COPD in primary care: Towards simple prediction of quality of life, exacerbations and mortality

Siebeling, L.

Citation for published version (APA):
CHAPTER 7

ACCURACY OF DIFFERENT METHODS TO ASCERTAIN COPD EXACERBATIONS AND IMPLICATIONS FOR TREATMENT EFFECT ESTIMATES AND SAMPLE SIZE

Anja Frei, Lara Siebeling, Callista Wolters, Patrick Muggensturm, Alexandra Strassmann, Marco Zoller, Gerben ter Riet & Milo A. Puhan

Submitted
Abstract

Background

COPD exacerbations are often ascertained through patient self-reports. Their accuracy is largely unknown. We evaluated the accuracy of COPD exacerbation ascertainment through patient self-reports and single expert adjudication against a reference standard of central event adjudication by a committee (AC). We also assessed the impact of endpoint misclassification on treatment effect estimates and sample size requirements.

Methods

We used an event-based definition of exacerbations. Exacerbations in 409 primary care COPD patients participating in ICE COLD ERIC study over 3 years were ascertained using 6-months follow-ups by patient self-reports and adjudication by one experienced physician, which were compared to consensus classifications by a central AC consisting of experienced physicians. We calculated sensitivities and specificities to adjust a published treatment meta-analysis for outcome misclassification. We delineated the consequences of AC deployment for sample size requirements in future trials.

Results

The AC identified 648 exacerbations (241 patients [58.9%] with ≥ 1 exacerbation; incidence rate 0.63 per patient-year). Sensitivity and specificity of patient self-reports were 84% and 76%, those of single expert adjudication between 89-96% and 87-99%. The pooled odds ratio in a meta-analysis of LABA versus placebo changed from 0.81 (95% CI 0.75-0.88) to 0.65 (0.52-0.81) after correction for misclassification. Avoiding exacerbation misclassification led up to 4-fold reductions in required sample sizes depending on the risk for exacerbations.

Conclusions

Patient-reports of exacerbations are only moderately accurate. This leads to underestimation of treatment effects. The use of several data sources with event adjudication will substantially reduce sample size requirements and possibly cost of studies.
Introduction

The accurate ascertainment of clinical endpoints is crucial to minimise endpoint misclassification when estimating effects in observational studies and randomised controlled trials (RCT). Because of concerns about misclassification of outcomes, endpoint adjudication committees have become common, particularly in large cardiovascular RCTs \(^1\)-\(^3\). The value of endpoint adjudication committees has been debated recently because they require considerable efforts and because some studies found no discrepancy between the effect estimates based on centrally adjudicated endpoints or adjudications made by single investigators \(^4\)-\(^6\). There is little evidence on when outcome adjudication committees are warranted in studies with patients with chronic obstructive pulmonary disease (COPD). Candidate outcomes for adjudication by a central committee in COPD studies include cause-specific mortality and exacerbations. The TORCH and UPLIFT trials \(^5\)-\(^9\) had adjudication committees for cause-specific mortality and found substantial disagreement between local investigators and adjudication committees. COPD exacerbations have rarely been adjudicated centrally by blinded experts, \(^10\)-\(^11\) despite their importance as outcome measure \(^12\)-\(^15\). The assessment of exacerbations is challenging because of mimicking differential diagnoses (e.g. worsening of heart failure or pulmonary embolism), since the cause of the exacerbations often cannot be determined and because several sources of information (patient self-reports, patient charts, emergency healthcare visits) are needed to avoid missing or misclassifying exacerbations.

Exacerbations are usually ascertained through patient self-reports or single physicians who review the available information about a patient and make a judgement about the occurrence of an exacerbation (expert adjudication). However, little is known about the accuracy of these reports. Non-differential misclassification of exacerbations, i.e. misclassification that is independent of treatment assignment or some other exposure generally leads to underestimation of treatment effects \(^16\). In addition, the required sample sizes are substantially larger if observed treatment effects are smaller than the true effects (i.e. without misclassification). Our aim was to evaluate the accuracy of the ascertainment of COPD exacerbations through patient self-reports and single physician adjudication against a reference standard of a central adjudication committee and to assess the implications of endpoint misclassification on treatment effect estimates and sample size requirements.

