Effect of e-Learning and Repeated Performance Feedback on Spirometry Test Quality in Family Practice: A Cluster Trial

Tjard R. Schermer, MSc, PhD
Reinier P. Akkermans, MSc
Alan J. Crockett, MD, PhD
Marian van Montfort, MD
Joke Grootens-Stekelenburg
Jim W. Stout, MD, PhD
Willem Pieters, MD, PhD

1Department of Primary and Community Care, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
2Discipline of General Practice, The University of Adelaide, Adelaide, Australia
3QUARTZ Integrated Care Support Service, Helmond, The Netherlands
4Department of Pediatrics, University of Washington, Seattle, Washington
5Department of Pulmonary Medicine, Eilkeriek Hospital, Helmond, The Netherlands

ABSTRACT

PURPOSE Spirometry has become an indispensable tool in primary care to exclude, diagnose, and monitor chronic respiratory conditions, but the quality of spirometry tests in family practices is a reason for concern. Aim of this study was to investigate whether a combination of e-learning and bimonthly performance feedback would improve spirometry test quality in family practices in the course of 1 year.

METHODS Our study was a cluster trial with 19 family practices allocated to intervention or control conditions through minimization. Intervention consisted of e-learning and bimonthly feedback reports to practice nurses. Control practices received only the joint baseline workshop. Spirometry quality was assessed by independent lung function technicians. Two outcomes were defined, with the difference between rates of tests with 2 acceptable and repeatable blows being the primary outcome and the difference between rates of tests with 2 acceptable blows being the secondary outcome. We used multilevel logistic regression analysis to calculate odds ratios (ORs) for an adequate test in intervention group practices.

RESULTS We analyzed 1,135 tests. Rate of adequate tests was 33% in intervention and 30% in control group practices (OR = 1.3; \(P = .605\)). Adequacy of tests did not differ between groups but tended to increase with time: OR = 2.2 (\(P = .057\)) after 3 and OR = 2.0 (\(P = .086\)) in intervention group practices after 4 feedback reports. When ignoring test repeatability, these differences between the groups were slightly more pronounced: OR = 2.4 (\(P = .033\)) after 3 and OR = 2.2 (\(P = .051\)) after 4 feedback reports.

CONCLUSIONS In the course of 1 year, we observed a small and late effect of e-learning and repeated feedback on the quality of spirometry as performed by family practice nurses. This intervention does not seem to compensate the lack of rigorous training and experience in performing spirometry tests in most practices.


INTRODUCTION

Spirometry has become an indispensable tool for primary care professionals to diagnose and monitor chronic respiratory conditions. 1,2 Recent studies indicate that, when implemented in primary care, spirometry is a valid test that leads to increased rates of respiratory diagnoses and may improve disease management in the United States and elsewhere. 4,7 Although spirometry can be made available in several ways, from a practical point of view having good-quality tests performed in the practice itself is the preferred mode. 8

In the Netherlands approximately 62% of family practices own a spirometer (the remaining practices having access to spirometry facilities elsewhere), and the rate of spirometry tests performed in family practices has tripled in the past couple of years. 9 Recent surveys in the United States indicate that 47% to 75% of family physicians use spirometry. 10,11
Lack of spirometry training appears to be an important barrier to the use of spirometry in US family practices. Performing good quality spirometry requires proper training and well-standardized procedures, which may be difficult to achieve in a family practice.

Using different sets of criteria, investigators have studied the quality of spirometry tests in family practices and reported rates of adequate tests ranging from only 3% to up to 80%. Although there are some indications that training and performance feedback may increase the quality of spirometry in nonlaboratory settings, only 1 controlled intervention trial has been published from a real-life family practice setting. In the study reported in this article, we investigated whether a combined intervention of e-learning and subsequent bimonthly performance feedback after a baseline spirometry workshop would improve test quality in family practices in the course of 1 year.

METHODS
Study Design
The study was a cluster controlled trial, with family practices as clusters for patients in whom 1 or more spirometry tests were performed during a 1-year observation period (Clinicaltrials.gov Protocol Registration System: NCT00962455; http://www.clinicaltrials.gov). All practices involved were nonacademic practices that have a working agreement with a local hospital (Elkerliek Hospital, Helmond, The Netherlands) regarding support in spirometry training, test execution, and interpretation. All practices have a PC-based spirometer (SpiroPerfect, Welch Allyn, Delft, The Netherlands). The spirometry tests are electronically submitted by the practices and can be accessed by the hospital’s lung function technicians and chest physicians. The Medical Ethics Committee of the Elkerliek Hospital approved the study (file number 07-393).

