Validation of Patient-Reported Outcomes for Clinical Trials in Allergic Rhinitis: A Systematic Review

Moises A. Calderon, Thomas Casale, Pascal Demoly

Validation of Patient-Reported Outcomes for Clinical Trials in Allergic Rhinitis: A Systematic Review

Moises A. Calderón MD, PhD[1], Thomas B. Casale MD[2], Pascal Demoly MD, PhD[3]

[1] Imperial College London - National Heart & Lung Institute, Royal Brompton Hospital NHS, London, United Kingdom.

[2] University of South Florida, Division of Allergy and Immunology, Tampa, FL 3361.

[3] Department of Pulmonology - Division of Allergy, Hôpital Arnaud de Villeneuve, University Hospital of Montpellier, Montpellier, France, and Sorbonne Universités, UMR-S 1136, IPLESP, Equipe EPAR, Paris, France

Corresponding author:

Professor Pascal Demoly MD, PhD. Allergy Division, Pulmonary Department, Hôpital Arnaud de Villeneuve, University Hospital of Montpellier, Montpellier, France

Tel.: +33-467-336-107

Fax: +33-467-042-708

E-mail: pascal.demoly@inserm.fr

Competing interests

Moises A. Calderón has received lecture fees from ALK, Stallergenes Greer, Merck, Allergopharma, HAL Allergy, and consultancy fees from ALK, Stallergenes Greer, Merck, HAL Allergy, and ASIT biotech.

Thomas Casale has been an investigator and consultant for Novartis and Genentech with all fees to his university employer.

Pascal Demoly has received consultancy fees from ALK, Stallergenes Greer, Allergy Therapeutics, YSlab, Sanofi, Bausch & Lomb, AstraZeneca and ThermoFisher Scientific; has received fees for participation in review activities from BTT; and has received lecture fees from MYLAN, Chiesi, ALK, and Stallergenes Greer.

List of abbreviations

AA: allergic asthma

AAQQ: Arabic Allergic Rhinitis Quality of Life Questionnaire

ACS: Allergy-Control-SCORE

AdSS: Adjusted Symptom Score

AIT: allergen immunotherapy

AOS: Allergy Outcome Survey

AR: allergic rhinitis

ARSQOL: Allergic Rhinitis-Specific Quality of Life

CARAT10: Control of Allergic Rhinitis and Asthma Test

COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments

CQ5: Congestion Quantifier Five-Item Screener

CQ7: Congestion Quantifier Seven-Item Test

EAACI: European Academy of Allergy and Clinical Immunology

EMA: European Medicines Agency

ESPRINT: Cuestionario ESPañol de Calidad de Vida en RINiTis

GA2LEN: Global Allergy and Asthma European Network

JRQLQ: Japan Rhinoconjunctivitis Quality of Life Questionnaire

MASK: Mobile Airways Sentinel Network

MCID: minimal clinically important difference.

MID: minimal important difference.

mini-RQLQ: Mini Rhinoconjunctivitis Quality of Life Questionnaire

NOSE: Nasal Obstruction Symptom Evaluation

NRQLQ: Nocturnal Rhinoconjunctivitis Quality of Life Questionnaire

PADQLQ: Pediatric Allergic Disease Quality of Life Questionnaire

PADQLQ: Pediatric Allergic Disease Quality of Life Questionnaire

PBI-AR: Patient Benefit Index - Allergic Rhinitis;

Ped-AR-QoL: Paediatric Allergic Rhinitis Quality of Life Questionnaire

PRO: patient-reported outcome

PRQLQ: Paediatric Rhinoconjunctivitis Quality of Life Questionnaire

RAPP: RhinAsthma Patient Perspective

RCAT: Rhinitis Control Assessment Test

RCSS: Rhinitis Control Scoring System

RHINASTHMA: a disease-specific QoL questionnaire

RQLQ(S): Standardized Rhinoconjunctivitis Quality of Life Questionnaire

RSUI: Rhinitis Symptom Utility Index

RTSS: Rhinoconjunctivitis Total Symptom Score

SCIT: subcutaneous allergen immunotherapy

SLIT: sublingual allergen immunotherapy

SoAR: Scale of Allergic Rhinitis

T4NSS: 4-item total nasal symptom score

T5SS: 5-item symptom score

T6SS: 6-item symptom score

Abstract

Although regulatory authorities have recently recommended the use of a combined symptom-medication score (CSMS) as a primary efficacy endpoint, none has been psychometrically validated. Here, we sought to determine to what extent allergic rhinitis (AR)-related patient-reported outcomes (symptom scores, medication scores, disease control scores, satisfaction or quality of life (QoL) scales) have been assessed for construct, content and/or criterion validity, reliability, responsiveness, and the minimal clinically important difference. We searched the PubMed database from January 1997 until June 2018 with logical combinations of keywords related to validation, AR, and patient-rated outcomes and scales. From a total of 1705 potentially relevant publications, 55 were reviewed. Despite the current emphasis on a CSMS for evaluating the efficacy of AIT in AR, symptom scores have not been extensively validated, and we did not find any publications describing the validation of a medication score. Disease control scales (mainly the Rhinitis Control Assessment Test, the Control of Allergic Rhinitis and Asthma Test, and the Allergic Rhinitis Control Test) and health-related QoL scales (mainly the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ) and the mini-RQLQ) have been extensively validated in AR but have some practical disadvantages as primary efficacy criteria in clinical trials.

Keywords: allergen immunotherapy, allergic rhinitis, score, validation, psychometric, reliability, responsiveness, minimally clinical important difference.

Introduction

The need for valid patient-reported outcomes (PROs)

Allergic rhinoconjunctivitis (AR) can impair respiratory health, quality of life (QoL), healthcare costs, social activities, sleep, academic/work performance (1-4), and can even increase the risk of traffic accidents (5). Furthermore, AR is acknowledged to be one step in the "allergic march" towards allergic asthma, with its potentially life-threatening and enormous societal cost consequences (6, 7). Although pharmacologic (symptomatic) medications can provide short-term symptom relief (8), they may not be effective or well tolerated by some patients – especially those with moderate-to-severe AR – and do not treat the underlying disease. The only currently available disease-modifying treatment for both AR and allergic asthma (i.e. with the potential to increase allergen tolerance and at least slow the allergic march) is allergen immunotherapy (AIT) (9-17).

In major markets such as the USA and the European Union, developers of new AIT formulations and other treatments for atopic diseases must demonstrate safety and efficacy in pivotal Phase III double-blind, placebo-controlled, randomized clinical trials (DBPC RCTs) (18). Safety and efficacy outcomes are influenced by many key aspects of clinical trial design, implementation and analysis. In the field of atopic disease, it is clear that many of these aspects require further optimization (10, 19-27). The choice of the criterion used to evaluate efficacy is crucial (10, 25, 28, 29)). From an ethical standpoint, one can even consider (somewhat provocatively) that the use of a non-validated PRO criterion to evaluate efficacy in a group of patients could even be equated to the use of a non-validated (i.e. non-approved) imaging method, functional testing method or medical device in clinical practice. Hence, establishing the validity of a PRO is ethically, legally and clinically important.

In the absence of a "gold standard", a large number of patient-reported scales, scores and instruments have been used to measure efficacy in the treatment of AR with AIT and/or symptomatic medications: nasal symptom scores, ocular symptom scores, total (nasal+ ocular) symptom scores, individual symptom scores, medication-adjusted symptom scores, medication scores, QoL scores, combined symptom-medication scores (CSMSs), visual analogue scales of overall symptom burden, disease control questionnaires, physician- and

patient-rated improvement/worsening, symptom-free days, "well days", "worst days", satisfaction questionnaires, nasal provocation tests, and conjunctival provocation tests (for comprehensive reviews, see (25, 28, 29)).

Regulatory authorities in both the USA and Europe have expressed a preference for a combined symptom-medication score (CSMS) that reflects both symptom severity and rescue medication intake (30, 31). However, a European Medicines Authority (EMA) guideline developed over the period 2007-2008 pointed out that "*Up to now, no validated symptom score exists, but the measurement of symptoms on a 4-point rating scale…. is generally accepted….. Likewise, no validated medication score exists…. So far, no validated system for balancing symptom and medication score exist.*" (30).

Outcomes for AIT trials have been described and commented on in depth (25, 28, 29)). However, we decided to focus on the three main quality domains applicable to PROs: validity, reliability and responsiveness (Figure 1), as described in more detail below. In the present review, we followed the definitions of validity (subdivided into content validity, construct validity, and criterion validity) given in the recent Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) (32-35).

Content validity

The COSMIN definition of content validity is "*the degree to which the content of an instrument is an adequate reflection of the construct to be measured*", i.e. the comprehensiveness of the PRO (33). In the present review, we considered that a group of researchers had assessed content validity if they had (i) explicitly stated this or (ii) had made a systematic or semi-systematic attempt to select and then test an instrument's constituent items.