Methods

Study design and population

This study was nested within the prospective ICE COLD ERIC cohort study and comprised a time period of 3 years; from 1 year before baseline assessment until
2 years after baseline assessment. ICE COLD ERIC is an international multi-site prospective cohort study where 409 primary care patients (≥ 40 years of age) from the Netherlands (n=258) and Switzerland (n=151) with COPD as determined by a ratio of postbronchodilator FEV₁/FVC<0.7 and a postbronchodilator FEV₁ below 80% predicted were included. At inclusion all patients had been free of exacerbations for ≥ 4 weeks. Exclusion criteria were a life expectancy of ≤ 12 months, dementia, psychosis or other psychiatric morbidity that might have invalidated assessment of patient-reported measures. Patients were enrolled between April 2008 and August 2009. All included patients have provided written informed consent. The study has been approved of by all local ethics committees and is registered on www.ClinicalTrials.gov (NCT00706602). Detailed information on the study design 17 and results 18 19 were published elsewhere.

**Definition of exacerbations**

We used an event-based definition of exacerbations that required 1) an unscheduled physician contact in a hospital, in private practice or by telephone for worsening of dyspnoea, cough, increased sputum production and/or a change in sputum colour AND 2) an electronic or hand-written documentation in the patient record of a new prescription or a dosage increase of systemic steroids and/or new prescription of an antibiotic 17 20.

**Ascertainment of exacerbations**

We used two different methods to ascertain exacerbations: 1) through patient self-reports and 2) through detailed review of patient charts and study case report forms for exacerbations by experienced physicians. We paid much attention to have identical protocols for adjudicating exacerbations in Switzerland and the Netherlands.

Experienced study nurses contacted patients to ask about exacerbations (in patient friendly language) every 6 months by telephone interview and, at baseline and 2 years follow-ups, by face-to-face interviews. The study nurses avoided the term “exacerbation” because patients often do not understand the term but carefully asked for unscheduled contacts with physicians or hospitalisations due to worsening of symptoms (shortness of breath, sputum and/or cough) and for new prescription or dosage increases of systemic steroids and/or new prescription of antibiotics.

Seven experts, experienced pulmonologists and general practitioners with experience in the pulmonary field, independently reviewed the patient charts from the general practitioners and the case report forms for exacerbations. The experts, each on his or her own, decided about the occurrence of exacerbations, date of onset and treatment setting (at home, in hospital). The onset of an exacerbation was determined as the date of prescription or dosage increase of the drug and, for patients with self-medication, the date when the first pill was taken. To distinguish between new exacerbations and slow-to-resolve ones or relapses from
previous exacerbations, an interval of at least one month was required between the start of subsequent exacerbations.

**Adjudication committee**
The aim of the adjudication committee meetings was to reach consensus by the experts on final classifications of the patients’ exacerbations. The meetings were organised by study staff who did, however, not participate in the discussion nor had any influence on the decisions made. Based on the individual expert decisions (see above) and re-review of patient charts and case report forms if necessary, the expert committee discussed their individual decisions for those instances where individual decisions were discrepant and reached a consensus on exacerbations for each patient. The adjudication process was piloted in March 2011. Three adjudication committee meetings were held, one in Switzerland (3 Swiss experts, September 2011) and two in the Netherlands (4 Dutch experts, November and December 2011).

**Statistical analysis and re-estimation of clinical treatment effects**
We tabulated the total number of patients’ self-reported exacerbations per patient against the total number of adjudicated exacerbations per patient (defined as reference standard) to illustrate agreement. We determined the percentage of over-reporting and under-reporting of exacerbations by patients. As measures of accuracy, we calculated sensitivity (proportion of patients with ≥ 1 exacerbations correctly identified by the test method, i.e. true positives) and specificity (proportion of patients with 0 exacerbation correctly identified by the test method, i.e. true negatives) of patient self-reports compared to centrally adjudicated exacerbations. We further calculated sensitivity and specificity of single expert judgements compared to centrally adjudicated exacerbations.