Recruitment, Sample Size, and Group Allocation
Practices were recruited through a mailing to all family practices that collaborate with the Elkerliek Hospital. Cluster sample size calculation showed that 19 practices were needed to demonstrate a 25% difference in the rate of adequate spirometry tests. Assumptions for the power calculation were as follows: 30% adequate spirometry tests in control practices, 12-month study duration; average of 2 spirometry tests per practice per week, intracluster correlation coefficient (ICC) of 0.15, α level of 0.05, 1−β of 0.80. Although we anticipated a learning curve in intervention effects over time, we did not take this into account in the sample size calculation. After recruitment, practices received a questionnaire to inquire about their spirometry details and were allocated to either the intervention or usual practice group through computerized minimization, a method of ensuring balance between groups for several prognostic factors, even in small samples. Allocation was stratified by the average weekly number of spirometry tests in the past year and by practice nurses’ years of spirometry experience.

Intervention and Control Conditions
All practice nurses involved in the study attended a 2.5-hour baseline workshop to refresh their spirometry knowledge and skills. Current criteria for adequate spirometry tests were discussed, and execution of tests was practiced. At the end of the workshop, all nurses were encouraged to continue or increase their usual frequency of spirometry testing. Intervention practice nurses stayed for an additional 30 minutes to (1) receive a copy of the e-learning CD-ROM Spirometry Fundamentals (University of Washington, Seattle, Washington), (2) be instructed in how to study the CD-ROM in the next 4 weeks, and (3) discuss details of the upcoming bimonthly feedback reports—including the way these reports should be interpreted and discussed during telephone calls with the lung function technician who had been assigned as their coach. The feedback reports focused on the spirometry tests performed in the past 2- to 2.5-month period and, after the first period, on comparison with previous periods. All practices were instructed that (1) only trained nurses should perform spirometry, (2) all tests performed as a part of routine patient care should be submitted, and (3) each submitted test should consist of the 3 best blows.

The observation period for the trial started immediately after the spirometry workshop (on November 6, 2007) and lasted exactly 12 months. A new feedback period started on the day the feedback reports that were based on the tests from the previous period had been e-mailed to the intervention practices. After electronic submission, the quality of prebronchodilator tests from intervention practices was assessed by one of the Elkerliek Hospital’s lung function technicians, using a checklist based on recent recommendations. Because the lung function technicians could invest a limited number of hours to assess spirometry tests for the study, 25 tests per practice per period was set as the maximum number of tests on which a feedback report could be based.

Spirometry Quality Assessment and Study Outcomes
Three experienced lung function technicians who were not otherwise involved in the study assessed de-identified printouts of all submitted tests. These technicians were instructed on how to score spirometry test quality, practiced on 10 tests, and discussed their scoring method.
and experiences. Two technicians volunteered to assess all tests for the trial. Interobserver agreement between these 2 technicians was assessed in a random sample of 80 tests and was deemed sufficient (Cohen’s $\kappa = 0.79$). In case of disagreement, the third technician assessed the test. Tests were presented to the technicians in random order, and they were blinded to each other’s assessments and to information on the practice or study period to which a test belonged.

We only used prebronchodilator tests from patients older than 10 years. The primary outcome for the study was the proportion of tests with 2 or more acceptable blows that were also repeatable for both FEV$_1$ (forced expiratory volume in 1 second) and FVC (forced vital capacity).

An adequate forced blow has a good start, satisfactory duration of exhalation, and is free from artifacts (e.g., cough, glottis closure, obstruction of mouthpiece). Because the feedback report focused strongly on acceptability of blows and less so on repeatability between blows, we studied differences in the proportions of tests with 2 or more acceptable blows as a secondary outcome.

Statistical Analysis

Multilevel logistic regression analysis with family practice as cluster level and a compound symmetry matrix correlation structure was used to test differences in the odds of an adequate spirometry test (GLIMMIX procedure, SAS 9.2, SAS Institute, Cary, North Carolina). Odds ratios (ORs) derived from the logistic regression models express the odds (and 95% confidence interval [CI]) of an adequate test in intervention relative to control practices. The logistic model also included the period in which the test had been performed and covariates related to test quality (i.e., age, sex, and severity of airflow obstruction). ICCs were calculated for the primary and secondary outcomes. Statistical tests were 2-sided, $P < .05$ was considered statistically significant.