Construct validity

According to COSMIN, construct validity is "*the degree to which the scores of an instrument are consistent with hypotheses (e.g. with regard to internal relationships, relationships to scores of other instruments, or*

differences between relevant groups)" (33). It notably "*includes structural validity, hypothesis testing, and cross-cultural validity*". In the present review, we considered that a group of researchers had assessed construct validity if they had (i) explicitly stated this, (ii) had related the AR score under study to other AR scores, or (iii) had looked at whether the AR score under study differentiated between a group of patients with AR (as defined by a physician, for example) and healthy controls.

Criterion validity

A PRO is said to have criterion validity if its results are an adequate reflection of a "gold standard" (33). However, there is debatably no universally accepted "gold standard" PRO for AR symptom and disease control scores, and there is overlap between criterion validity and construct validity. In the present review, we considered that a group of researchers had assessed criterion validity if they had explicitly stated this.

Reliability and responsiveness

Reliability is defined by COSMIN as "*the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health-related PRO (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)*" (33). In the present review, we considered that a group of researchers had assessed reliability if they had performed test-retest, inter-rater and intra-rater analyses or studied internal consistency (Figure 1) (33). Responsiveness is the ability of an HR-PRO instrument to detect a change over time in the construct to be measured (33). A key component of responsiveness is the "minimal clinically important difference" (MCID), which denotes the smallest score or change in score likely to be important from the patient's perspective (36-38). The MCID can be determined using an anchor-based method (often relative to a 15-point global rating of change scale (GRCS)) or the change in the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ)), a distribution-based method (often 0.5 of the standard deviation of the mean PRO value) or an expert consensus (36-40). We considered that a group of researchers had assessed responsiveness if they had

assessed the MCID or, more generally, the ability of an instrument to detect a change over time in the construct to be measured.

Hence, the objective of the present review was to contribute to the ongoing debate on validated efficacy criteria in AIT trials by determining to what extent AR-related PROs (symptom scales, medication scales, disease control scales, and rhinitis-specific QoL scales) have been validated, i.e. assessed for validity, reliability and responsiveness (41, 42).

Methods

Using the US National Library of Medicine's online tool, we searched the PubMed database from January 1997 until June 2018 with logical combinations of the following keywords: allerg*, rhin*, valid*, validat*, score*, scale*, control*; outcome*, psychometr*, MCID, MID, relevan*, anchor-based, distribution-based, VAS, quality, and QoL. The search was limited to publications in English. Review articles, position papers, "rostrum" articles, and abstract publications were also excluded. The search period was limited to the last 20 years because this corresponds to the period for the clinical development and marketing of "modern" AIT formulations, such as sublingual tablets. Abstracts were obtained and screened for relevance. If an abstract was found to be relevant, the full publication was retrieved and read. The reference lists of retrieved full publications were also screened for other relevant publications. Lastly, we excluded publications on instruments that could not, in our opinion, be scored on a daily or even weekly basis before and then throughout a pivotal DBPC RCT in AR (e.g. treatment satisfaction questionnaires and illness perception questionnaires) (43). The selected publications were rated as to whether they included assessments of construct validity (convergent, discriminant, cross-sectional, longitudinal or language validity), content validity (the comprehensiveness of item selection), criterion validity (concurrent or predictive, vs. a gold standard or an outcome), reliability (inter-/intra-rater, test-retest, or internal consistency), and responsiveness (ability to detect change, including the MCID), defined according to COSMIN (32-35).

Results and Discussion

Search results

Our search initially yielded a total of 1705 potentially relevant publications (Figure 2). Many of these publications were spurious hits unrelated to AR; they were produced by our decision to use the keyword root "rhin*" (to include "rhinasthma", for example), rather than "rhini*". After screening the initial list of publications and applying our exclusion criteria, we examined 102 full-text publications in detail. We excluded (i) PROs tested exclusively in rhinosinusitis or asthma, (ii) tools using to diagnose AR or estimate

the prevalence of AR in population studies (such as the Score for Allergic Rhinitis (44, 45)) or classify types of AR (such as the Allergic Rhinitis and its Impact on Asthma (ARIA) classification) (46, 47), (iii) PROs used to measure satisfaction with treatment or the treatment outcome alone (such as the ESPIA questionnaire) (48), and (iv) illness perception questionnaires (43, 49)

After further exclusions, 55 full-text publications were included in the present review (Supplementary Table 1).

Symptom scores

Despite the publication of the EMA's observations in 2008, we found that symptom scores have not undergone extensive validation, and - with the possible exception of nasal congestion scores - have been far less extensively validated than disease control scores and QoL scores (Supplementary Tables 1 and 2). Hence, the EMA's call for more validation is as relevant now as it was in 2008 (30). This lack of validation contrasts greatly with the high frequency with which total symptom scores (such as the Rhinoconjunctivitis Total Symptom Score (RTSS)) have been used in major DBPC RCTs, and the regulatory authorities' recommendations of these scores. Indeed, we did not find any specific references to the psychometric validation of the frequently reported RTSS, other than a study of the MCID. Devillier et al. used anchor-based and SD-based methods to determine the MCID in the RTSS (six symptoms scored from 0 to 3, giving a total score ranging from 0 to 18) as part of a multicentre study of 806 patients (253 children, 250 adolescents and 303 adults) consulting for grass-pollen-induced AR (50) (Supplementary Table 1). The researchers suggested that the MCID could be rounded down to 1 but emphasized that any subsequently determined MCIDs for AR induced by other allergens may not be the same as those determined for grass-pollen-induced AR (50). Despite the lack of specific psychometric validation studies for the RTSS, other reports have highlighted the strong or moderate correlations between this and other commonly used PROs, such as the VAS. In a further analysis of the above-mentioned study population of 806 patients, Devillier et al. showed that weekly changes in RQLQ were correlated with changes in the VAS and RTSS scores (51). In all three age groups, the VAS score was well correlated with the weekly mean RTSS score (Pearson's r:

0.79–0.88) and moderately correlated with the weekly mean RQLQ score (Pearson's r: 0.64–0.80); in other words, Devillier et al.'s study can be interpreted as an assessment of construct validity for the RTSS.

Higaki et al. calculated the MCID for a six-item symptom score (T6SS) that had been extracted from data on the JRQLQ and comprised the same six individual symptom scores as the RTSS (sneezing, rhinorrhoea, nasal congestion, itchy nose, itchy eyes, and watery eyes) (52). The data came from a study of 55 patients with Japanese cedar/cypress pollinosis (17 males, 38 females; mean (range) age 53.1 (23-79)). The JRQLQ rates each individual symptom score from 0 (no symptoms) to 4 (very severe symptoms), whereas the RTSS rates each score from 0 to 3. Furthermore, clinical change was rated against a 5-point "face scale" as the anchor, rather than the commonly used 15-point GRCS. In the 2009 and 2010 Japanese cedar and cypress pollen seasons, the MCID was 0.686 and 0.531 units per item, respectively. Even after considering the differences in the rating scales, these values are markedly higher than those calculated by Devillier et al. for the RTSS (i.e. ~0.167 per item) in a European population of patients with grass-pollen-induced AR (50). Higaki et al. also calculated the MCID for a T5SS with three nasal symptoms (sneezing, rhinorrhoea, and nasal congestion) and two ocular symptoms (watery and itchy eyes) rated on a 5-point scale. The per-item MCID was much lower for the T5SS than for the T6SS (0.285 in 2009 and 0.288 in 2010) (52).

We found that the most extensively validated symptom-related scores are the Nasal Obstruction Symptom Evaluation (NOSE) scale and the 5- and 7-item versions of the CQ5 and CQ7 (53-55). The NOSE scale is a short, easy-to-complete PRO. Although the NOSE's name refers to "symptom evaluation", it might be considered as a measure of disease control because the patient is asked "Over the past 1 month, how much of a problem were the following conditions for you?". The NOSE scale also includes a VAS of nasal obstruction. It was assessed for construct, content, and criterion validity, reliability and responsiveness in patients presenting with drug-refractory symptoms of chronic nasal obstruction for least 3 months, rather than in a context of AR (53) (Supplementary Table 2). Despite its extensive validation, the NOSE scale is essentially an individual symptom score with regard to AR, and is probably not a broad enough measure to be of value in clinical trials in this field.

