Asthma control measurement using five different questionnaires: A prospective study

Francois Vermeulena,*, Isabelle de Meulder a, Marianne Paesmans b, Inge Muyllea, Marie Bruyneela, Vincent Ninanea

a Chest Service, CHU Saint-Pierre, Rue Haute 322, 1000 Brussels, Belgium
b Data Centre, Institut Jules Bordet, Boulevard de Waterloo 121, 1000 Brussels, Belgium

Received 25 March 2013; accepted 3 July 2013
Available online 26 July 2013

KEYWORDS
Asthma control; Questionnaires; Agreement; Global initiative for asthma

Summary
Questionnaires play a key place in the assessment of asthma control. Different questionnaires have been developed. However, it remains largely unknown whether they can be used interchangeably. We wondered whether the panel of frequently used scores would give similar measurement of asthma control. The present study aimed to assess the agreement between five specific questionnaires.

Methods: In this prospective study, ninety-nine patients completed five commonly used asthma control scores: the GINA, the Asthma Control Test, the Royal College of Physician score, the Asthma Therapy Assessment Questionnaire (ATAQ), and the Asthma Control Questionnaire (ACQ). The kappa coefficient was used to assess the agreement between questionnaires.

Results: The agreement between the GINA and other scores was only moderate (kappa coefficients amounted from 0.41 to 0.60). With respect to the "controlled" level, all the other scores gave higher results than GINA. All other scores also tended to underestimate GINA "uncontrolled level". For the "partly controlled level" defined by 3 of the 5 questionnaires, ACQ identified the same percentage of patients than GINA while ATAQ overestimated this percentage.

Conclusion: This study shows only moderate agreement between five commonly used asthma control scores. The GINA score showed the lowest percentage of controlled and the highest
percentage of uncontrolled asthma. As a consequence, all these scores do not seem to evaluate the same symptoms.

Trial Registration number: NCT01350661.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Asthma control was defined as the extent to which the various manifestations of asthma have been reduced or removed by treatment [1]. This evaluation has taken a key place in the management of asthma. Indeed, recommendations emphasize the importance of asthma control evaluation to guide the adaptation of drug therapy [2] and this is mainly performed using various composites scores. Whether these scores are categorical or continuous, predicted values have been proposed to define two or three levels of control. Although several scores have been published [2–10], they differ by the factors that were considered and the weight that was given to each item [1]. They are mainly used according to local preference for the everyday management of asthma but are also the pillar of asthma assessment in many studies on asthma treatment or monitoring modalities.

A limited number of studies have assessed the level of agreement of the different available scores [11–14] and suggest that these scores are not interchangeable. These studies however have limitations related to the fact that they generally compare only two different scores between them, most often taking the GINA score as the comparator and to the fact that they were performed retrospectively using data from prospective studies that included strictly selected population. In addition, only two studies [14,15] have used the kappa coefficient to accurately assess the interobserver agreement between these questionnaires.

We wondered whether the panel of frequently used scores would give similar measurement of asthma control in a prospective study with asthma patients. With this in mind, we have assessed control of asthma in all consecutive patients referred to our outpatient asthma clinic during a one-year period using five frequently used questionnaires, namely the GINA score [2], the Asthma Control Test™ (ACT) [3], the Asthma Control Questionnaire® (ACQ) [4], the Asthma Therapy Assessment Questionnaire (ATAQ) [5] and, the Royal College of Physician score (RCP) [6].

Methods and material

Patients

This prospective observational study was performed in the outpatient’s clinic of asthma from the University hospital Saint-Pierre, Brussels, Belgium. During a one-year period (June 2010–June 2011), the principal investigator proposed to each consecutive asthma patient coming for a routine visit to participate. The only prerequisite was a diagnosis of asthma and an age older than 18 and younger than 70 years. The diagnosis of asthma included a clinical history of asthma for at least 3 months. Patients also had to demonstrate a FEV₁ reversibility exceeding 12% of the absolute value and 200 mL in response to inhalation of a short-acting beta 2-agonist.