RESULTS

Practices, Patients, and Spirometry Tests

Figure 1 displays the recruitment and allocation of practices and reasons for excluding spirometry tests from the analysis. One of the 9 intervention practices dropped out before the first feedback report was provided. A total of 490 prebronchodilator spirometry tests from intervention and 645 tests from usual prac-
SPIROMETRY TEST QUALITY

Table 1. Characteristics of Family Practices and Patients Tested With Spirometry

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention</th>
<th>Control</th>
<th>P Value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of practices</td>
<td>9</td>
<td>10&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Practice type, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solo</td>
<td>2 (22)</td>
<td>2 (22)</td>
<td></td>
</tr>
<tr>
<td>Duo</td>
<td>1 (11)</td>
<td>3 (33)</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>5 (56)</td>
<td>2 (22)</td>
<td></td>
</tr>
<tr>
<td>Multidisciplinary health care center</td>
<td>1 (11)</td>
<td>2 (22)</td>
<td></td>
</tr>
<tr>
<td>Physicians per practice, mean (SD), n</td>
<td>3.2 (1.6)</td>
<td>2.3 (1.7)</td>
<td>.582</td>
</tr>
<tr>
<td>Practice size, patients, mean (SD), n</td>
<td>6,521 (3,428)</td>
<td>4,754 (3,012)</td>
<td>.263</td>
</tr>
<tr>
<td>Spirometry test operators per practice, mean (SD), n</td>
<td>2.0 (0.9)</td>
<td>2.3 (0.9)</td>
<td>.461</td>
</tr>
<tr>
<td>Spirometry experience, mean (SD), y</td>
<td>5.6 (3.0)</td>
<td>5.0 (3.0)</td>
<td>.673</td>
</tr>
<tr>
<td>Spirometry tests per month, mean (SD), n</td>
<td>13.7 (15.0)</td>
<td>16.6 (10.6)</td>
<td>.644</td>
</tr>
<tr>
<td>Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spirometry tests, n</td>
<td>490</td>
<td>645</td>
<td></td>
</tr>
<tr>
<td>Unique patients, n</td>
<td>457</td>
<td>615</td>
<td></td>
</tr>
<tr>
<td>Age, mean (SD), y</td>
<td>53.3 (16.1)</td>
<td>53.9 (17.9)</td>
<td>.550</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>245 (50)</td>
<td>303 (47)</td>
<td>.421&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;, mean (SD), L&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2.56 (0.97)</td>
<td>2.55 (0.98)</td>
<td>.829</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;, as % of predicted value, mean (SD)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>84.5 (23.4)</td>
<td>84.0 (20.8)</td>
<td>.719</td>
</tr>
</tbody>
</table>

FEV<sub>1</sub> = forced expiratory volume in 1 second; FVC = forced vital capacity.
<sup>a</sup> From Student t test for independent samples.
<sup>b</sup> One baseline questionnaire was not returned in the usual practice group (after 2 reminders).
<sup>c</sup> From Pearson χ<sup>2</sup> test.
<sup>d</sup> From prebronchodilator spirometry test.
<sup>e</sup> Predicted values from the European Community for Steel and Coal (ECSC), 1993.

Table 2. Reasons for Lack of Acceptability of Forced Expiratory Blows in the Intervention and Control Groups