According to its developers, the Congestion Quantifier (CQ) was originally formulated as a screening tool for nasal congestion but can distinguish between patients with different levels of severity of AR symptoms, and between cases and controls (54). It is also sensitive to changes in a patient's condition. The 7-item CQ (CQ7) was assessed for construct and content validity, reliability and responsiveness in a study of 129 patients with confirmed AR, 155 patients with nasal congestion, and 70 control subjects recruited at a total of 49 sites (54) (Supplementary Table 2). Although the CQ focuses on the preceding week (rather than the preceding or current day), we consider that this tool might be a suitable measure of clinical efficacy over a long-term RCT in AR.

Medication scores

We did not find any publications describing the validation of a "pure" medication score. Again, this complicates the regulatory authorities' preference for a CSMS as the recommended score in pivotal clinical trials. Despite the presence of the word "control" in its name, the Allergy-Control-SCORE (ACS) is a CSMS (56, 57). The medication score component is based on a list of 745 commercially available preparations of drugs with relevance in AR and allergic asthma, each of which is attributed a target organ (nose, eyes or lungs) and a number of points. In a study of 81 patients with AR and/or allergic asthma and 40 healthy controls, the ACS was assessed for construct and content validity, and reliability with regard to nasal, ocular and lung symptoms (56) and then with regard to nasal and ocular symptoms only (57) (Supplementary Table 2).

Disease control scores

We found that disease control scales have been extensively validated in AR (Supplementary Table 1). However, one should bear in mind that there is no single definition of "disease control"; nevertheless, PROs referred to as disease control scales usually assess most or all of the following domains - albeit to varying extents: (i) symptom intensity and/or frequency (i.e. symptom scoring), (ii) physical impairments (such as air flow), (iii) the frequency and consequences of exacerbations, and (iv) the disease burden (i.e. the

278 "bothersomeness" of symptoms, and disease–related impairments in social, physical or professional
279 activities) (42). The Rhinitis Control Scoring System (RCSS) rates five symptoms (sneezing, rhinorrhoea,
280 nasal obstruction, nasal itching, and conjunctivitis) in terms of severity and frequency over the previous 7
281 days, but does not probe bothersomeness or exacerbations; despite the presence of "control" in the score's
282 name, it should probably be classified as a retrospective symptom score (58). The RCSS has nevertheless
283 been thoroughly validated (construct, content and possibly criterion validity, reliability and internal
284 consistency) in a study of 50 patients with AR, and warrants further investigation as an efficacy criterion.
285
286 We found that the most extensively validated disease control PROs were the Rhinitis Control Assessment
287 Test (RCAT) and the Control of Allergic Rhinitis and Asthma Test (CARAT) (Supplementary Tables 1 and
288 2). In a study of data from the placebo arm of two DBPC RCTs (on 29 grass-pollen-allergic patients and 76
289 birch-pollen-allergic patients), Liedtke et al. found strong correlations between the RCAT and RQLQ scores
290 in grass-pollen-allergic patients and birch-pollen-allergic patients (r = -0.871 and -0.795, respectively), a
291 strong correlation between the RCAT score and the RTSS in grass-pollen-allergic patients (r = -0.811) and a
292 moderate correlation between the RCAT score and the RTSS in birch-pollen-allergic patients (r = -0.539)
293 (59). These findings suggest that efficacy in clinical trials may be evidenced by disease control scores.
294 The CARAT was initially developed as a 17-question test in Portuguese, by using a comprehensive set of
295 methodological steps (including expert and patient panels) to verify the design quality, face validity and
296 content validity (60). A 10-item version was then produced (using descriptive analysis, exploratory factor
297 analysis and internal consistency) and validated in a cross-sectional study of 193 adults with AR and allergic
298 asthma from 15 outpatient clinics in Portugal (61). On the basis of the studies reviewed here, the CARAT10
299 has been evaluated for construct and criterion validity, reliability, responsiveness, and the MCID
300 (Supplementary Table 2). Furthermore, the paediatric version of the CARAT (CARAT Kids) has been
301 evaluated for construct and content validity, reliability, responsiveness, and the MCID (Supplementary
302 Table 2). Another extensively validated disease control score is the ARCT, for which construct and content
303 validity, reliability and responsiveness to change have been examined (62-65). The ARCT has notably been

304 applied to the management of "step-up" and "step-down" approaches to the prescription of symptomatic
305 medications in patients with AR, as a function of the level of disease control achieved (63, 65).
306
307 *QoL scales*
308 Health-related QoL scales (and especially the RQLQ and the mini-RQLQ) have been extensively validated
309 in AR – notably due to the pioneering work by the RQLQ's developers. Indeed, the mini-RQLQ and the
310 RQLQ have been considered by some investigators as "gold standards" in AR against which the criterion
311 validity of other PROs can be assessed (66) (67).
312 The RQLQ has been used as a secondary efficacy criterion in large, multicentre trials of sublingual grass
313 pollen AIT formulations, where changes in the QoL were correlated with changes in a symptom score
314 (RTSS) (68, 69). However, the use of the RQLQ as a primary endpoint to be scored daily in a pivotal DBPC
315 RCT in AR is compromised by the relatively long administration time (around 10 minutes). Further, the
316 recall period in the RQLQ is 7 days, and so this PRO is not an instantaneous score. Furthermore, the use of
317 instruments with recall periods of 1 week or more may mean that some relevant days with symptoms will be
318 missed (29). The mini-RQLQ is a more rapid alternative, and has also been extensively validated. Several
319 other AR-related QoL tools have been partially or fully validated in specific contexts and languages; these
320 include the RHINASTHMA disease-specific QoL questionnaire, the RhinAsthma Patient Perspective
321 (RAPP, derived from the RHINASTHMA questionnaire), the Arabic Allergic Rhinitis Quality of Life
322 Questionnaire (AAQQ), the Pediatric Allergic Disease Quality of Life Questionnaire (PADQLQ), and the
323 Cuestionario ESPañol de Calidad de Vida en RINiTis (ESPRINT) questionnaire (70-72) (73, 74).
324
325 *Other scores, and perspectives for future research*
326 Lastly, we decided to review the Allergy Outcome Survey (AOS) (75). This is something of a hybrid score,
327 since it combines items related to disease control, symptom severity, medication use, and treatment
328 outcome. Kemker et al. concluded that the AOS was a brief, reliable, easy-to-use performance metric that
329 could evaluate change with therapy when used alone or combined with other PROs (75). However, the

AOS's feasibility as a regularly scored PRO in an RCT is debatable. It might also be of value to explore the wider use of illness perception scales (such as the Brief Illness Perception Questionnaire) and treatment satisfaction questionnaires (43, 48) - at least as secondary endpoints. On the basis of our assessments of validation status, the strengths and weakness of the various types of score are summarized in Table 1.

As mentioned above, the present systematic review was prompted by regulatory calls for a validated symptom score that could be used in pivotal clinical trials. By definition, PROs are reported by patients, and patients spend most of their time away from medical facilities. Hence, it would be particularly interesting to explore the combination of a validated PRO in AR with the "m-health" and "e-health" information technology tools that are now widely available among the general public (e.g. smartphones with Internet access). Many different apps have been developed for patients with asthma, although the focus is usually self-management and not PRO data collection *per se* (76, 77). Some research in the field of AR () has already been described. A recent review identified 72 apps related to AR; although most were related to patient education and pollen count, 14 (19.4%) were related to symptom tracking (78). By way of an example, the Mobile Airways Sentinel network (MASK) rhinitis tool is used to implement care pathways for AR by using a smartphone-/tablet-based application and a clinical decision support system (79-81). The user self-evaluates his/her level of AR control on a VAS. The system's developers consider that the MASK-rhinitis VAS app (currently available in 16 languages and 23 countries) is a reliable, valid tool for assessing AR allergic control at the population level (79).

Conclusion

Despite the regulatory authorities' current emphasis on CSMSs as the basis for evaluating efficacy of AIT in AR, neither of the latter's two components (the symptom score or the medication score) has been extensively validated. The most frequently used symptom score (RTSS) has not been thoroughly validated, although the MCID has been determined in patients with grass-pollen-induced AR. The validity of medication scores in AR has not been studied. In contrast, disease control scales and health-related QoL

scales have been extensively validated. However, the adoption of a disease control or QoL scale as a primary endpoint in clinical trials will require a paradigm shift within the clinical community and regulatory authorities. The regulatory authorities' current requirement for daily symptom score data makes it difficult to envisage the use of retrospective or reflective disease control scores or disease-related QoL scores as primary efficacy endpoints. Nevertheless, some experts consider that a paradigm shift in this field is required, i.e. with further emphasis on the long-term control (over several months or years) of atopic diseases rather than on symptom control during a clinical trial that sometimes may last for just a few weeks or months (e.g. a single pollen season) (82). In the field of asthma, the Asthma Control Questionnaire (83) and the Asthma Control Test (84) have recently been recommended for use in both primary care and other clinical settings (85).

