting mild obstruction (n = 1); moderate obstruction (n = 1); severe obstruction (n = 2); rare pattern of restriction (n = 1); fixed upper airways obstruction (n = 1); mixed pattern of both obstruction and restriction (n = 1); incorrect test manoeuvres (n = 2); and normal curves (n = 3). The participating GPs worked through two sets of six cases each, which were assessed in random order within a period of one year. Randomisation codes were prepared by a fellow-researcher who was not involved in the study and stored in sealed envelopes until use. Data were collected in the period July 1999 through April 2001.

A research assistant visited the GPs in their practice. For each case, a concise medical history and results of the physical examination were presented to the GP first. Subsequently, absolute and predicted postbronchodilator spirometry test results (including FEV1, FEV1/FVC and flow–volume curves) were provided. After having assessed a case, GPs had to select one spirometric diagnosis.
from a preformulated list. An example of the case structure is depicted in Fig. 1.

Before the study, the 12 paper cases had been judged by an independent expert panel consisting of two pulmonologists, a pulmonary physiologist and a GP with specific expertise in the pulmonary field. The panel reached consensus on the spirometric and clinical diagnoses of the paper cases during a panel discussion meeting, while no cases were excluded. The panel meeting was audiotaped and independently scored by two of the authors (NC and TS) in order to establish the panels’ final diagnosis and—as when applicable—alternative diagnoses for each case. There was 100% agreement between the two observers with respect to the panels’ final and alternative diagnoses. The panel consensus diagnoses served as ‘the gold standard’ in the subsequent evaluation of the GPs’ diagnostic achievements.

Outcome measures

To assess the diagnostic achievements of GPs with regard to interpretation of spirometry, the following four outcome categories were considered most relevant and contrasting from a clinical point of view: (1) bronchial obstruction (from mild to severe); (2) rare respiratory pathology (i.e., restriction, fixed upper airways obstruction, mixed pattern); (3) normal lung function; and (4) incorrect test manoeuvre.

In addition, the impact of spirometry on the GPs’ decision-making process was assessed using four indicators: (1) diagnostic uncertainty (size of differential diagnosis, i.e. the number of alternative diagnoses considered by the GPs while assessing a case); (2) probability of prescribing respiratory medication; (3) probability of initiating a diagnostic prednisolone course, a commonly used test (albeit its’ value is uncertain); and (4) probability of referral to a pulmonologist and/or cardiologist. These process indicators of GP decision-making were assessed before and after the results of spirometry were shown to the GPs (Fig. 1).

Statistical analyses

First, the agreement between the GPs’ interpretations and the expert panels’ gold standard diagnoses was investigated univariately using the SPSS software package (Version 9.0 for Windows). Subsequently, multi-level linear and logistic modelling was used to account for the intra-cluster correlation induced by the fact that each GP

Figure 1  Schematic representation of case structure 'mild to moderate COPD'.
assessed more than one case. SAS® software (Release 6.12 for Windows) was used for these multi-level analyses.

The following diagnostic test characteristics were calculated for each outcome measure: positive and negative predictive values (further referred to as PPV and NPV, respectively), positive and negative likelihood ratios (LR + and LR −, respectively) and the diagnostic odds ratio (DOR). PPV expresses the probability of disease in subjects with a positive test result, NPV the probability of absence of disease in subjects with a negative test result. LR + is the ratio of the probability of a positive test in subjects with disease and the probability of a positive test in subjects without disease. Conversely, LR − is the ratio of the probability of a negative test in subjects with disease and the probability of a negative test in subjects without disease. A diagnostic test is better the more LR differs from 1, that is, greater than 1 for LR + and lower than 1 for LR −. Finally, the DOR summarises the overall discrimination of a diagnostic test with a dichotomous outcome. In fact, it is the ratio of LR + and LR −. Therefore, a diagnostic test is useless if DOR = 1.21

After the four indicators of decision-making were dichotomised (1 vs. > 1 diagnosis; 0 vs. 1 diagnostic prednisolone course; 0 vs. 1 or more referrals; 0 vs. 1 or more prescriptions), the before-after spirometry measurements were compared using multi-level logistic regression analysis and expressed as odds ratios with 95% confidence intervals (95% CI).

Results

General practitioners

Thirty-nine (39) GPs participated in the study. Three GPs dropped out during the study, one because of early retirement, the others due to loss of interest. These three GPs completed one set of six cases, instead of both sets. Table 1 shows that the study population consisted predominantly of middle-aged male doctors who had been using spirometry in their daily practice for a mean of 4.3 (SD 3.7) years, having received 4.2 (SD 4.9) hours of spirometry training in the year preceding the experiment.

Diagnostic achievements by GPs

Altogether, the GPs assessed 444 cases. Table 2 shows the agreement between GP judgements and expert panel for each of the diagnostic outcome categories. Concordance with the expert panel regarding obstruction was present in 91.3% [95% CI 86.8, 95.8] of cases, followed by normal spirometry obtaining 77.9% [95% CI 70.2, 85.6] correct answers, while incorrect manoeuvres reached a score of 64.9% [95% CI 54.0, 75.8], and rare pathological curves were recognised in 41.3% [95% CI 32.1, 50.5] of cases.