<table>
<thead>
<tr>
<th>Reason</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
<th>Period 5</th>
<th>All Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-Learning + feedback, blow (test), n</td>
<td>306 (102)</td>
<td>330 (113)</td>
<td>261 (87)</td>
<td>273 (91)</td>
<td>291 (97)</td>
<td>1470 (490)</td>
</tr>
<tr>
<td>Usual practice, blow (test), n</td>
<td>348 (116)</td>
<td>456 (152)</td>
<td>348 (116)</td>
<td>405 (135)</td>
<td>378 (126)</td>
<td>1935 (645)</td>
</tr>
<tr>
<td>Blows with poor start</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-Learning + feedback, n (%)</td>
<td>21 (20.6)</td>
<td>23 (20.4)</td>
<td>19 (21.8)</td>
<td>16 (17.6)</td>
<td>16 (16.5)</td>
<td>95 (19.4)</td>
</tr>
<tr>
<td>Usual practice, n (%)</td>
<td>27 (23.3)</td>
<td>33 (21.7)</td>
<td>28 (24.1)</td>
<td>37 (27.4)</td>
<td>25 (19.8)</td>
<td>150 (23.3)</td>
</tr>
<tr>
<td>P value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.857</td>
<td>0.931</td>
<td>0.933</td>
<td>0.095</td>
<td>0.664</td>
<td>0.474</td>
</tr>
<tr>
<td>Blows with artifacts during exhalation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-Learning + feedback, n (%)</td>
<td>11 (10.8)</td>
<td>23 (20.4)</td>
<td>18 (20.7)</td>
<td>17 (18.7)</td>
<td>14 (14.4)</td>
<td>83 (16.9)</td>
</tr>
<tr>
<td>Usual practice, n (%)</td>
<td>11 (9.5)</td>
<td>17 (11.2)</td>
<td>13 (11.2)</td>
<td>23 (17.0)</td>
<td>14 (11.1)</td>
<td>78 (12.1)</td>
</tr>
<tr>
<td>P value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.644</td>
<td>0.026</td>
<td>0.214</td>
<td>0.987</td>
<td>0.757</td>
<td>0.311</td>
</tr>
<tr>
<td>Blows with unsatisfactory exhalation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abrupt end</td>
<td>36 (5.9)</td>
<td>15 (13.3)</td>
<td>10 (11.5)</td>
<td>5 (5.5)</td>
<td>5 (5.2)</td>
<td>41 (8.4)</td>
</tr>
<tr>
<td>Usual practice, n (%)</td>
<td>12 (10.3)</td>
<td>9 (5.9)</td>
<td>3 (2.6)</td>
<td>9 (6.7)</td>
<td>5 (4.0)</td>
<td>38 (5.9)</td>
</tr>
<tr>
<td>P value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.190</td>
<td>0.410</td>
<td>0.082</td>
<td>0.710</td>
<td>0.727</td>
<td>0.702</td>
</tr>
<tr>
<td>Duration &lt;6 sec or no plateau in volume-time curve</td>
<td>35 (34.3)</td>
<td>38 (33.6)</td>
<td>30 (34.5)</td>
<td>29 (31.9)</td>
<td>28 (28.9)</td>
<td>160 (32.7)</td>
</tr>
<tr>
<td>e-Learning + feedback, n (%)</td>
<td>55 (47.5)</td>
<td>66 (41.4)</td>
<td>44 (37.9)</td>
<td>69 (51.1)</td>
<td>49 (38.9)</td>
<td>280 (43.4)</td>
</tr>
<tr>
<td>Usual practice, n (%)</td>
<td>55 (47.5)</td>
<td>63 (41.4)</td>
<td>44 (37.9)</td>
<td>69 (51.1)</td>
<td>49 (38.9)</td>
<td>280 (43.4)</td>
</tr>
<tr>
<td>P value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.180</td>
<td>0.380</td>
<td>0.749</td>
<td>0.171</td>
<td>0.854</td>
<td>0.371</td>
</tr>
</tbody>
</table>

<sup>a</sup> From multilevel logistic regression analysis.
quality criteria. Overall, poor start (21.6%) and insufficient duration of blows (38.8%) were the predominant reasons for test inadequacy, but no statistically significant differences between intervention and control group practices were observed. For the primary outcome, the rate of adequate tests in the first 2 months (ie, after the baseline workshop and initial e-learning in intervention practices, but before the first feedback report) was 30.4% in intervention and 25.0% in control practices. The odds ratio for an adequate test in intervention relative to control practices was 1.3 (95% CI, 0.5-3.2; \( P = .605 \)). For all 5 study periods combined, the rate of adequate tests was 32.9% in the intervention group and 29.8% in the control group (OR = 1.2, 95% CI, 0.6-2.5; \( P = .663 \)). Figure 2 (panel A) shows that the adequacy of tests gradually increased over the consecutive study periods in the intervention group relative to the control group. In the fifth and final study period, rates of adequate tests for the primary outcome were 43.3% and 34.1% for intervention and control practices, respectively. The odds ratio of an adequate spirometry test in the intervention group was not statistically significant during the whole test period, but approached the threshold for significance after 3 feedback reports: OR = 2.2 (95% CI, 1.0-5.0; \( P = .057 \)) in period 4 and OR = 2.0 (95% CI, 0.9-4.5; \( P = .086 \)) in period 5.

When repeatability between blows was ignored in the definition of test adequacy, the differences between the study groups were more pronounced: OR = 2.4 (95% CI, 1.1-5.3; \( P = .033 \)) in period 4 and OR = 2.2 (95% CI, 1.0-4.8; \( P = .051 \)) in period 5 (Figure 2, panel B). Supplemental Figure 1 (available at http://www.annfammed.org/cgi/content/full/9/4/330/DC1) displays the rate of adequate spirometry tests per practice and per period for the primary outcome.