In conclusion, we have the following closing comments and recommendations: (i) the RTSS and a CSMS should undergo full psychometric validation (notably with regard to construct, content and criterion validity) in a well-powered clinical trial in AR as a matter of some urgency; (ii) in collaboration with all the relevant stakeholders, the regulatory authorities should evaluate the possible use of a disease control score as the primary efficacy criterion in a Phase III clinical trial; (iii) whenever possible, a Phase III clinical trial inAR should have a QoL score as a secondary efficacy criterion, and (iv) PROs validated in AR should be combined with "m-health" and "e-health" tools for symptom/disease control tracking.

Competing interests

Moises A. Calderón has received lecture fees from ALK, Stallergenes Greer, Merck, Allergopharma, HAL Allergy, and consultancy fees from ALK, Stallergenes Greer, Merck, HAL Allergy, and ASIT biotech.

Thomas Casale has been an investigator and consultant for Novartis and Genentech with all fees to his university employer.

Pascal Demoly has received consultancy fees from ALK, Stallergenes Greer, Allergy Therapeutics, YSlab, Sanofi, Bausch & Lomb, AstraZeneca and ThermoFisher Scientific; has received fees for participation in review activities from BTT; and has received lecture fees from MYLAN, Chiesi, ALK, and Stallergenes Greer.

Authors' contributions

MC and PC conceived the study and collected the data. All the authors analyzed the data and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements.

We thank David Fraser D.Phil. (Biotech Communication SARL, Ploudalmezeau, France) for providing copy-editing assistance (funded by ALK-Abelló A/S (Hørsholm, Denmark) and Stallergenes Greer plc (London, UK)).

Table 1. Strengths and weaknesses of the various types of PROs assessed for validation status.

Type of score	Strengths	Weaknesses
Total symptom scores	• Familiar to clinicians and researchers, due to its extensive use in the literature • The PRO that corresponds closely to a "physiological" score, reflecting the effect of disease on the body • Can be scored rapidly once or more a day	• **Not extensively validated** • Subject to inter-individual differences in the perception of symptom severity
Nasal congestion scores	• **Extensively validated** • The PRO that corresponds closely to a "physiological" score, reflecting the effect of disease on the body • Can be scored rapidly once or more a day	• Does not reflect all the symptoms of AR • Subject to inter-individual differences in the perception of symptom severity
Combined symptom and medication scores	• Conceptually attractive, since the use of rescue medical testifies to a need for symptom relief • Can be scored rapidly once or more a day	• **Not extensively validated** • Arbitrary balance between the "symptom" and the "medication" parts of the score • Subject to inter-individual differences in the perception of symptom severity
Disease control scores	• **Extensively validated** • Takes account of inter-individual differences in the perception of symptom severity • May correspond most closely to "real-life" efficacy	• Relatively long administration time • Poor time resolution (retrospective/reflective scoring over weeks, typically).
Quality of life scores	• **Extensively validated**	• Relatively long administration time • Far removed from a clinical effect • Poor time resolution (retrospective/reflective scoring over weeks, typically).

Table 2. Summary of validation status for symptom scores, disease control scores, QoL scores and other PROs in AR, based on the publications listed in Supplementary Table 1.

Name of instrument	Type of instrument	Study population(s)	References	Construct validity	Content validity	Criterion validity	Reliability	Responsiveness	MCID (anchor)	MCID (SD)
NOSE	symptom score	children and adults	Stewart 2004 (53), Zicari 2015 (86)	✓	✓	✓	✓	✓	–	–
CQ5 and CQ7	symptom score	adults	Stull 2007 (54), Stull 2010 (55), Valero 2011 (87)	✓	✓	–	✓	✓	–	–
RASS	symptom score	adults	Wasserfallen 1997 (88)	✓	✓	–	✓	✓	–	–
NtSS	symptom score	teens and adults	Santanello 2006 (89)	✓	–	–	✓	✓	–	–
RSUI	symptom score	adults	Revicki 1998 (90), Lo 2006 (91)	✓	–	✓	✓	–	–	–

Name of instrument	Type of instrument	Study population(s)	References	Construct validity	Content validity	Criterion validity	Reliability	Responsiveness	MCID (anchor)	MCID (SD)
PBI-AR	symptom score	adults	Franzke 2011 (92), Demoly 2015 (67)	✓	–	✓	✓	–	–	✓
SoAR	symptom score	adults	Johnson 2004 (93)	✓	–	–	✓	–	–	–
RTSS	symptom score	children, teens and adults	Devillier 2014 (50)	–	–	–	–	–	✓	✓
AdSS	symptom score	children, teens and adults	Grouin 2011 (94)	✓	–	–	–	–	–	–
A T4NSS	symptom score	adults	Barnes 2010 (36)	–	–	–	–	–	✓	✓
A T5SS *(in Japanese cedar/cypress pollinosis)*	symptom score	adults	Higaki 2013 (52)	–	–	–	–	–	✓	–
A T6SS *(in Japanese cedar/cypress pollinosis)*	symptom score	adults	Higaki 2013 (52)	–	–	–	–	–	✓	–
ACS	combined symptom and medication score	teens and adults	Hafner 2011 (56), Hafner 2012 (57)	✓	✓	–	✓	–	–	–

Name of instrument	Type of instrument	Study population(s)	References	Construct validity	Content validity	Criterion validity	Reliability	Responsiveness	MCID (anchor)	MCID (SD)
CARAT10	disease control score	adults	Fonseca 2010) (61), Fonseca 2012 (95), van der Leeuw 2015 (96)	✓	–	✓	✓	✓	✓	✓
CARATkids	disease control score	children and teens	Linhares 2014 (47), Borrego 2014 (97), Amaral 2017 (98), Emons 2017 (99)	✓	✓	–	✓	✓	✓	✓
RCAT	disease control score	teens and adults	Schatz 2010 (100), Schatz 2012 (101), Meltzer 2013 (102), Fernandes 2016 (103)	✓	✓	–	✓	–	✓	✓
RCSS	disease control score	adults	Boulay 2016 (58)	✓	✓	✓	✓	–	–	–

Name of instrument	Type of instrument	Study population(s)	References	Construct validity	Content validity	Criterion validity	Reliability	Responsiveness	MCID (anchor)	MCID (SD)
ARCT	disease control score	children, teens and adults	Demoly 2011 (62), Wang 2014 (64), Wang 2016 (63), Zhu 2018 (65)	✓	✓	–	✓	✓		
VAS for the burden of AR in primary care	disease control score	adults	Demoly 2013 (104)	✓	–	–	–	✓	–	–
VASs for the control of overall, nasal, ocular, and asthma symptoms	disease control score	not stated	Caimmi 2017 (81)	–	–	–	✓	–	–	–
RQLQ	QoL score	adults	Leong 1999 (105), Leong 2005 (106), Okuda 2009 (107), Yüksel 2009 (108)	✓	✓	–	✓	✓	–	–
mini-RQLQ	QoL score	adults	Yüksel 2009 (108), Barnes 2010 (36), Schatz 2012 (101)	✓	✓	✓	✓	✓	✓	✓

Name of instrument	Type of instrument	Study population(s)	References	Construct validity	Content validity	Criterion validity	Reliability	Responsiveness	MCID (anchor)	MCID (SD)
RAPP	QoL score	teens and adults	Braido 2012 (109), Molinengo 2017 (70)	✓	✓	✓	✓	✓	–	✓
AAQQ	QoL score	adults	AbuRuz 2009 (71)	✓	✓	–	✓	✓	✓	–
PADQLQ	QoL score	children and teens	Roberts 2003 (72), Kiotseridis 2011 (110)	✓	–	–	✓	✓	✓	✓
ESPRINT	QoL score	adults	Esprint Study Investigators 2007 (73)	✓	✓	–	✓	✓	–	–
RQLQ(S)	QoL score	adults	Juniper 1999 (111)	✓	–	–	✓	✓	✓	–
RHINASTHMA	QoL score	adults	Baiardini 2003 (74)	✓	–	✓	✓	✓	–	–
JRQLQ	QoL score	adults	Okuda 2005 (107), Higaki 2013 (52)	✓	–	–	✓	–	✓	–

Name of instrument	Type of instrument	Study population(s)	References	Construct validity	Content validity	Criterion validity	Reliability	Responsiveness	MCID (anchor)	MCID (SD)
NRQLQ	QoL score	adults	Juniper 2003 (112)	✓	–	–	✓	✓	–	–
PRQLQ	QoL score	children and teens	Juniper 1998 (113), Yüksel 2009 (114)	✓	–	–	✓	✓	–	–
Ped-AR-QoL	QoL score	children and teens	Mavroudi 2016 (115)	✓	–	–	✓	–	–	–

Age classes of study subjects are defined as children: younger than 13; teens: aged 13 to 17; adults: aged 18 and over

Abbreviations (in alphabetical order)

AAQQ: Arabic Allergic Rhinitis Quality of Life Questionnaire

ACS: Allergy-Control-SCORE

AdSS: Adjusted Symptom Score

AOS: Allergy Outcome Survey

ARSQOL: Allergic Rhinitis-Specific Quality of Life

CARAT10: Control of Allergic Rhinitis and Asthma Test

CQ5: Congestion Quantifier Five-Item Screener

CQ7: Congestion Quantifier Seven-Item Test

ESPRINT: Cuestionario ESPañol de Calidad de Vida en RINiTis

JRQLQ: Japan Rhinoconjunctivitis Quality of Life Questionnaire

MCID: minimal clinically important difference.

