Use of a composite symptom score during challenge in patients with suspected aspirin-exacerbated respiratory disease

Kevin A. Cook, MD *; Brian D. Modena, MD *; Nathan E. Wineinger, PhD 1; Katharine M. Woessner, MD 1; Ronald A. Simon, MD 1; Andrew A. White, MD 1

* Division of Allergy, Asthma and Immunology, Scripps Clinic, San Diego, California
1 Scripps Translational Science Institute, La Jolla, California

A R T I C L E   I N F O

Article history:
Received for publication December 30, 2016.
Received in revised form February 8, 2017.
Accepted for publication February 16, 2017.

A B S T R A C T

Background: Aspirin-exacerbated respiratory disease is characterized by asthma, chronic rhinosinusitis, nasal polyposis, and sensitivity to aspirin and other nonsteroidal anti-inflammatory drugs. Confirmation of the diagnosis requires provocation challenge with resulting upper and/or lower airways reactivity. Currently, determination of a positive challenge result is based solely on clinical judgment that synthesizes subjective symptoms and objective measures, as a concomitant increase in nasal or bronchial airways resistance is measured in only half of patients.

Objective: To describe a quantitative scoring system, based on symptoms typically reported during provocation challenge, used to identify a positive challenge result.

Methods: A total of 115 patients were asked to record 10 symptoms, rated on a scale from 1 (mild) to 10 (most severe), at regular intervals during intranasal ketorolac with modified oral aspirin challenge performed in our office. Composite scores, a simple sum of all individual scores, were calculated at each time point and compared with baseline, prechallenge values.

Results: One hundred of the 115 patients were determined to have a positive challenge result. A statistically significant difference in composite scores was observed in reactors vs nonreactors. All nonreactors recorded an increase in composite score of less than 5, whereas 69% of reactors recorded an increase of 5 or more.

Conclusion: Our 10-symptom composite score provides a quantitative and comparable measure of symptoms that typically present during a challenge with a positive result. Although an external validation is needed to confirm its diagnostic performance characteristics, a change in composite score of 5 or more appears to be specific to reactors.

© 2017 American College of Allergy, Asthma & Immunology. Published by Elsevier Inc. All rights reserved.

Introduction

Aspirin-exacerbated respiratory disease (AERD) is a condition characterized by a combination of asthma, chronic rhinosinusitis with nasal polyposis, and sensitivity to aspirin and other cyclooxygenase 1–inhibiting nonsteroidal anti-inflammatory drugs (NSAIDs). AERD occurs in 7% of patients with asthma but is estimated to affect up to 14% of patients with severe asthma.1 AERD is linked to greater severity of symptoms, more frequent exacerbations requiring systemic steroids, and an increased rate of emergency intubation compared with aspirin-tolerant patients with asthma.2–8

Currently, there is no biomarker or objective test to diagnose AERD.9 The diagnosis is made based on clinical history of chronic rhinosinusitis with nasal polyps and asthma, along with a confirmatory positive provocation challenge result to aspirin or other NSAIDs. Historically, challenges were performed using aspirin, with a reported 43% of suspected AERD challenge patients experiencing bronchial reactions.10 Bronchial reactions have been typically confirmed and quantified by spirometry, performed at regular intervals throughout the challenge procedure. Although a decrease in forced expiratory volume in 1 second (FEV1) of 20% or greater has an estimated 45% to 69% sensitivity in detecting aspirin reactors,11,12 a decrease of 15% or greater has been used as a diagnostic threshold in several more recent studies.13–15 Both threshold values have insufficient sensitivity to be used as the sole measure of reactivity and would miss many AERD reactions.

The increased use of inhaled corticosteroids and leukotrienemodifying drugs (LTMDs) before an aspirin or NSAID challenge has reduced the number and severity of bronchial reactions.10 In...
addition, intranasal ketorolac with modified oral aspirin challenge has proven to be a safe, effective alternative to using oral aspirin for provocation challenge. This protocol further reduces lower respiratory and extrapulmonary reactions. In fact, 65% of patients now experience only naso-ocular symptoms. Peak nasal inspiratory flow (pNIF) has been established as an objective measure of naso-ocular reactions. A decline in flow rate of greater than 25% is highly specific to identify a positive reaction, although the sensitivity of this measurement is less robust. Furthermore, clinical symptoms representing an upper airways reaction, such as rhinorrhea and pruritus, often occur without measurable increase in nasal resistance. Thus, the determination of a positive challenge result is still substantially dependent on patient-reported symptoms and the overall judgment of an experienced physician. We developed a composite score that consisted of 10 symptoms most commonly reported during in-office challenge, including nasal congestion, runny nose, sneezing, itchy nose, itchy or watery eyes, itchy ears, itchy tongue or mouth, throat tightness, wheezing, and chest tightness. In a retrospective analysis, we report differences in scores between aspirin reactors and nonreactors and those reactors with concomitant objective findings.