Table 3 shows that normal and obstructive curves were characterised by high DORs: 65.0 and 48.9, respectively. By contrast, rare pathological curves obtained a low DOR of 3.8. Scoring of an incorrect test manoeuvre generated an intermediate DOR of 24.4. The negative predictive values (probability of rightly ruling out disease) varied between 0.93 and 0.96, except for rare pathology, which reached 0.82. Positive predictive values (probability of rightly labelling disease) however, revealed a range of values between 0.87 (normal curves) and 0.49 (rare pathology).

Indicators of GPs decision-making

Before spirometry, GPs considered an average of 2.05 diagnoses per case, with a maximum of eight, while after spirometry this was reduced to a mean of 1.35, with a maximum of six. Table 4 quantifies this significant reduction of diagnostic uncertainty: > 1 diagnosis is considered in 59.6% [55.1, 64.1] of cases before spirometry, while after spirometry > 1 diagnosis is considered in 31.2% [26.9, 35.5] of cases (OR 0.266 [0.200, 0.353]). Conversely, spirometry significantly increases the number of diagnostic prednisolone courses and the referral rate, while the proportion of cases where medication is prescribed increases, but not significantly. The probability of diagnostic prednisolone testing rises three-fold, from 8.0% [5.5, 10.5] to 27.6% [23.5, 31.7] per case (OR 4.55 [3.12, 6.64]) as a result of

---

**Table 1** Characteristics of participating general practitioners (n = 36)*.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/female</td>
<td>33/3</td>
</tr>
<tr>
<td>Age, years</td>
<td>48 (5.1)</td>
</tr>
<tr>
<td>Patients enlisted per GP, number</td>
<td>2086 (712)</td>
</tr>
<tr>
<td>Surgery hours per week, hours</td>
<td>43.6 (12.7)</td>
</tr>
<tr>
<td>Use of spirometer in daily patient care, yes/no</td>
<td>35/1</td>
</tr>
<tr>
<td>Duration of spirometry utilisation, years</td>
<td>4.3 (3.7)</td>
</tr>
<tr>
<td>Spirometry training in previous year, hours</td>
<td>4.2 (4.9)</td>
</tr>
</tbody>
</table>

Values are means (standard deviation) unless stated otherwise.

*Data missing on 3 GPs who dropped out during the study.
Table 2  Agreement between expert panel and GP judgement for the presence (or absence) of obstructive disease (A), rare pathology (B), normal spirometry (C), and incorrect manoeuvre (D).

<table>
<thead>
<tr>
<th>Expert panel judgement</th>
<th>Obstruction</th>
<th>No obstructiona</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) GP judgement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obstruction</td>
<td>136 (31)</td>
<td>52 (12)</td>
</tr>
<tr>
<td>No obstructiona</td>
<td>13 (3)</td>
<td>243 (55)</td>
</tr>
<tr>
<td></td>
<td>149 (34)</td>
<td>295 (66)</td>
</tr>
</tbody>
</table>

| Rare pathology         | 45 (10)     | 52 (12)         |
|                       | 64 (14)     | 283 (64)        |
|                       | 109 (25)    | 335 (75)        |

| Normal spirometry      | 88 (20)     | 17 (4)          |
|                        | 25 (6)      | 314 (71)        |
|                        | 113 (25)    | 331 (75)        |

| Incorrect Manoeuvre    | 48 (11)     | 26 (6)          |
|                        | 26 (6)      | 344 (77)        |
|                        | 74 (17)     | 370 (83)        |

Percentages of total number of cases within parenthesis.

aEither rare pathology, normal spirometry, or incorrect manoeuvre.
bEither obstruction, normal spirometry, or incorrect manoeuvre.
cEither obstruction, rare pathology, or incorrect manoeuvre.
dEither obstruction, rare pathology, or normal spirometry.

Table 3  Predictive values, likelihood ratios and diagnostic odds ratios for general practitioners diagnoses.

<table>
<thead>
<tr>
<th></th>
<th>PPV</th>
<th>NPV</th>
<th>LR+</th>
<th>LR−</th>
<th>DOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal spirometry</td>
<td>0.87</td>
<td>0.93</td>
<td>15.16</td>
<td>0.23</td>
<td>65.0</td>
</tr>
<tr>
<td>Obstructive disease</td>
<td>0.75</td>
<td>0.96</td>
<td>5.18</td>
<td>0.11</td>
<td>48.9</td>
</tr>
<tr>
<td>Incorrect manoeuvre</td>
<td>0.68</td>
<td>0.93</td>
<td>9.23</td>
<td>0.38</td>
<td>24.4</td>
</tr>
<tr>
<td>Rare pathology</td>
<td>0.49</td>
<td>0.82</td>
<td>2.66</td>
<td>0.70</td>
<td>3.8</td>
</tr>
</tbody>
</table>

PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR−: negative likelihood ratio; DOR: diagnostic odds ratio.