MID: minimal important difference.

mini-RQLQ: Mini Rhinoconjunctivitis Quality of Life Questionnaire

NOSE: Nasal Obstruction Symptom Evaluation

NRQLQ: Nocturnal Rhinoconjunctivitis Quality of Life Questionnaire

PADQLQ: Pediatric Allergic Disease Quality of Life Questionnaire

PADQLQ: Pediatric Allergic Disease Quality of Life Questionnaire

PBI-AR: Patient Benefit Index - Allergic Rhinitis;

Ped-AR-QoL: Paediatric Allergic Rhinitis Quality of Life Questionnaire

PRQLQ: Paediatric Rhinoconjunctivitis Quality of Life Questionnaire

RAPP: RhinAsthma Patient Perspective

RCAT: Rhinitis Control Assessment Test

RCSS: Rhinitis Control Scoring System

RHINASTHMA: a disease-specific QoL questionnaire

RQLQ(S): Standardized Rhinoconjunctivitis Quality of Life Questionnaire

RSUI: Rhinitis Symptom Utility Index

RTSS: Rhinoconjunctivitis Total Symptom Score

SoAR: Scale of Allergic Rhinitis

T4NSS: a 4-item total nasal symptom score.

References

1. Blaiss MS, Hammerby E, Robinson S, Kennedy-Martin T, Buchs S. The burden of allergic rhinitis and allergic rhinoconjunctivitis on adolescents: A literature review. Ann Allergy Asthma Immunol. 2018;121(1):43-52 e3.
2. Linneberg A, Dam Petersen K, Hahn-Pedersen J, Hammerby E, Serup-Hansen N, Boxall N. Burden of allergic respiratory disease: a systematic review. Clin Mol Allergy. 2016;14:12.
3. Ozdoganoglu T, Songu M. The burden of allergic rhinitis and asthma. Ther Adv Respir Dis. 2012;6(1):11-23.
4. Maspero J, Lee BW, Katelaris CH, Potter PC, Cingi C, Lopatin A, et al. Quality of life and control of allergic rhinitis in patients from regions beyond western Europe and the United States. Clin Exp Allergy. 2012;42(12):1684-96.
5. Demoly P, Maigret P, Elias Billon I, Allaert FA. Allergic rhinitis increases the risk of driving accidents. J Allergy Clin Immunol. 2017;140(2):614-6.
6. Shaker M. New insights into the allergic march. Curr Opin Pediatr. 2014;26(4):516-20.
7. Thomsen SF. Epidemiology and natural history of atopic diseases. Eur Clin Respir J. 2015;2.
8. Ridolo E, Montagni M, Melli V, Braido F, Incorvaia C, Canonica GW. Pharmacotherapy of allergic rhinitis: current options and future perspectives. Expert Opin Pharmacother. 2014;15(1):73-83.
9. Porcaro F, Corsello G, Pajno GB. SLIT's Prevention of the Allergic March. Curr Allergy Asthma Rep. 2018;18(5):31.
10. Pfaar O, Alvaro M, Cardona V, Hamelmann E, Mosges R, Kleine-Tebbe J. Clinical trials in allergen immunotherapy: current concepts and future needs. Allergy. 2018;73(9):1775-83.
11. Dhami S, Nurmatov U, Arasi S, Khan T, Asaria M, Zaman H, et al. Allergen immunotherapy for allergic rhinoconjunctivitis: A systematic review and meta-analysis. Allergy. 2017;72(11):1597-631.
12. Jutel M, Agache I, Bonini S, Burks AW, Calderon M, Canonica W, et al. International Consensus on Allergen Immunotherapy II: Mechanisms, standardization, and pharmacoeconomics. J Allergy Clin Immunol. 2016;137(2):358-68.
13. Roberts G, Pfaar O, Akdis CA, Ansotegui IJ, Durham SR, Gerth van Wijk R, et al. EAACI Guidelines on Allergen Immunotherapy: Allergic rhinoconjunctivitis. Allergy. 2018;73(4):765-98.
14. Jutel M, Agache I, Bonini S, Burks AW, Calderon M, Canonica W, et al. International consensus on allergy immunotherapy. J Allergy Clin Immunol. 2015;136(3):556-68.
15. Martignago I, Incorvaia C, Ridolo E. Preventive actions of allergen immunotherapy: the facts and the effects in search of evidence. Clin Mol Allergy. 2017;15:13.
16. Pfaar O, Bonini S, Cardona V, Demoly P, Jakob T, Jutel M, et al. Perspectives in allergen immunotherapy: 2017 and beyond. Allergy. 2018;73 Suppl 104:5-23.
17. Burks AW, Calderon MA, Casale T, Cox L, Demoly P, Jutel M, et al. Update on allergy immunotherapy: American Academy of Allergy, Asthma & Immunology/European Academy of Allergy and Clinical Immunology/PRACTALL consensus report. J Allergy Clin Immunol. 2013;131(5):1288-96 e3.
18. Bonertz A, Roberts GC, Hoefnagel M, Timon M, Slater JE, Rabin RL, et al. Challenges in the implementation of EAACI guidelines on allergen immunotherapy: A global perspective on the regulation of allergen products. Allergy. 2018;73(1):64-76.
19. Baiardini I, Bousquet PJ, Brzoza Z, Canonica GW, Compalati E, Fiocchi A, et al. Recommendations for assessing patient-reported outcomes and health-related quality of life in clinical trials on allergy: a GA(2)LEN taskforce position paper. Allergy. 2010;65(3):290-5.
20. Baiardini I, Braido F, Bindslev-Jensen C, Bousquet PJ, Brzoza Z, Canonica GW, et al. Recommendations for assessing patient-reported outcomes and health-related quality of life in patients with urticaria: a GA(2) LEN taskforce position paper. Allergy. 2011;66(7):840-4.
21. Brozek JL, Bousquet J, Agache I, Agarwal A, Bachert C, Bosnic-Anticevich S, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines-2016 revision. J Allergy Clin Immunol. 2017;140(4):950-8.
22. Calderon MA, Casale TB, Togias A, Bousquet J, Durham SR, Demoly P. Allergen-specific immunotherapy for respiratory allergies: from meta-analysis to registration and beyond. J Allergy Clin Immunol. 2011;127(1):30-8.
23. Calderon MA, Kleine-Tebbe J, Linneberg A, De Blay F, Hernandez Fernandez de Rojas D, Virchow JC, et al. House Dust Mite Respiratory Allergy: An Overview of Current Therapeutic Strategies. J Allergy Clin Immunol Pract. 2015;3(6):843-55.