To demonstrate the implication of exacerbation misclassification on effect estimates, we repeated a recent meta-analysis on long-acting beta-agonists (LABA) as first-line maintenance therapy vs. placebo for patients with moderate or more severe COPD 21 without and with correction for non-differential misclassification (i.e. we assumed that misclassification was the same for the patients with LABA or placebo). We used the logitem command in Stata (that considers the sensitivity and specificity of measuring the outcome, here exacerbations). We considered the treatment effect estimates (Odds Ratio [OR]) from that meta-analysis of RCTs that used an event-based definition for moderate or severe exacerbations according to the ATS/ERS Task Force 20. Finally, we estimated sample size requirements for randomised trials that use patient self-reports (OR for uncorrected treatment effect estimate) or adjudication by a consensus committee (OR for corrected treatment effect estimate) to ascertain exacerbations (sampsi command in Stata) 22. For all sample size calculations, we assumed a power of 80% and a significance level of 5% (two-sided). We conducted all analyses using Stata for Windows (version 12.0, College Station, Tx).
Table 7.1 Total number of exacerbations per patient: patient self-reports compared to adjudicated exacerbations by an adjudication committee (reference standard)

<table>
<thead>
<tr>
<th>Total number of centrally adjudicated exacerbations per patient</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>127*</td>
<td>24</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>165</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>40</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>77</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>17</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

*Dark grey shaded cells: number of correct self-reported exacerbations. E.g. see “Total number of exacerbations per patient, adjudicated in AC”, column “1 exacerbation”: 40 patients reported correctly that they had 1 exacerbation, 24 patients under-reported their exacerbations (they reported no exacerbation while they had 1 according to the adjudication committee), 37 patients over-reported their exacerbations (reported they had >1 exacerbations while they had 1 according to the adjudication committee)
Table 7.2 Sensitivities and specificities of patient self-reports and single experts’ judgements compared to centrally adjudicated exacerbations by an adjudication committee (reference standard) (n=409)

<table>
<thead>
<tr>
<th>Source*</th>
<th>Number of patients</th>
<th>Sensitivity (%)†</th>
<th>Specificity (%)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient self-reports</td>
<td>409</td>
<td>84.2</td>
<td>75.6</td>
</tr>
<tr>
<td>Expert 1 CH</td>
<td>151</td>
<td>88.8</td>
<td>94.4</td>
</tr>
<tr>
<td>Expert 2 CH</td>
<td>151</td>
<td>93.8</td>
<td>97.2</td>
</tr>
<tr>
<td>Expert 3 CH</td>
<td>151</td>
<td>91.3</td>
<td>98.6</td>
</tr>
<tr>
<td>Expert 1 NL</td>
<td>258</td>
<td>94.4</td>
<td>96.9</td>
</tr>
<tr>
<td>Expert 2 NL</td>
<td>257</td>
<td>95.0</td>
<td>86.6</td>
</tr>
<tr>
<td>Expert 3 NL</td>
<td>258</td>
<td>96.3</td>
<td>94.8</td>
</tr>
<tr>
<td>Expert 4 NL</td>
<td>258</td>
<td>90.7</td>
<td>96.9</td>
</tr>
</tbody>
</table>

*CH= Switzerland; NL= The Netherlands
†Exacerbations categorised: no exacerbation vs. ≥ 1 exacerbation

Results

Patient characteristics and exacerbations

57.5% (233/409) of the patients were male. Their mean age was 67.3 (SD 10.0) years, FEV₁ in % predicted 55.4 (SD 16.6) and MRC score 1.9 (SD 1.5). 261 (63.8%) patients were classified in GOLD stage II, 89 (21.8%) in GOLD stage III and 59 (14.4%) in GOLD stage IV.

In total, for the 409 patients, over 3 years, the adjudication committee identified 648 exacerbations; 241/409 patients (58.9%) suffered from at least 1 exacerbation, 168/409 patients (41.1%) had no exacerbation. The length of follow-up was 2.78 years (SD=0.57), total person-years of follow-up was 1135.7 (SE=11.5) and the incidence rate was 0.63 exacerbations/patient-year (SD=0.91). 94% (n=606) of the exacerbations were treated at home, 6% (n=38) in a hospital setting and in 4 exacerbations the setting was unclear.
Agreement between patient self-reports and single expert judgements with adjudicated exacerbations by an adjudication committee (reference standard).