Table 4  Impact of flow-volume curve on indicators of the decision-making process in general practitioners.

<table>
<thead>
<tr>
<th>Process indicators</th>
<th>Before F/V-curve</th>
<th>After F/V-curve</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic uncertaintya</td>
<td>59.6% (55.1, 64.1)</td>
<td>31.2% (26.9, 35.5)</td>
<td>0.266 (0.200, 0.353)</td>
</tr>
<tr>
<td>Prednisolon course</td>
<td>8.0% (5.5, 10.5)</td>
<td>27.6% (23.5, 31.7)</td>
<td>4.55 (3.12, 6.64)</td>
</tr>
<tr>
<td>Referral rate</td>
<td>6.0% (3.8, 8.2)</td>
<td>31.7% (27.4, 36.0)</td>
<td>7.26 (4.71, 11.2)</td>
</tr>
<tr>
<td>Medication prescription</td>
<td>36.5% (32.0, 41.0)</td>
<td>39.4% (34.9, 43.9)</td>
<td>1.14 (0.87, 1.50)</td>
</tr>
</tbody>
</table>

Numbers are percentages with 95% confidence intervals.
aProportion of >1 diagnoses in the differential diagnosis.
spirometry. The probability of referral changes from 6.0% [3.8, 8.2] to 31.7% [27.4, 36.0] as a result of spirometry (OR 7.26 [4.71, 11.2]).

Discussion

The present study demonstrates for the first time the reasonable diagnostic achievements of trained GPs with regard to commonly encountered spirometric patterns. Curves of obstructive airways disease as well as the physiological can be considered the more prevalent conditions, as opposed to patterns suggestive of restriction or fixed upper airways obstruction, which GPs can be expected to be less familiar with. On the whole, the positive predictive values are lower than the negative predictive values. This reflects the fact that in primary care it remains more difficult to label a disease than to exclude it, due to the lower a priori probability. The relatively low positive predictive value of an incorrect test manoeuvre illustrates the need to emphasise the importance of quality assessment of the flow-volume curve, which should precede interpretation. The low diagnostic achievements in the less prevalent categories points out the paradoxical necessity of recognising patterns one does not understand. Another explanation might be that dynamic spirometry is of limited use in differentiating between normal and restrictive disease, thus contributing to the low diagnostic yield. These elements should receive considerable attention in future spirometry courses, professional supervision or automated supportive software.

The significant influence of the flow-volume curve on the trained GPs’ decision-making is expressed in a reduction of the number of alternative diagnoses but an increase in referral rates and diagnostic prednisolone courses. Thus, the flow-volume curve seems to support establishing a diagnosis in patients with respiratory morbidity, but probably leads to an increased use of additional diagnostic procedures or specialist care, at least initially. This could partially reflect the relatively high prevalence of pathology in this specific case-set, warranting further work-up. Another explanation could be that the number of options was limited; for example, an option to repeat spirometry to verify correctness of the manoeuvre was missing, possibly leading to increased prednisolone testing or referral instead. The current design does not allow us to deduct if this initial increase would be sustained in time, nor does it predict the exact effect size in daily practice.

Spirometers are increasingly available but seem underused, while doctors have been observed to overestimate their actual interpretative skills in spirometry, as well as in ECGs. This underlines that training is a prerequisite for meaningful implementation of advanced diagnostic tools in primary care. Both quality assessment and pattern recognition have been part of our standardised spirometry training course, which took place 6–9 months preceding the measurements, allowing the primary care physicians to integrate skills in daily practice. The format and duration of the training were directly derived from a common postgraduate spirometry course, which has been attended by large numbers of Dutch primary care physicians in the past few years.

The results of the present study reflect the ability of trained GPs to diagnose this specific case-set. Therefore, we do not pretend to reflect actual prevalences of the disease patterns within the constitution of the cases. By analysing spirometric patterns separately this over-representation is corrected for. Consequently, the multi-level analysis was performed to account for intra-cluster correlation within the GPs. However, it remains to be investigated what the results will be in a real-life setting, with actual patients and less or even untrained GPs. The case-set structure allowed us to compare the level of pattern recognition quite precisely with an expert panel, which was confronted with the identical set of cases. Moreover, the expert panel scored cases preceding the study, independent of the results of the primary care physicians, thereby eliminating a potential bias which might have been overlooked in previous studies.

In this study we demonstrated that the novel method of combining standardised case material with techniques of multi-level analysis may be useful to evaluate complex diagnostic tools, like spirometry. We conclude that trained GPs were able to differentiate between normal and obstructive disease patterns, while F–V curves suggestive of rare and mixed pathology were often missed. Spirometry seems to influence the decision-making process of the GP by reducing diagnostic uncertainty but increasing use of additional diagnostics and referral to specialist care. Whether this represents an initial or a more sustained effect remains to be evaluated in studies of daily primary care practice.

Acknowledgements

The authors are grateful to Dr. JWM Muris, GP, and Dr. GJ Wesseling, pulmonologist, who kindly assisted the preparation of the cases used in this.
study. We also wish to thank all participating GPs for their interpretative efforts.

References