24. Canonica GW, Baena-Cagnani CE, Bousquet J, Bousquet PJ, Lockey RF, Malling HJ, et al. Recommendations for standardization of clinical trials with Allergen Specific Immunotherapy for respiratory allergy. A statement of a World Allergy Organization (WAO) taskforce. Allergy. 2007;62(3):317-24.
25. Nelson HS, Calderon MA, Bernstein DI, Casale TB, Durham SR, Andersen JS, et al. Allergen Immunotherapy Clinical Trial Outcomes and Design: Working Toward Harmonization of Methods and Principles. Curr Allergy Asthma Rep. 2017;17(3):18.
26. Papadopoulos NG, Agache I, Bavbek S, Bilo BM, Braido F, Cardona V, et al. Research needs in allergy: an EAACI position paper, in collaboration with EFA. Clin Transl Allergy. 2012;2(1):21.
27. Passalacqua G. Recommendations for appropriate sublingual immunotherapy clinical trials. World Allergy Organ J. 2014;7(1):21.
28. Pfaar O, Anders C, Klimek L. Clinical outcome measures of specific immunotherapy. Curr Opin Allergy Clin Immunol. 2009;9(3):208-13.
29. Pfaar O, Demoly P, Gerth van Wijk R, Bonini S, Bousquet J, Canonica GW, et al. Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI Position Paper. Allergy. 2014;69(7):854-67.
30. EMA. European Medicines Agency (EMA) Guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases [Internet]. Accessed at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003605.pdf European Medicines Agency: London; 2008.
31. FDA. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Allergic Rhinitis: Developing Drug Products for Treatment Guidance for Industry, Draft Guidance [Internet]. Accessed at http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071293.pdf. Center for Drug Evaluation and Research; 2016.
32. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006;6:2.
33. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther. 2016;20(2):105-13.
34. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.
35. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.
36. Barnes ML, Vaidyanathan S, Williamson PA, Lipworth BJ. The minimal clinically important difference in allergic rhinitis. Clin Exp Allergy. 2010;40(2):242-50.
37. Meltzer EO, Wallace D, Dykewicz M, Shneyer L. Minimal Clinically Important Difference (MCID) in Allergic Rhinitis: Agency for Healthcare Research and Quality or Anchor-Based Thresholds? J Allergy Clin Immunol Pract. 2016;4(4):682-8 e6.
38. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102-9.
39. Juniper EF, Guyatt GH. Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis. Clin Exp Allergy. 1991;21(1):77-83.
40. Juniper EF, Guyatt GH, Griffith LE, Ferrie PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. J Allergy Clin Immunol. 1996;98(4):843-5.
41. Dietz de Loos DA, Segboer CL, Gevorgyan A, Fokkens WJ. Disease-specific quality-of-life questionnaires in rhinitis and rhinosinusitis: review and evaluation. Curr Allergy Asthma Rep. 2013;13(2):162-70.
42. Demoly P, Calderon MA, Casale T, Scadding G, Annesi-Maesano I, Braun JJ, et al. Assessment of disease control in allergic rhinitis. Clin Transl Allergy. 2013;3(1):7.
43. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006;60(6):631-7.

44. Annesi-Maesano I, Didier A, Klossek M, Chanal I, Moreau D, Bousquet J. The score for allergic rhinitis (SFAR): a simple and valid assessment method in population studies. Allergy. 2002;57(2):107-14.
45. Lam SC, Yeung CCY, Chan JHM, Lam DWC, Lam AHY, Annesi-Maesano I, et al. Adaptation of the Score for Allergic Rhinitis in the Chinese Population: Psychometric Properties and Diagnostic Accuracy. Int Arch Allergy Immunol. 2017;173(4):213-24.
46. Jauregui I, Davila I, Sastre J, Bartra J, del Cuvillo A, Ferrer M, et al. Validation of ARIA (Allergic Rhinitis and its Impact on Asthma) classification in a pediatric population: the PEDRIAL study. Pediatr Allergy Immunol. 2011;22(4):388-92.
47. Linhares DV, da Fonseca JA, Borrego LM, Matos A, Pereira AM, Sa-Sousa A, et al. Validation of control of allergic rhinitis and asthma test for children (CARATKids)--a prospective multicenter study. Pediatr Allergy Immunol. 2014;25(2):173-9.
48. Justicia JL, Cardona V, Guardia P, Ojeda P, Olaguibel JM, Vega JM, et al. Validation of the first treatment-specific questionnaire for the assessment of patient satisfaction with allergen-specific immunotherapy in allergic patients: the ESPIA questionnaire. J Allergy Clin Immunol. 2013;131(6):1539-46.
49. Pesut D, Raskovic S, Tomic-Spiric V, Bulajic M, Bogic M, Bursuc B, et al. Gender differences revealed by the Brief Illness Perception Questionnaire in allergic rhinitis. Clin Respir J. 2014;8(3):364-8.
50. Devillier P, Chassany O, Vicaut E, de Beaumont O, Robin B, Dreyfus JF, et al. The minimally important difference in the Rhinoconjunctivitis Total Symptom Score in grass-pollen-induced allergic rhinoconjunctivitis. Allergy. 2014;69(12):1689-95.
51. Devillier P, Bousquet PJ, Grassin-Delyle S, Salvator H, Demoly P, Bousquet J, et al. Comparison of outcome measures in allergic rhinitis in children, adolescents and adults. Pediatr Allergy Immunol. 2016;27(4):375-81.
52. Higaki T, Okano M, Kariya S, Fujiwara T, Haruna T, Hirai H, et al. Determining minimal clinically important differences in Japanese cedar/cypress pollinosis patients. Allergol Int. 2013;62(4):487-93.
53. Stewart MG, Witsell DL, Smith TL, Weaver EM, Yueh B, Hannley MT. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. Otolaryngol Head Neck Surg. 2004;130(2):157-63.
54. Stull DE, Krouse J, Meltzer EO, Roberts L, Kim S, Frank L, et al. Development and validation of the Congestion Quantifier seven-item test (CQ7): a screening tool for nasal congestion. Value Health. 2007;10(6):457-65.
55. Stull DE, Meltzer EO, Krouse JH, Roberts L, Kim S, Frank L, et al. The congestion quantifier five-item test for nasal congestion: refinement of the congestion quantifier seven-item test. Am J Rhinol Allergy. 2010;24(1):34-8.
56. Hafner D, Reich K, Matricardi PM, Meyer H, Kettner J, Narkus A. Prospective validation of 'Allergy-Control-SCORE(TM)': a novel symptom-medication score for clinical trials. Allergy. 2011;66(5):629-36.
57. Hafner D, Reich K, Zschocke I, Lotzin A, Meyer H, Kettner J, et al. Prospective validation of the "rhino conjunctivitis allergy-control-SCORE(c)" (RC-ACS(c)). Clin Transl Allergy. 2012;2(1):17.
58. Boulay ME, Boulet LP. The Rhinitis Control Scoring System: Development and validation. Am J Rhinol Allergy. 2016;30(1):54-9.
59. Liedtke JP, Mandl A, Kother J, Chwieralski J, Shah-Hosseini K, Raskopf E, et al. RCAT reflects symptom control and quality of life in allergic rhinoconjunctivitis patients. Allergy. 2018;73(5):1101-9.
60. Nogueira-Silva L, Martins SV, Cruz-Correia R, Azevedo LF, Morais-Almeida M, Bugalho-Almeida A, et al. Control of allergic rhinitis and asthma test--a formal approach to the development of a measuring tool. Respir Res. 2009;10:52.
61. Fonseca JA, Nogueira-Silva L, Morais-Almeida M, Azevedo L, Sa-Sousa A, Branco-Ferreira M, et al. Validation of a questionnaire (CARAT10) to assess rhinitis and asthma in patients with asthma. Allergy. 2010;65(8):1042-8.
62. Demoly P, Jankowski R, Chassany O, Bessah Y, Allaert FA. Validation of a self-questionnaire for assessing the control of allergic rhinitis. Clin Exp Allergy. 2011;41(6):860-8.
63. Wang Y, Chen H, Zhu R, Liu G, Huang N, Li W, et al. Allergic Rhinitis Control Test questionnaire-driven stepwise strategy to improve allergic rhinitis control: a prospective study. Allergy. 2016;71(11):1612-9.
64. Wang Y, Zhu R, Liu G, Li W, Chen H, Daures JP, et al. Prevalence of uncontrolled allergic rhinitis in Wuhan, China: a prospective cohort study. Am J Rhinol Allergy. 2014;28(5):397-403.
65. Zhu R, Wang J, Wu Y, Yang Y, Huang N, Yang Y, et al. The Allergic Rhinitis Control Test Questionnaire Is Valuable in Guiding Step-Down Pharmacotherapy Treatment of Allergic Rhinitis. J Allergy Clin Immunol Pract. 2018.
66. Juniper EF, Thompson AK, Ferrie PJ, Roberts JN. Development and validation of the mini Rhinoconjunctivitis Quality of Life Questionnaire. Clin Exp Allergy. 2000;30(1):132-40.