**Methods**

**Patients**

A total of 115 patients had a provocation challenge in our clinic between November 2011 and December 2015, with recording of the 10-symptom score throughout the challenge procedure. Our retrospective analysis of the results was approved by the Scripps institutional review board. All patients had similar features suggestive of AERD before challenge, including nasal congestion, runny nose, sneezing, itchy nose, itchy or watery eyes, itchy ears, itchy mouth or tongue, tight throat, wheezing, and chest tightness. In a retrospective analysis, we report differences in scores between aspirin reactors and nonreactors and those reactors with concomitant objective findings.

**Provocation Challenge**

All patients had undergone an identical protocol using intranasal ketorolac with modified oral aspirin, which has been previously described in detail. This protocol is considered diagnostic and therapeutic, with the goal of confirming aspirin sensitivity while simultaneously desensitizing the patient. On day 1, patients were challenged with increasing doses of ketorolac intranasal spray (1.26, 2.53, 5.05, and 7.58 mg) at 30-minute intervals, followed by oral aspirin (60 and 60 mg) separated by 90 minutes. On day 2, patients were challenged with oral aspirin only (150 and 325 mg) separated by 90 minutes. A pNIF meter (In Check Nasal; Clement Clarke International, Essex, England) was used to measure nasal inspiratory flow at baseline (before challenge), 30-minute intervals during intranasal ketorolac challenge, and 60- to 90-minute intervals during modified oral aspirin challenge. FEV1 was recorded at baseline, 30 minutes after completion of ketorolac challenge, at 60- and 90-minute intervals during modified oral aspirin challenge, and more often if the patient reported onset of bronchial symptoms.

**Determination of Positive Challenge Result**

A physician experienced in the diagnosis and management of AERD made the final decision on positive reactors based the new onset or worsening upper or lower respiratory tract symptoms, including conjunctivitis (tearing, pruritus, or injection), rhinitis (rhinorrhea, congestion, pruritus, or sneezing), laryngospasm, or bronchospasm (wheezing, chest tightness) and/or a decrease in FEV1 of 15% or greater in bronchial reactors. Specific threshold values for 10-symptom composite scores and pNIF measurements were not used by the physician to make the diagnosis of a positive challenge result.

**Symptoms Scoring**

Symptom scores were also recorded at baseline, at 30-minute intervals during ketorolac challenge, and at 60- and 90-minute intervals during modified oral aspirin challenge. Patients self-recorded the following symptoms on a scale of 0 to 10: nasal congestion, runny nose, sneezing, itchy nose, itchy watery eyes, itchy ears, itchy mouth or tongue, tight throat, wheezing, and chest tightness. Verbal guidelines for symptom scoring were given to each patient as follows: 0, no symptoms; 1 to 3, mild, barely noticeable; 4 to 7, moderate, somewhat troublesome; and 8 to 10, much worse than you generally experience, making you very uncomfortable. Composite scores, a simple sum of all individual scores, was calculated at each time point and compared with baseline, pretest values. A sample symptom score card is provided in Figure 1.

**Statistical Analysis**

A change in the 10-symptom composite score was calculated as the difference between baseline and the time of reaction for aspirin and NSAID reactors. For nonreactors, change in composite score was calculated as the difference between baseline and their highest recorded composite score. Similarly, changes in FEV1 and pNIF were calculated as the percentage change between baseline and the time of reaction for reactors and the difference between baseline and the lowest measurements for nonreactors.

A \( \chi^2 \) test was used to assess significance related to differences in categorical variables. Similarly, unpaired Welch 2-tailed \( t \) tests were used to assess significance related to differences in continuous variables, including changes in 10-symptom composite scores and

---

**Figure 1.** Sample symptom score flowsheet. This hypothetical patient had a baseline composite score of 12. At 9:30 his change in composite score exceeded a diagnostic threshold of 5, indicating a probable reaction.
the percent changes in pNIF and FEV1. $P < .05$ was considered statistically significant. The Spearman rank correlation coefficient was used to correlate nasal congestion score vs change in pNIF.