Methods

After having signed an informed consent, each patient was asked to complete successively in a random order, the following five asthma control scores: GINA score [2], ACT [3], ACQ [4], ATAQ [5], and the RCP [6] (Fig. 1). For each patient, a randomization table created in MS Excel (Microsoft, Redmond, USA) determined the order of the questionnaires. The principal investigator could give clarifications to the patient during questionnaire completion. The GINA questionnaire includes a question about the occurrence of exacerbation in the last year. The principal investigator questioned the patient on exacerbation...
occurrence during the last year and, if possible, confirmed timing by analysis of medical records.

On the same day, a spirometry was performed using the Jaeger Masterscreen Pneumo (CareFusion Corp., San Diego, USA), as well as measurement of fractional exhaled nitric oxide (FeNO), using the Niox Flex (Aerocrine AB, Solna, Sweden). Patients were blinded with respect to the results of spirometry and exhaled NO until the total completion of scores. The study has been approved by the Institutional Ethics Committee of CHU Saint-Pierre.

**Questionnaires**

The main characteristics of the five questionnaires are summarized in Table 1.

The GINA questionnaire includes six questions about symptoms, the forced expiratory volume in one second (FEV1), and the occurrence of exacerbation. If an exacerbation occurred in the last week, the patient is considered to be uncontrolled, whatever the other results. According to GINA guidelines, an exacerbation is defined as an acute and severe loss of control that requires urgent treatment. This questionnaire is proposed by experts and has not been any validated.

The ACT includes five questions related to symptoms, reliever use, and self-rating of the level of asthma control. To answer, a Likert-scale from 1 to 5 is proposed. Results of each question are summed and validation has been reached according to the following: a score $\leq 19$ was defined as characterizing a not controlled asthma. This value was determined because it corresponds to the best sensitivity and specificity (respectively 71.3% and 78.2%) when comparing overall agreement between ACT and specialist’s rating of patient’s level of control. Validation of reliability was performed using the Cronbach’s alpha coefficients calculated for the entire sample of patients, for the group of "not controlled" patients and for the group of "controlled" patients with good results (respectively 0.84, 0.83, and 0.79) [3].

The RCP score is the result of a consensus (without further validation) issued from a multidisciplinary seminar organized by the Royal College of Physicians (UK) in 1998 [6]. It includes three questions related to nocturnal symptoms, diurnal symptoms, and limitation of activity. For each item, two answers are possible: yes or no. Each "yes" answer counts as 1 point. A score $\geq 1$ means that the asthma is not controlled.

The ACQ and encompasses questions about symptoms, use of reliever, and lung function [4]. Results of the seven questions are averaged to obtain the final score. In order to determine the level of control, two cut-off values have been described [15]. A score $\leq 0.75$ means that asthma is controlled, whereas a score $\geq 1.5$ means that asthma is uncontrolled. A minimal clinical important difference value of 0.5 was also determined [16]. The process of validation, applied to 50 patients followed during 9 weeks, has shown that the ACQ has good longitudinal validity (intraclass correlation coefficient in the stable group about 0.90) and responsiveness to change in asthma control ($p < 0.0001$) [4].

The ATAQ is a four questions score that assesses limitation of activity, nocturnal awakening, self-assessment of asthma control level, and reliever use [5]. Possible answers include "yes", "no" or "uncertain". It should be noted that the answer "uncertain" counts for one point in favour of the level "not controlled". In addition, if the patient takes less than 5 "reliever" puffs/day, the value is 0 in favour of "controlled" asthma. No validation process was applied.

**Analysis**

Three possible levels of asthma control were considered, namely "controlled", "partly controlled" or "not controlled". The agreement was assessed using the kappa coefficient [17].