67. Demoly P, Aubier M, de Blay F, Wessel F, Clerson P, Maigret P. Evaluation of patients' expectations and benefits in the treatment of allergic rhinitis with a new tool: the patient benefit index - the benefica study. Allergy Asthma Clin Immunol. 2015;11(1):8.
68. Didier A, Malling HJ, Worm M, Horak F, Jager S, Montagut A, et al. Optimal dose, efficacy, and safety of once-daily sublingual immunotherapy with a 5-grass pollen tablet for seasonal allergic rhinitis. J Allergy Clin Immunol. 2007;120(6):1338-45.
69. Durham SR, Yang WH, Pedersen MR, Johansen N, Rak S. Sublingual immunotherapy with once-daily grass allergen tablets: a randomized controlled trial in seasonal allergic rhinoconjunctivitis. J Allergy Clin Immunol. 2006;117(4):802-9.
70. Molinengo G, Baiardini I, Braido F, Loera B. RhinAsthma patient perspective: A Rasch validation study. J Asthma. 2018;55(2):119-23.
71. AbuRuz SM, Bulatova NR, Tawalbeh MI. Development and validation of the Arabic allergic rhinitis quality of life questionnaire. Saudi Med J. 2009;30(12):1577-83.
72. Roberts G, Hurley C, Lack G. Development of a quality-of-life assessment for the allergic child or teenager with multisystem allergic disease. J Allergy Clin Immunol. 2003;111(3):491-7.
73. Group ES, Investigators, Valero A, Alonso J, Antepara I, Baro E, et al. Development and validation of a new Spanish instrument to measure health-related quality of life in patients with allergic rhinitis: the ESPRINT questionnaire. Value Health. 2007;10(6):466-77.
74. Baiardini I, Pasquali M, Giardini A, Specchia C, Passalacqua G, Venturi S, et al. Rhinasthma: a new specific QoL questionnaire for patients with rhinitis and asthma. Allergy. 2003;58(4):289-94.
75. Kemker BJ, Corey JP, Branca J, Gliklich RE. Development of the allergy outcome survey for allergic rhinitis. Otolaryngol Head Neck Surg. 1999;121(5):603-5.
76. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev. 2013(11):CD010013.
77. Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T. The Potential of Mobile Apps for Improving Asthma Self-Management: A Review of Publicly Available and Well-Adopted Asthma Apps. JMIR Mhealth Uhealth. 2017;5(8):e113.
78. Zhou AH, Patel VR, Baredes S, Eloy JA, Hsueh WD. Mobile Applications for Allergic Rhinitis. Ann Otol Rhinol Laryngol. 2018;127(11):836-40.
79. Bousquet J, Arnavielhe S, Bedbrook A, Bewick M, Laune D, Mathieu-Dupas E, et al. MASK 2017: ARIA digitally-enabled, integrated, person-centred care for rhinitis and asthma multimorbidity using real-world-evidence. Clin Transl Allergy. 2018;8:45.
80. Bousquet J, Caimmi DP, Bedbrook A, Bewick M, Hellings PW, Devillier P, et al. Pilot study of mobile phone technology in allergic rhinitis in European countries: the MASK-rhinitis study. Allergy. 2017;72(6):857-65.
81. Caimmi D, Baiz N, Tanno LK, Demoly P, Arnavielhe S, Murray R, et al. Validation of the MASK-rhinitis visual analogue scale on smartphone screens to assess allergic rhinitis control. Clin Exp Allergy. 2017;47(12):1526-33.
82. Bieber T, Vieths S, Broich K. New opportunities and challenges in the assessment of drugs for atopic diseases. Allergy. 2016;71(12):1662-5.
83. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. Eur Respir J. 1999;14(4):902-7.
84. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol. 2004;113(1):59-65.
85. Kocks JWH, Seys SF, van Duin TS, Diamant Z, Tsiligianni IG. Assessing patient-reported outcomes in asthma and COPD patients: which can be recommended in clinical practice? Curr Opin Pulm Med. 2018;24(1):18-23.
86. Zicari AM, Occasi F, Montanari G, Indinnimeo L, De Castro G, Tancredi G, et al. Intranasal budesonide in children affected by persistent allergic rhinitis and its effect on nasal patency and Nasal Obstruction Symptom Evaluation (NOSE) score. Curr Med Res Opin. 2015;31(3):391-6.
87. Valero A, Mullol J, Herdman M, Rosales MJ, Spanish CQSG. Measuring outcomes in allergic rhinitis: psychometric characteristics of a Spanish version of the congestion quantifier seven-item test (CQ7). Health Qual Life Outcomes. 2011;9:14.
88. Wasserfallen JB, Gold K, Schulman KA, Baraniuk JN. Development and validation of a rhinoconjunctivitis and asthma symptom score for use as an outcome measure in clinical trials. J Allergy Clin Immunol. 1997;100(1):16-22.

89. Santanello NC, DeMuro-Mercon C, Shah SR, Schenkel EJ, Ratner PH, Dass SB, et al. Validation of the nighttime symptoms score as a clinically relevant measure of allergic rhinitis. Allergy Asthma Proc. 2006;27(3):231-9.
90. Revicki DA, Leidy NK, Brennan-Diemer F, Thompson C, Togias A. Development and preliminary validation of the multiattribute Rhinitis Symptom Utility Index. Qual Life Res. 1998;7(8):693-702.
91. Lo PS, Tong MC, Revicki DA, Lee CC, Woo JK, Lam HC, et al. Rhinitis Symptom Utility Index (RSUI) in Chinese subjects: a multiattribute patient-preference approach. Qual Life Res. 2006;15(5):877-87.
92. Franzke N, Schafer I, Jost K, Blome C, Rustenbach SJ, Reich K, et al. A new instrument for the assessment of patient-defined benefit in the treatment of allergic rhinitis. Allergy. 2011;66(5):665-70.
93. Johnson B, Harrington R, Perz C. Validation of the Scale of Allergic Rhinitis. Psychology, Health & Medicine. 2004;9(2):217-25.
94. Grouin JM, Vicaut E, Jean-Alphonse S, Demoly P, Wahn U, Didier A, et al. The average Adjusted Symptom Score, a new primary efficacy end-point for specific allergen immunotherapy trials. Clin Exp Allergy. 2011;41(9):1282-8.
95. Fonseca JA, Nogueira-Silva L, Morais-Almeida M, Sa-Sousa A, Azevedo LF, Ferreira J, et al. Control of Allergic Rhinitis and Asthma Test (CARAT) can be used to assess individual patients over time. Clin Transl Allergy. 2012;2(1):16.
96. van der Leeuw S, van der Molen T, Dekhuijzen PN, Fonseca JA, van Gemert FA, Gerth van Wijk R, et al. The minimal clinically important difference of the Control of Allergic Rhinitis and Asthma Test (CARAT): cross-cultural validation and relation with pollen counts. NPJ Prim Care Respir Med. 2015;25:14107.
97. Borrego LM, Fonseca JA, Pereira AM, Pinto VR, Linhares D, Morais-Almeida M. Development process and cognitive testing of CARATkids - Control of Allergic Rhinitis and Asthma Test for children. BMC Pediatr. 2014;14:34.
98. Amaral R, Carneiro AC, Wandalsen G, Fonseca JA, Sole D. Control of Allergic Rhinitis and Asthma Test for Children (CARATKids): Validation in Brazil and cutoff values. Ann Allergy Asthma Immunol. 2017;118(5):551-6 e2.
99. Emons JA, Flokstra BM, de Jong C, van der Molen T, Brand HK, Arends NJ, et al. Use of the Control of Allergic Rhinitis and Asthma Test (CARATkids) in children and adolescents: Validation in Dutch. Pediatr Allergy Immunol. 2017;28(2):185-90.
100. Schatz M, Meltzer EO, Nathan R, Derebery MJ, Mintz M, Stanford RH, et al. Psychometric validation of the rhinitis control assessment test: a brief patient-completed instrument for evaluating rhinitis symptom control. Ann Allergy Asthma Immunol. 2010;104(2):118-24.
101. Schatz M, Zeiger RS, Chen W, Yang SJ, Stanford RH, Garris CP. A comparison of the psychometric properties of the Mini-Rhinitis Quality of Life Questionnaire and the Rhinitis Control Assessment Test. Am J Rhinol Allergy. 2012;26(2):127-33.
102. Meltzer EO, Schatz M, Nathan R, Garris C, Stanford RH, Kosinski M. Reliability, validity, and responsiveness of the Rhinitis Control Assessment Test in patients with rhinitis. J Allergy Clin Immunol. 2013;131(2):379-86.
103. Fernandes PH, Matsumoto F, Sole D, Wandalsen GF. Translation into Portuguese and validation of the Rhinitis Control Assessment Test (RCAT) questionnaire. Braz J Otorhinolaryngol. 2016;82(6):674-9.
104. Demoly P, Bousquet PJ, Mesbah K, Bousquet J, Devillier P. Visual analogue scale in patients treated for allergic rhinitis: an observational prospective study in primary care: asthma and rhinitis. Clin Exp Allergy. 2013;43(8):881-8.
105. Leong KP, Chan SP, Tang CY, Yeak SC, Saurajen AS, Mok PK, et al. Quality of life of patients with perennial allergic rhinitis: preliminary validation of the Rhinoconjunctivitis Quality of Life Questionnaire in Singapore. Asian Pac J Allergy Immunol. 1999;17(3):163-7.
106. Leong KP, Yeak SC, Saurajen AS, Mok PK, Earnest A, Siow JK, et al. Why generic and disease-specific quality-of-life instruments should be used together for the evaluation of patients with persistent allergic rhinitis. Clin Exp Allergy. 2005;35(3):288-98.
107. Okuda M, Ohkubo K, Goto M, Okamoto H, Konno A, Baba K, et al. Comparative study of two Japanese rhinoconjunctivitis quality-of-life questionnaires. Acta Otolaryngol. 2005;125(7):736-44.
108. Yuksel H, Yilmaz O, Alkan S, Bayrak Degirmenci P, Kirmaz C. Validity and reliability of Turkish version of rhinitis and mini-rhinitis quality of life questionnaires. Allergol Immunopathol (Madr). 2009;37(6):293-7.
109. Braido F, Baiardini I, Stagi E, Scichilone N, Rossi O, Lombardi C, et al. RhinAsthma patient perspective: a short daily asthma and rhinitis QoL assessment. Allergy. 2012;67(11):1443-50.