**Results**

**Study Population**

Medical records from 115 patients were included, and 100 patients were noted to be aspirin or NSAID reactors. Nonreactors were found to be older than the reactor group (mean age, 55 vs 47 years; $P < .05$). Otherwise, there were no statistical differences in other baseline patient characteristics (Table 1), including sex, race, atopy, or number of sinus operations. Nearly all patients (97%) were taking an LTMD at the time of their challenge. A subset of patients were noted to be aspirin or NSAID reactors. Nonreactors atopy, or number of sinus operations. Nearly all patients (97%) were

**Symptom Scores**

Composite scores were significantly higher at the time of reaction vs highest composite score of nonreactors (15.3 vs 7.8, $P = .003$). The change in composite score from baseline was also significantly higher in reactors vs nonreactors, with mean (SD) changes of 9.4 (7.9) and 1.4 (1.6), respectively ($P < .001$). When reactors were subdivided into those with objective airways obstruction (decline in FEV1 $>15\%$, n = 24) and those with only subjective symptoms of reaction (N = 76), no significant difference was noted in change in composite score, with mean values of 11.1 and 8.8, respectively ($P = .20$). Both reactor groups differed significantly from nonreactors (both $P < .001$). None of the nonreactors had a change in composite score of 5 or greater. Distribution of change in composite scores is depicted in Figure 2, showing clear differences in composite symptoms scores based on reactor status.

Of the 10 symptoms, all individual symptom scores were higher in the reactor group compared with the nonreactors, with the difference reaching statistical significance for nasal congestion, runny nose, sneezing, and tight throat (Fig 3). Interestingly, the highest symptom scores were recorded for nasal congestion in reactors and nonreactors, although the finding was still statistically greater in reactors (3.4 vs 2.3, $P = .04$). A moderate inverse correlation ($r_s = -0.511$) was noted between change in nasal congestion score and decline in pNIF, suggesting relative agreement between perceived nasal symptoms and objective measurement of nasal resistance.

Although extrapulmonary symptoms were not included in symptom scoring and were not universally recorded in challenge flowsheets, at least 5 aspirin reactors experienced a gastrointestinal reaction characterized by nausea, vomiting, and/or abdominal cramping. Two of these patients recorded little or no other symptoms of an airways reaction, with a change in composite score from baseline of less than 5. At least 2 aspirin reactors experienced a rash, both of which had significant accompanying airways symptoms.

$pNIF$ and $FEV_1$

The decrease in pNIF was significantly greater in reactors vs nonreactors at a mean (SD) of 24% (23%) and 14% (10%), respectively ($P = .006$). Similarly, the decrease in $FEV_1$ was greater in reactors vs nonreactors at a mean (SD) of 7% (11%) vs 4% (4%), respectively ($P = .049$). Twenty-four reactors experienced a decrease in $FEV_1$ of 15% or more.

Performance characteristics of established diagnostic threshold values for pNIF and $FEV_1$ were assessed. Using the physician diagnosis of a positive challenge result as the standard, a decrease in pNIF of 25% or more was found to be 42% sensitive and 80% specific for making the diagnosis. Similarly, a decrease in $FEV_1$ of 15% or more was only 24% sensitive but 100% specific. Combined use of objective diagnostic thresholds, including a decrease in pNIF of 25% or more and/or a decrease in $FEV_1$ of 15% or more yielded a 54% sensitivity and 80% specificity.

Thirty-one aspirin reactors recorded minimal change in composite scores (score change of 0–4) at the time a reaction was noted by the supervising physician. In this group, 5 patients had a decrease in $FEV_1$ greater than or equal to 15%, indicating aspirin sensitivity. An additional 7 patients had a decrease in pNIF greater than or equal to 25%, aiding the physician in identifying a reaction.

**Discussion**

The diagnosis of AERD requires suggestive historical features, with confirmation of aspirin or NSAID sensitivity by provocation challenge. Objective measurements, including pNIF and $FEV_1$, are often insufficient to identify all reactors. In fact, only 54% of our patients would have met diagnostic criteria for a positive challenge result based solely on $FEV_1$ and/or pNIF threshold values. Thus, patient-reported symptoms must be considered during provocation challenge. Intuitively, symptom scoring makes sense because we know that patients with AERD should experience symptoms of airways inflammation when challenged with aspirin or other NSAIDs. A symptom score represents a logical step toward quantifying symptoms, which provides a means to compare reactions within and between institutions. On the basis of our experience, we used 10 symptoms that patients with AERD often report during challenge to create a 10-symptom composite score.