Table 7.1 shows the total number of exacerbations per patient as assessed by patient self-reports compared to the total number of centrally adjudicated exacerbations. 48% (196/409) of the patients correctly reported on the total number of exacerbations; the majority of them (127/196; 65%) had had no exacerbations during the studied period. More patients over-reported exacerbations than under-reported exacerbations (139/409, 34% vs. 74/409, 18%) (Table 7.1). Thus there was more false positive than false negative misclassification.

Using the centrally adjudicated exacerbations as a reference standard, patient self-reports of exacerbations achieved a sensitivity of 84% and a specificity of 76%. Similarly, exacerbation adjudication by single experts (7 individual experts) had sensitivities ranging from 89 to 96% and specificities ranging from 87 to 99% (Table 7.2).

Meta-analysis of randomised trials on LABA vs. placebo without and with correction for misclassification of exacerbations

In 9 of 16 RCTs (56%) included in the meta-analysis of Wang patient self-reports of exacerbations were explicitly reported, while the method of exacerbation ascertainment was not reported in 7 trials (44%) but assumed to be patient reported. None of the trials used an adjudication committee. Figure 7.1 shows the meta-analyses of trials assessing the effects of LABAs vs. placebo on exacerbations. When we corrected for misclassification of exacerbations by considering a sensitivity of 84% and a specificity of 76% of patient self-reports, the pooled OR changed from 0.81 (95% CI 0.75-0.88) to 0.65 (0.52-0.81).

Sample size requirements for randomised trials using patient self-reports and consensus committee adjudication for exacerbations

We performed sample size calculations for three hypothetical RCTs that would enrol patients at low (arbitrarily set 10% for illustration), moderate (50%) or high (50%) risk for exacerbations over the course of one year. Figure 7.2 shows the sample size requirements for trials without an adjudication committee comparable to those included in the recent meta-analysis. For COPD populations at low, moderate and high risk the required number of patients to detect an OR of 0.81 with 80% power would range from 1458 to 4399 per trial arm. Instead, if misclassification were minimised through an adjudication committee, a true OR of 0.65 could be expected and sample size requirements would range from 363 to 1189 per trial arm. Thus minimising misclassification would reduce sample size requirements by around 4-fold depending on the particular risk for exacerbations in the study population at issue.
Figure 7.1 Re-estimation of meta-analysis of trials assessing the effectiveness of long-acting bronchodilators vs. placebo with exacerbations as the outcome Corrected for misclassification by considering sensitivity 84% and specificity of 76% of patient self-reports of exacerbation; same misclassification assumed for patients with LABA or placebo

Figure 7.2 Difference in sample size requirements with and without an adjudication committee for COPD exacerbations, the example of LABA vs. placebo
Discussion

Our study showed that COPD exacerbations as measured by patient self-reports are only moderately accurate. The assessments of comprehensive patient information from several data sources by single experts were more accurate than patient self-reports. The recalculation of a recent meta-analysis on LABA vs. placebo showed that effect estimates are substantially larger when corrected for misclassification of patient-reported exacerbations. The use of adjudication committees may greatly reduce sample size requirements for RCTs. Patients tended to over-report on exacerbations. During the adjudication committee meetings, we learned that a major reason for over-reporting was that patients remembered physician encounters or hospital visits but that these were often unrelated to COPD exacerbations. For example, patients received antibiotics or were admitted to hospitals because of urinary tract infection or injuries such as a broken leg. These events were clearly documented in the patient charts of general practitioners, who usually receive reports from hospitals or from other physicians if treatment did not occur at the general practitioner. The use of such data sources highlights the advantage of going beyond single sources like patient self-reports for outcomes that are difficult to ascertain. Some misclassifications concerned differential diagnoses like pulmonary embolism and some over-reporting occurred because patients referred to new exacerbations whereas the adjudication committee did not classify them as new but as slow recoveries from previous exacerbations.

Little is known about the accuracy of patient self-reports and single physician judgements, the two most common methods to ascertain COPD exacerbations. Aaron et al. (2008) \(^{10}\) compared exacerbations assessed by a patient symptom questionnaire in the Canadian Optimal Therapy of COPD Trial \(^{23}\) against a blinded adjudication committee who was provided with relevant patient records. That study also showed over-reporting of patient-reported exacerbations, which would have been missed without adjudication. A recently published study regarding physical activity and risk of moderate and severe exacerbations and COPD-related hospitalisations used medical records and two blinded investigators to verify patient reports on exacerbations but did not report on agreement of expert adjudication with patient self-reports \(^{11}\).