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Characteristics of 5 asthma control scores. (NA: not available; ○: Quantitative items; ●: Qualitative items).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recall time, week(s)</strong></td>
<td>GINA</td>
</tr>
<tr>
<td><strong>n items</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>n levels of control</strong></td>
<td>6</td>
</tr>
<tr>
<td><strong>Minimal important difference</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>Subjective items</strong></td>
<td>NA</td>
</tr>
<tr>
<td><strong>Daytime symptoms</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Shortness of breath</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Wheeze</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Usual symptoms</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Nighttime symptoms</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Symptoms at wake up in the morning</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Nighttime symptoms and at wake up in the morning</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Limitation of activities</strong></td>
<td>●</td>
</tr>
<tr>
<td><strong>Self-assessment</strong></td>
<td>●</td>
</tr>
<tr>
<td><strong>Semi-objective items</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Reliever use</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Objective items</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Lung function</strong></td>
<td>○</td>
</tr>
<tr>
<td><strong>Exacerbation</strong></td>
<td>○</td>
</tr>
</tbody>
</table>
calculated using Statistical Package for Social Sciences 19 (IBM; Chicago, USA). As all scores do not have the same number of asthma control levels (two or three), the partly controlled and uncontrolled levels were combined when calculating the Kappa coefficient. As kappa coefficient only allows comparing two scores between them, the GINA score was initially compared to the four other scores and the remaining side-by-side comparisons were made thereafter. All kappa values were characterized according to the nomenclature proposed by Landis and Koch[18]. They suggested arbitrary words to describe the importance of agreement between observers (<0.00 poor; 0.00–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1.00 almost perfect). The Kruskal-Wallis test was used to compare the scores of each level of control respectively. A p value < 0.05 was considered as significant.

As no data allowed us to generate hypotheses about score distributions and/or differences between scores, this study was considered as exploratory and analysis was mainly descriptive. Therefore, we did not carry out a priori sample size estimation but we chose to include a maximum number of patients in a reasonably short time period (one year accrual).

Results

During the 12 months period, 105 patients were invited to participate in the study. Ninety-nine were included. Six patients refused to participate because they reported lack of time to fill out questionnaires. The characteristics of included patients, classified according to the GINA control levels, are presented in Table 2. Uncontrolled asthma was not associated with differences in anthropometric characteristics and frequency of comorbidities but with significantly lower FEV1 and a tendency toward increased doses of inhaled corticosteroids. No difference in FeNO measurements was observed.

The agreement between the GINA score, on one side and the ATAQ, the ACT, RCP and the ACQ, on the other side was moderate with kappa coefficient observed values that amounted to 0.49, 0.41, 0.52 and 0.60 respectively (Table 3).

### Table 2

Characteristics of included patients according to GINA score of asthma control. ICS dosages are expressed in beclomethasone dipropionate equivalent. (FEV1: Forced expiratory volume in one second; FeNO: fraction of exhaled nitric oxide; BMI: body mass index; ICS: inhaled corticosteroids; GERD: gastro-oesophageal reflux disease; SABA: short-acting β2-agonist; LABA: long-acting β2-agonist; LTRA: leukotriene receptor antagonist; OCS: oral corticosteroids; IgE: immunoglobulin E).

<table>
<thead>
<tr>
<th>GINA classification</th>
<th>n</th>
<th>Partly controlled</th>
<th>Uncontrolled</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Controlled</strong></td>
<td>28</td>
<td>21</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Sex F/1M</td>
<td>3.0</td>
<td>1.6</td>
<td>1.6</td>
<td>0.614</td>
</tr>
<tr>
<td>Age yrs</td>
<td>43.8 ± 14.5</td>
<td>39.3 ± 19.0</td>
<td>43.1 ± 17.2</td>
<td>0.501</td>
</tr>
<tr>
<td>Body mass index kg/m²</td>
<td>26.9 ± 6.1</td>
<td>27.0 ± 7.7</td>
<td>26.9 ± 6.0</td>
<td>0.945</td>
</tr>
<tr>
<td>Time from diagnosis yrs</td>
<td>9.85 ± 16.4</td>
<td>6.3 ± 6.3</td>
<td>11.7 ± 14.0</td>
<td>0.101</td>
</tr>
<tr>
<td>FEV1 L</td>
<td>2.88 ± 0.74</td>
<td>3.13 ± 0.68</td>
<td>2.47 ± 0.87</td>
<td>0.002</td>
</tr>
<tr>
<td>FEV1% of predicted value</td>
<td>97 ± 12</td>
<td>100 ± 12</td>
<td>80 ± 19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FeNO ppb</td>
<td>34.1 ± 33.3</td>
<td>31.6 ± 24.8</td>
<td>38.9 ± 33.4</td>
<td>0.566</td>
</tr>
</tbody>
</table>