---

**Table 1 Clinical Characteristics of the Study Population**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Reactors (n = 100)</th>
<th>Nonreactors (n = 15)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>.52</td>
</tr>
<tr>
<td>Female</td>
<td>51 (51.0)</td>
<td>9 (60.0)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49 (49.0)</td>
<td>6 (40.0)</td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>46.8 (11.7)</td>
<td>55.0 (10.5)</td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>48 (19–70)</td>
<td>56 (31–64)</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White/non-Hispanic</td>
<td>76 (76.0)</td>
<td>10 (66.7)</td>
<td></td>
</tr>
<tr>
<td>White/Hispanic</td>
<td>4 (4.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>3 (3.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2 (2.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Unknown or refused</td>
<td>15 (15.0)</td>
<td>5 (33.3)</td>
<td>.07</td>
</tr>
<tr>
<td>Atopic by skin testing or specific IgE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>63 (63.0)</td>
<td>5 (33.3)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>37 (37.0)</td>
<td>10 (66.7)</td>
<td></td>
</tr>
<tr>
<td>Unknown or not tested</td>
<td>22 (22.0)</td>
<td>7 (46.7)</td>
<td></td>
</tr>
<tr>
<td>Prior sinus operations, mean (SD)</td>
<td>2.8 (2.1)</td>
<td>2.4 (1.8)</td>
<td>.52</td>
</tr>
<tr>
<td>LTMD</td>
<td>96 (96.0)</td>
<td>15 (100)</td>
<td>.43</td>
</tr>
<tr>
<td>Systemic corticosteroids</td>
<td>22 (22.0)</td>
<td>3 (20.0)</td>
<td>.86</td>
</tr>
<tr>
<td>Omalizumab</td>
<td>17 (17.0)</td>
<td>1 (6.7)</td>
<td>.30</td>
</tr>
<tr>
<td>Baseline $FEV_1$, mean (SD), % predicted</td>
<td>88.1 (15.2)</td>
<td>80.5 (16.5)</td>
<td>.11</td>
</tr>
<tr>
<td>Baseline pNIF, mean (SD), L/min</td>
<td>163.1 (50.0)</td>
<td>140.3 (38.3)</td>
<td>.052</td>
</tr>
<tr>
<td>Baseline composite score</td>
<td>6.5 (9.0)</td>
<td>5.2 (4.7)</td>
<td>.42</td>
</tr>
<tr>
<td>Mean (range)</td>
<td>4 (0–59)</td>
<td>3 (3–15)</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** FEV1, forced expiratory volume in 1 second; LTMD, leukotriene modifier drug; NIF, nasal inspiratory flow.

*a Data are presented as number (percentage) of patients unless otherwise indicated.
A significant difference in composite score was noted between reactors and nonreactors, with reactors averaging a more than 6-fold higher increase from baseline. We find the patient-recorded, 10-symptom AERD composite score an easily interpreted supplement to objective measures of airways resistance that can be implemented without additional equipment or burden to medical staff.

Increasing use of inhaled corticosteroids, LTMDs, and challenge protocols using intranasal ketorolac have reduced the number of bronchial reactions, and most patients now experience only naso-ocular symptoms. pNIF, a measurement of nasal resistance and a marker of nasal congestion, has been established as the most widely used, objective metric of nasal reactions, although limited by overall low sensitivity. Supportively, we found that a decrease in pNIF of 25% or greater was only 42% sensitive but 80% specific for determining a positive challenge result. Nonetheless, our study reveals that subjective complaints of nasal congestion moderately correlate in an inverse manner with a decrease in pNIF ($r_s = -0.511$), validating the utility of monitoring patient-reported symptoms.

Similarly, although lower respiratory and extrapulmonary symptoms have been reported during intranasal ketorolac and modified oral aspirin challenge, the rates may be much lower compared with protocols using oral aspirin alone. This finding is consistent with the poor diagnostic sensitivity (24%) of a decrease in FEV$_1$ to identify reactors in our study population compared with prior reported performance characteristics during standard oral aspirin protocols (45%–69% sensitivity).