**DISCUSSION**

**Summary of Main Findings**

In a cluster controlled trial we investigated whether a combined intervention of e-learning and bimonthly performance feedback was able to improve the quality of spirometry tests in Dutch family practices that had
already implemented spirometry within their regular patient care. Although we did not observe an intervention effect for the overall 12-month observation period, a modest increase in the rate of adequate tests emerged after 3 feedback reports. This increase led to a trend toward an approximately 10% higher rate of adequate tests in the final 4 months of the observation period in the intervention group. As in previous studies, not meeting end-of-test criteria was the predominant reason for test inadequacy.

Comparison With Existing Literature
Certified and skilful lung function technicians can be expected to produce adequate spirometry tests in 80% or more of adult patients tested. Although our data show that some family practice nurses may be able to achieve rates of adequate tests in 50% to 75% of tests (Supplemental Figure 1), it is unlikely that most practices will be able to achieve and maintain such a performance level. A recent Australian study showed that delegating spirometry to well-trained and experienced visiting nurses substantially improved spirometry adequacy. Adopting this (or a similar) model may overcome practice nurses’ lack of training, experience, and routine—which are essential factors for good-quality spirometry. Studies conducted in specialized settings suggest that performance feedback is able to improve spirometry test quality. Our findings support such an approach, although expectations should be fairly modest, and it may take a rather long breath and more intensive (certified) training programs to achieve satisfactory results.

In our view, obtaining sufficiently reliable and clinically useful spirometry tests—not necessarily perfect tests—is what family practices should be striving for. The impact of inadequate spirometry tests on patients’ diagnoses and management in family practice is currently unclear.

Strengths and Limitations
A particular strength of our study was that it was undertaken in a real-life setting in which a group of family practices had implemented spirometry for several years. Compared with national figures, our group of practices was representative for the Netherlands in terms of spirometry experience, but smaller (ie, single-handed and duo) practices were underrepresented. All practices involved in the evaluation participated in the working agreement with the local hospital, which may limit the generalizability of our findings. Other strengths are the high participation rate, the individualized feedback, the rigorous method of outcome assessment using multiple blinded experts, the clustered trial design and multilevel analyses, and the stratification by spirometry experience of practice nurses. A limitation was the 12-month duration of the study. Although 12 months would seem like a sufficient observation period, the intervention appeared to start having effect after 3 feedback reports. A few additional months of follow-up might have shown either a continuing upward trend, a plateau, or a regression of intervention effects.

During the course of 1 year we found a rather small and late effect of baseline e-learning and repeated performance feedback on the quality of spirometry tests by practice nurses who perform spirometry as a part of regular patient care in family practices. This intervention does not seem to be able to fully compensate for their lack of rigorous training and experience in performing spirometry tests. Other models to provide family practices with good-quality spirometry should be explored.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/9/4/330.

Key words: Spirometry; respiratory function tests; family practice; primary health care; quality assurance, health care; randomized controlled trial; multicenter study; education; pulmonary disease, chronic obstructive; asthma

Submitted September 3, 2010; submitted, revised, January 5, 2011; accepted January 18, 2011.

Funding support: This study was funded by QUARTZ, a regional, integrated care support service; the Elkerliek Hospital; and the Department of Primary and Community Care of the Radboud University Nijmegen Medical Centre, the Netherlands.

An oral presentation at the International Primary Care Respiratory Group (IPCRG) Conference in Toronto, Canada, 2-5 June 2010.

Acknowledgements: The authors wish to express their gratitude to the family practice nurses, family practitioners, chest physicians, and patients who contributed to our study. The authors owe much gratitude to Cyril Welten and Marloes Beekwilder, the lung function technicians from the Elkerliek Hospital in Helmond, who assessed the spirometry tests from the intervention practices and coached the practice nurses from these practices. Without their enthusiasm and commitment we would not have been able to conduct this study. The authors very much appreciate the time and effort put in the study by Oda van Heeswijk, André Hijmans and Kees Lamers, the lung function technicians from the Radboud University Nijmegen Medical Centre, who judged the spirometry tests for the outcome assessment. The authors also appreciate the contributions to the preparatory and supportive work on the study by Janny Wildschut, Wilfried van Huisseling, Rasha Putros, Nicol Orbon, Daan Schellekens, and Hans Vlek.

References