#### Comorbidities

- Smoking: 2 (7.1%) 2 (9.5%) 10 (20.0%) 0.237
- GERD: 9 (32.1%) 11 (52.3%) 17 (34.0%) 0.277
- Chronic rhinosinusitis: 15 (53.5%) 11 (52.3%) 26 (52.0%) 0.991
- Sleep apnoea syndrome: 2 (7.1%) 1 (4.7%) 2 (4.0%) 0.831
- Obesity (BMI > 30): 6 (21.4%) 4 (19.0%) 16 (32.0%) 0.420
- Allergies: 22 (78.5%) 17 (80.9%) 35 (70.0%) 0.541

#### ICS dose prescribed

- No ICS: 4 (14.2%) 4 (19.0%) 7 (14.0%) 0.855
- < 400 mcg: 10 (35.7%) 3 (14.3%) 8 (16.0%) 0.179
- 400 – 799 mcg: 7 (25.0%) 6 (28.6%) 9 (18.0%) 0.464
- > 799 mcg: 7 (25%) 8 (38.1%) 26 (52%) 0.065

#### Medical treatment

- No treatment: 2 (7.1%) 1 (4.8%) 3 (6%) 0.737
- SABA: 18 (64.3%) 17 (81%) 38 (76%) 0.209
- LABA: 0 (0.0%) 1 (4.8%) 1 (2.0%) 0.252
- ICS: 7 (25.0%) 6 (28.6%) 13 (26.0%) 0.785
- ICS + LABA: 17 (60.7%) 12 (57.1%) 37 (74%) 0.806
- Theophyline: 0 (0.0%) 1 (4.8%) 0 (0.0%) NA
- LTRA: 7 (25%) 7 (33.3%) 13 (26%) 0.533
- OCS: 0 (0.0%) 0 (0.0%) 5 (10.0%) NA
- Anti-IgE: 1 (3.6%) 0 (0.0%) 1 (2.0%) 0.678
The ACQ and ATAQ had the worse agreement with a kappa value of 0.37. On the other hand, ACQ and respectively RCP and ACT had only a moderate agreement (respective kappa values of 0.57 and 0.58). ACT and RCP as well as ATAQ and RCP had also a moderate agreement (respective kappa value of 0.53 and 0.58). The best agreement was observed between ACT and ATAQ with a kappa value of 0.62.

Fig. 1 shows a graphical view of the comparison between the GINA score and the ACQ, ACT, RCP and ATAQ results. With respect to the "controlled" level, all scores gave higher results than GINA (ACQ +12.1%, ATAQ +16.2%, ACT +21.2%, RCP +4.0%). For the "partly controlled" level "ACQ identified the same percentage of patients than GINA while ATAQ overestimated this percentage (+15.2%). Concerning the "uncontrolled" level of control, all scores underestimated the GINA result (ACQ –12.1%, ATAQ –31.3%, ACT –21.2%, RCP –4.0%). Fig. 2 shows a similar graphical view of the different side-by-side comparisons between the ACQ, ACT, RCP and ATAQ and comparisons of results showed significant differences between ACQ and RCP (p < 0.001), ACQ and ATAQ (p = 0.002), ACT and ATAQ (p = 0.025), and ACQ and ACT (p = 0.049). No significant difference was observed between ATAQ and RCP (p = 0.067) or between ACQ and RCP (p = 0.074).

Comparisons of the percentage of patients in different levels of control (Table 4) also showed a statistically significant difference between GINA and respectively ACQ, ATAQ and ACT scores (p < 0.001) but not with the RCP score (p = 0.374). Differences were, however, often important. As an example, for the level "controlled", this percentage ranged from 4.0 to 21.2%, for the level "partly controlled" from 0.0 to 15.2%, and for the level "uncontrolled" from 4.0 to 31.3%.