Furthermore, many reactions that occur during provocation challenge cannot be measured objectively, such as watery eyes, pruritus of various locations, or throat tightness. Our 10-symptom composite score may provide broader capture of varied reaction types. All 10 individual symptom scores were higher in reactors vs nonreactors, although only 5 reached statistical significance. Weighting of individual scores by multiples of 0, 1, or 2 was attempted to maximize differentiation between reactors and nonreactors.
nonreactors but provided only marginal benefit. Accordingly, equal weighting of symptoms with calculation of the composite score as a simple sum was chosen to improve ease of clinical use and ensure capture of a range of symptoms. Although requiring validation in an external cohort, our 10-symptom composite score, or a future iteration, may one day prove useful as a diagnostic metric. All nonreactors had an increase in composite score less than 5, whereas 69% of reactors scored 5 or more (Fig 2). This finding suggests that a diagnostic threshold of 5 would be moderately sensitive but highly specific to reactors. Unfortunately, a true validation of this score requires an objective variable independent of patient-reported symptoms that can be used as the gold standard of diagnosis. Because a positive challenge result is based mostly on the subjective diagnosis by a trained physician, this sort of gold standard simply does not exist yet. Notwithstanding, if experienced physicians at different sites independently find a diagnostic threshold of 5 or greater to be highly specific to aspirin reactors, this may indicate some degree of validity. Of note, 31 patients were identified as aspirin reactors despite minimal increase in their composite symptom score (scores of 0–4). Twelve (39%) of these patients had a decline in FEV1 and/or pNIF to aid in identifying their reactions. Others recorded increases in 1 or 2 symptom scores with composite scores not markedly increased. For example, a patient may experience an increase in “runny nose” from a self-recorded value of 4 to 8. Although the composite score increased by only 4, the supervising physician would be inclined to identify this as a reaction. Others recorded further increases in composite score after their reaction was identified, which would have meet a threshold value of 5. An additional 2 patients experienced gastrointestinal reactions without significant accompanying airways symptoms. Patients may also have objective evidence of reaction, such as sneezing or frequent throat clearing, despite underreporting symptoms caused by poor perception or high symptom tolerance. Thus, although the composite score may assist in identifying aspirin reactors, individual symptom scores, extrapulmonary symptoms, direct patient observation, and objective measures of airways obstruction must all be considered.

Similar to the use of pNIF and FEV1 values, we used a change in symptom scores measured from baseline rather than absolute composite scores. This strategy was used because of the observed variation in baseline symptoms in all patients. Many patients, both reactors and nonreactors, recorded significant symptoms with high composite scores at baseline, and the use of absolute threshold values would prohibit useful comparison between patients. Use of values measured as change from baseline identifies new and/or increasing symptoms.

Limitations of the study include exclusive use of the intranasal ketorolac and modified oral aspirin challenge protocol, which may hinder generalizability to other protocols. For example, protocols using oral aspirin only may lead to different patterns of symptoms, affecting composite score values. Use of premedications, including LTMDs and inhalers, may also influence the development and severity of symptoms, potentially decreasing the diagnostic value of symptom scoring. This would also be expected for other diagnostic metrics, including pNIF and FEV1. In addition, the composite symptom score may be less useful in patients who are poor symptom perceivers. This scoring tool does not include domains for cutaneous or gastrointestinal reactions, although these would be potentially useful additions in future iterations.

This study is also limited by its retrospective design and lack of a placebo control. Although a prospective controlled trial would be ideal for validation, logistics and cost of performing masked 2-day challenges limit feasibility. Alternatively, we feel that nonreactors in this study provide a reasonable group of comparison.

Nonreactors reported similar histories and baseline symptoms, suggesting comparable potential for nonspecific airways hyperresponsiveness to intranasal sprays.

Lastly, as mentioned, an objective and universally accepted method for identification of aspirin reactors simply does not exist. Urinary leukotriene E4 levels, platelet leukocyte aggregates, and other biomarkers have all been suggested but proven to be of equivocal utility and likely prohibitively expensive. Currently, the clinical interpretation and judgment of the supervising allergist is required to identify positive challenge results with some inevitable subjectivity. Nonetheless, the allergists supervising these challenges have collectively evaluated, desensitized, and longitudinally managed several hundred patients with AERD, and physician diagnosis remains the most accepted method for confirming aspirin sensitivity. Our 10-symptom composite score aims to quantify patient-reported symptoms and provides a means to compare reactions within and between institutions.

In conclusion, our 10-symptom AERD composite score provides an easily implementable, quantifiable measure of patient-reported symptoms during ketorolac with modified oral aspirin challenge. Our data suggest that a change in composite score of 5 or more may be a moderately sensitive and highly specific means to identify aspirin or NSAID reactors during provocation challenge, acknowledging multiple limitations in the study design and that further validation is needed. We propose that this scoring system can be helpful to clinicians performing aspirin challenges and desensitizations to compare severity of reaction between groups of patients with AERD in research studies and in comparative research among different institutions.

References