Investigators of the Women’s Health Initiative trial looked at the sensitivity to ascertain cardiovascular events. They found that the sensitivity of patient self-reports, compared against single expert assessment, was highly dependent on the type of outcome. The sensitivities of self-reported events like angina, peripheral vascular disease and congestive heart failure (38-49%) were substantially lower than for hard endpoints like coronary bypass surgery or angioplasty (84-90%). The accuracy of single expert judgements compared to central adjudication was higher, with sensitivities and specificities similar to those observed in our study \(^{7}\).
The current evidence suggests that the value of adjudication committees depends on the type of outcome. COPD exacerbations, episodes of angina (coronary heart disease) and other symptom-based events are rather “soft” outcomes that are difficult to define and to measure. Therefore, they are more prone to misclassification than more clearly defined events like surgery. Given the importance of symptom-based events as outcomes in RCTs, our study suggests that it may be well worth while to employ a central adjudication for "soft" events or at least an assessment by single experts. A prerequisite for adjudication is, however, that information about potential events is available from several data sources like patient reports, patient records of hospitals or GP practices and study-specific case report forms.

Our study also shows that adjudication may pay off substantially in terms of feasibility and cost of RCTs and observational studies. Non-differential misclassification of an outcome generally leads to underestimation of the true treatment effect 24. While this is well appreciated in observational research, there seems to be less awareness of this phenomenon in clinical trials. It is important to recognise that inaccurate measurement of outcomes does not just reduce the accuracy of effect estimates, but, more importantly, the validity of effect estimates 16. Our study illustrates that effect estimates corrected for misclassification can be substantially larger. As a consequence, adjudication may result in a reduction of sample size requirement and cost. If central adjudication committees for exacerbations are impossible or infeasible to organize, adjudication by single expert may be an attractive alternative because they are likely to be more accurate than patient self-reports. As Figure 7.2 shows, the reduction in sample size requirements depends much on the baseline risk for the outcome. Future research should explore, for example using simulation, how to design and conduct RCTs and observational studies in the most efficient way depending on the type of outcome, type of event ascertainment and adjudication, baseline risk of the population, the cost per patient enrolled and the cost for ascertaining and adjudicating events. As a result, investigators would be informed when and how to organise event adjudication.

Strengths of the study include that our population represents a large and diverse group of COPD patients recruited from primary care. Furthermore, we used a clearly defined event based definition of exacerbations in the ICE COLD ERIC cohort study. Exacerbations were assessed very carefully by experienced and well trained study nurses who used patients’ language and who had regularly contact with the patients during the study. Also the adjudication committees were carefully conducted in a standardised way by experienced physicians to ensure that events met pre-stated study definitions.

A limitation of our study is that even though we put great effort in the accurate assessment and adjudication of exacerbations, we still may have missed or mis-
classified some exacerbations. No standardised definition and measurement methods of exacerbations exist and different definitions may lead to different results. We used an event-based definition that required not only documented worsening of symptoms but also a documented dosage increase or new prescription of systemic corticosteroids and/or antibiotics. We therefore may have missed mild exacerbations (as defined by an ATS/ERS Task Force) which involve an increase in respiratory symptoms that can be controlled by an increase of the usual medication. Also, our time criterion of at least one month between single events to distinguish between separate exacerbations and relapses or continuation of previously recorded exacerbations is arbitrary to some extent. We used this time interval because recent data using the EXACT instrument showed that it often takes patients weeks to recover from an exacerbation. Finally, the sensitivities and specificities we provide for single experts may be somewhat inflated due to the fact that these experts were members of the adjudication committee, thereby violating the strict criterion that index test and reference standard have to be independent.

In conclusion, patient-reports of exacerbations are only moderately accurate and are likely to lead to underestimation of treatment effects. The use of several data sources with event adjudication by a committee or single experts could reduce sample size requirements and cost of RCTs and observational studies substantially.
References