### Discussion

In the present study, the level of agreement between five commonly used scores of asthma control was assessed prospectively in a population of ambulatory asthma patients. The results show an only moderate overall agreement between them. In addition, for each given level of asthma control, the proportion of patients varies according to the scores. All together, these results suggest that these questionnaires are not interchangeable.

Previous studies have already compared questionnaires between them. Koolen et al. [11] studied the ability of ACT to identify patients with uncontrolled asthma in comparison with GINA score. They found that ACT with a cut-off value 20 shows a sensitivity of 66% and a specificity of 100% and then suggested to increase the ACT cut-off value in order to increase sensitivity. Alvarez-Gutierrez et al. [12] also compared ACT and GINA scores and noticed that the ACT, for each GINA levels of control, classified only 59.4, 54.1 and 63.6% of the patients correctly. It is noteworthy that they proposed an additional, second cut-off for the ACT, such that it would include three levels of control. Finally, Thomas et al. [13] assessed the relationship between RCP and ACQ, and its course over 12 weeks. Despite the relatively small sample and a single cut-off to 1.0 for the ACQ, they observed a strong relationship (0.79, p < 0.001) between these scores. However, even though the RCP identified 94% of poor controlled episodes, the authors observed that 35% of the cases were considered as "controlled" using the ACQ.

![Figure 2](image_url)  Percentage of patients identified by each score for each level of control. GINA (solid line) is considered as the reference score. For ACT and RCP, with only two levels of control, the GINA 'partly controlled' level was included in the 'uncontrolled one' and expressed as GINA2 in the figure.
Consequently, they suggested that RCP may over-estimate inadequate control. Altogether, these studies already suggested that these questionnaires are not interchangeable. Their limitations, however, are related to the fact that they were most often retrospective, comparing a limited number of questionnaires, in a selected population of asthma patients and often without kappa coefficient analysis.

Indeed, the classical statistical test designed to assess the interobserver agreement is the kappa coefficient [17] that gives a value between 0 and 1. A value of 1 indicates the perfect agreement whereas, on the opposite, a value of 0 means that agreement is equivalent to chance. To date, only two studies assessed the agreement between scores using this coefficient. Thomas et al. [19] retrospectively compared the ACT and GINA results in more than 2900 patients and found a kappa coefficient of 0.42. They observed that a value of $\text{ACT} \leq 19$ is a good predictor of partly controlled/uncontrolled level, according to GINA since this cut-off identified 93.9% of patients. However, this percentage only amounted to 51.3% when controlled patients were considered. More recently, O’Byrne et al. [14] made a post hoc analysis of three studies comparing three scales, GINA, ACQ-5 (a shorter validated version of the ACQ) and the Gaining Optimal Asthma Control (GOAL) index [10]. They indeed found that GINA and GOAL have a substantial agreement with a kappa of 0.80 but this was not the case between GINA and ACQ since kappa value was only 0.61 when comparing GINA uncontrolled level and ACQ-5 $\geq 1.5$.

These previous observations as well as our results showing moderate agreement between five frequently used scores can be explained by different causes. The first one is related to differences in properties and validation. Indeed, few questionnaires meet all described quality criteria to assess the measurement properties of health status questionnaires [20–22]. Interestingly, the GINA score, which could be considered as a reference, has not been formally validated. This is also the case for the RCP. As far as ATAQ is concerned, authors showed good correlations between ATAQ, quality of life, and health care utilization but no formal validation was published [5]. Only the ACQ and ACT have been fully validated [3,4]. A second potential explanation is linked to the content of scores. Most of the time when creating a score, an initial set of items submitted to a panel of experts is gradually reduced and finally validated. Experts can also directly propose items in a final version. Even if expert opinion can be considered as valid at first sight, it is interesting to note that only one item (limitation of activity) is found in each of the five questionnaires (Table 1). A third explanation is related to accuracy of questionnaires. Indeed, the number of questions included in each score and the number of possible answers for each question vary according to questionnaires. This implies that the change of a single response will influence the final outcome to a different extent in different scores (Fig. 3). For example, changing an answer in the RCP modifies the final result by an amount that is equal to about one third. On the other hand, the change of a single answer in the ACQ, changes the final result by only 2.4%. A fourth reason is related to differences in the timing of recall: GINA and ACQ require a recall time of 7 days. In contrast, for the ACT and ATAQ, the patient should remember the last four weeks. RCP is less clear because it accounts for the last week or last month. These two periods are typically used, although their choice seems to be based on relatively empirical data [1,23]. In addition, it has been shown that the severity of symptoms at the time of completing a questionnaire affects the perception that one can have from the same symptoms appeared a few days earlier [24]. A last reason is specific to the GINA questionnaire. This was the only one to assess previous exacerbations. The occurrence of an exacerbation during the last week determines an uncontrolled asthma independently of the results to other questions. One may expect that this would explain the higher incidence of uncontrolled asthma using this questionnaire. In the present study, 8 patients have had an exacerbation during the week before the inclusion. This corresponds to 16.0% of uncontrolled patients according to GINA. As a comparison, for the four other questionnaires, the number of patients exacerbated in the last week and the percentage of uncontrolled patients according to each questionnaire was 6 (12.0%) for the ACT, 6 (15.8%) for the ACQ, 5 (26.3%) for the ATAQ, and 8 (11.9%) for the RCP.

All together, these factors may partly explain the variability of agreements between questionnaires observed in the present study.

This study also shows some potential limitations. At first, the number of included patients may seem low in comparison with the large number of patients from retrospective studies [14,19]. However, these ones comparing pairs of questionnaires (GINA vs ACQ [14], GINA vs ACT [19]) observed kappa values identical to those of the present study, a finding which suggests that the sample size of our study is likely to be large enough.

Figure 3 Relative values of each level of control for each score. (White, solid line: controlled asthma; Grey: partly controlled asthma; Black: uncontrolled asthma; White, dotted line: impact in percentage of the total score of one single unitary change in answer to question).
In the present study, a reference questionnaire was required for the pertinence and comparison of Kappa analysis. No questionnaire had clear advantages over the others and the GINA questionnaire was then selected on the reason that GINA guidelines are the most frequently and globally used guidelines for the management of asthma. We believe however that the selection of another reference would not have affected the general conclusion of the present study.

Questions included in the different questionnaires are often similar and are also part of the regular questions that are asked by clinicians outside questionnaire use to assess asthma. We do not believe that this may induce any limitation and it must be stressed that none of the patients from the present study had to complete any of these structured questionnaires before and the order of questionnaire administration was randomized.

Additional scores could also have been included in our study. The GOAL index [10] was not considered as it involves filling a daily diary with symptoms and PEF values during the whole week before the assessment. Given the fact that other scores do not provide this diary, we decided to not include GOAL. Boulet et al. [9] proposed a score that included assessment of symptoms, FEV₁, and sputum eosinophilia. Despite its usefulness [25], induced sputum is currently not routinely used and this led us not to take into account this score in the present analysis.

RCP and ACT scores provide only 2 categories, namely “controlled” or “not controlled” state. In order to compare both these scores with other three scores, we decided to combine the categories “partly controlled” and “uncontrolled”. This has however the disadvantage of masking patients who are “partially controlled” such that the calculated kappa values relate only the “controlled” category. Radar charts (Fig. 1) let show the distribution of patients who are “partially controlled” such that the present data do not allow making quantitative assessment of asthma control. These results demonstrate that these scores are not interchangeable. Cautious is then required when achieving literature reviews or meta-analyses that include studies assessing the level of asthma control. In the clinical setting, the present findings also suggest that the same score should be used and repeated in a given patient in order to allow longitudinal follow-up and comparison.

COI disclosure

All authors have reported that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Acknowledgements

This study was supported by a grant from the IRIS-Recherches ASBL, Brussels, Belgium.

References