110. Kiotseridis H, Cilio CM, Bjermer L, Aurivillius M, Jacobsson H, Tunsater A. Swedish translation and validation of the Pediatric Allergic Disease Quality of Life Questionnaire (PADQLQ). Acta Paediatr. 2011;100(2):242-7.
111. Juniper EF, Thompson AK, Ferrie PJ, Roberts JN. Validation of the standardized version of the Rhinoconjunctivitis Quality of Life Questionnaire. J Allergy Clin Immunol. 1999;104(2 Pt 1):364-9.
112. Juniper EF, Rohrbaugh T, Meltzer EO. A questionnaire to measure quality of life in adults with nocturnal allergic rhinoconjunctivitis. J Allergy Clin Immunol. 2003;111(3):484-90.
113. Juniper EF, Howland WC, Roberts NB, Thompson AK, King DR. Measuring quality of life in children with rhinoconjunctivitis. J Allergy Clin Immunol. 1998;101(2 Pt 1):163-70.
114. Yuksel H, Yilmaz O, Sogut A, Eser E. Validation and reliability study of the Turkish version of the Pediatric Rhinitis Quality of Life Questionnaire. Turk J Pediatr. 2009;51(4):361-6.
115. Mavroudi A, Chrysochoou EA, Boyle RJ, Papastergiopulos A, Karantaglis N, Karagiannidou A, et al. Validation study of the pediatric allergic rhinitis quality of life questionnaire. Asian Pac J Allergy Immunol. 2016;34(2):159-65.
116. Nascimento Silva M, Naspitz C, Sole D. Evaluation of quality of life in children and teenagers with allergic rhinitis: adaptation and validation of the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ). Allergol Immunopathol (Madr). 2001;29(4):111-8.
117. Larrosa F, Roura J, Dura MJ, Guirao M, Alberti A, Alobid I. Adaptation and validation of the Spanish version of the Nasal Obstruction Symptom Evaluation (NOSE) Scale. Rhinology. 2015;53(2):176-80.
118. van Zijl F, Timman R, Datema FR. Adaptation and validation of the Dutch version of the nasal obstruction symptom evaluation (NOSE) scale. Eur Arch Otorhinolaryngol. 2017;274(6):2469-76.
119. Marro M, Mondina M, Stoll D, de Gabory L. French validation of the NOSE and RhinoQOL questionnaires in the management of nasal obstruction. Otolaryngol Head Neck Surg. 2011;144(6):988-93.
120. Karahatay S, Tasli H, Karakoc O, Aydin U, Turker T. Reliability and validity of the Turkish Nose Obstruction Symptom Evaluation (NOSE) scale. Turk J Med Sci. 2018;48(2):212-6.
121. Onerci Celebi O, Araz Server E, Yigit O, Longur ES. Adaptation and validation of the Turkish version of the Nasal Obstruction Symptom Evaluation scale. Int Forum Allergy Rhinol. 2018;8(1):72-6.
122. Bezerra TF, Padua FG, Pilan RR, Stewart MG, Voegels RL. Cross-cultural adaptation and validation of a quality of life questionnaire: the Nasal Obstruction Symptom Evaluation questionnaire. Rhinology. 2011;49(2):227-31.
123. Amer MA, Kabbash IA, Younes A, Elzayat S, Tomoum MO. Validation and cross-cultural adaptation of the arabic version of the nasal obstruction symptom evaluation scale. Laryngoscope. 2017;127(11):2455-9.
124. Kanatani KT, Slingsby BT, Mukaida K, Kitano H, Adachi Y, Haefner D, et al. Translation and Linguistic Validation of the Allergy-CONTROL-Score(TM) for Use in Japan. Allergol Int. 2013;62(3):337-41.
125. Flokstra-de Blok BMJ, Baretta HJ, Fonseca JA, van Heijst E, Kollen BJ, de Kroon J, et al. Control of Allergic Rhinitis and Asthma Test with 1-week recall: Validation of paper and electronic version. Allergy. 2018.

Figure legends.

Figure 1. The main measurement properties of instruments for the evaluation of health-related patient-reported outcomes. MCID: minimal clinically important difference.

Figure 2. Study flow chart.

Figure 1. The main measurement properties of instruments for the evaluation of health-related patient-reported outcomes. MCID: minimal clinically important difference.

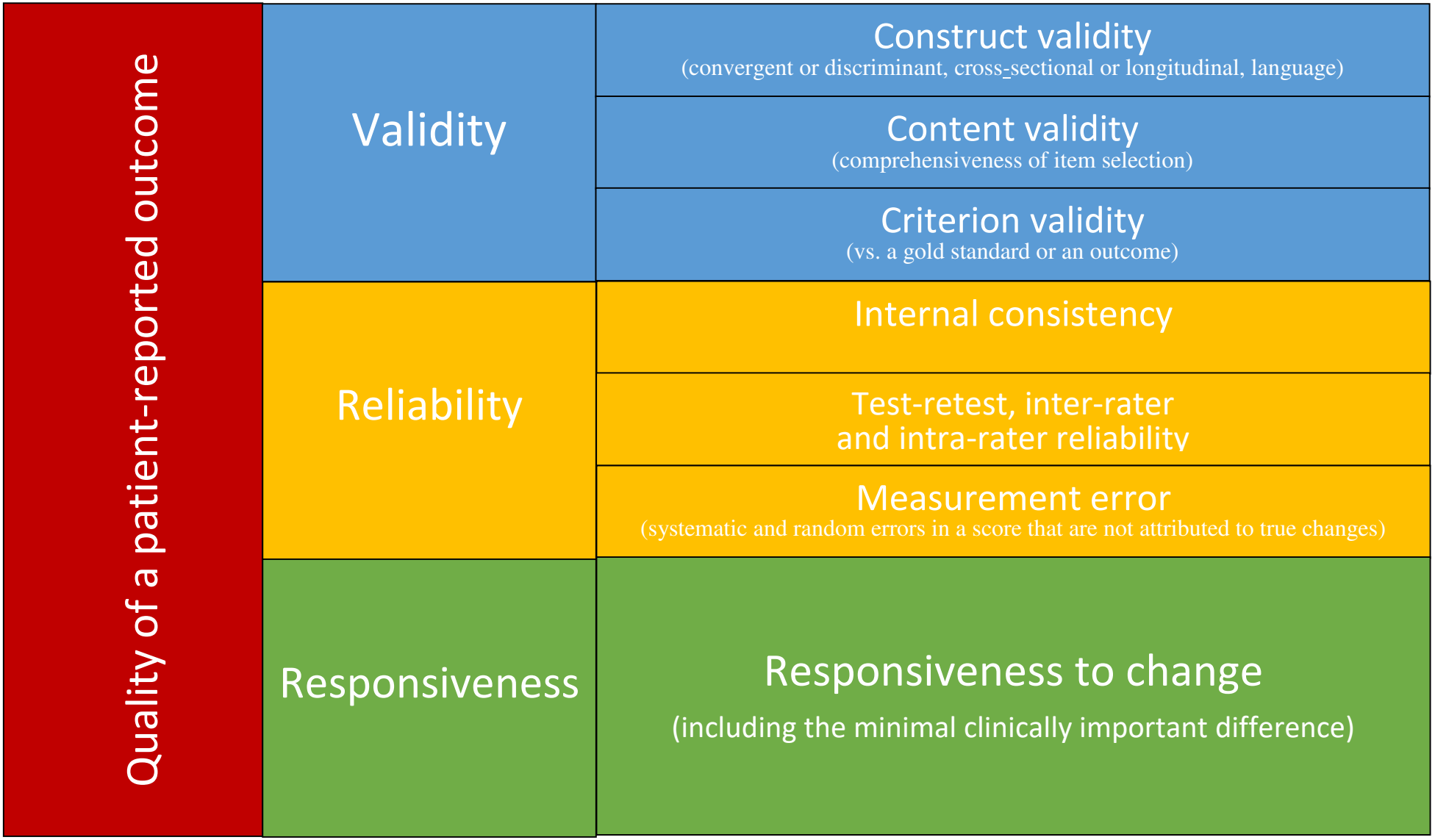

Figure 2. Study flow chart.

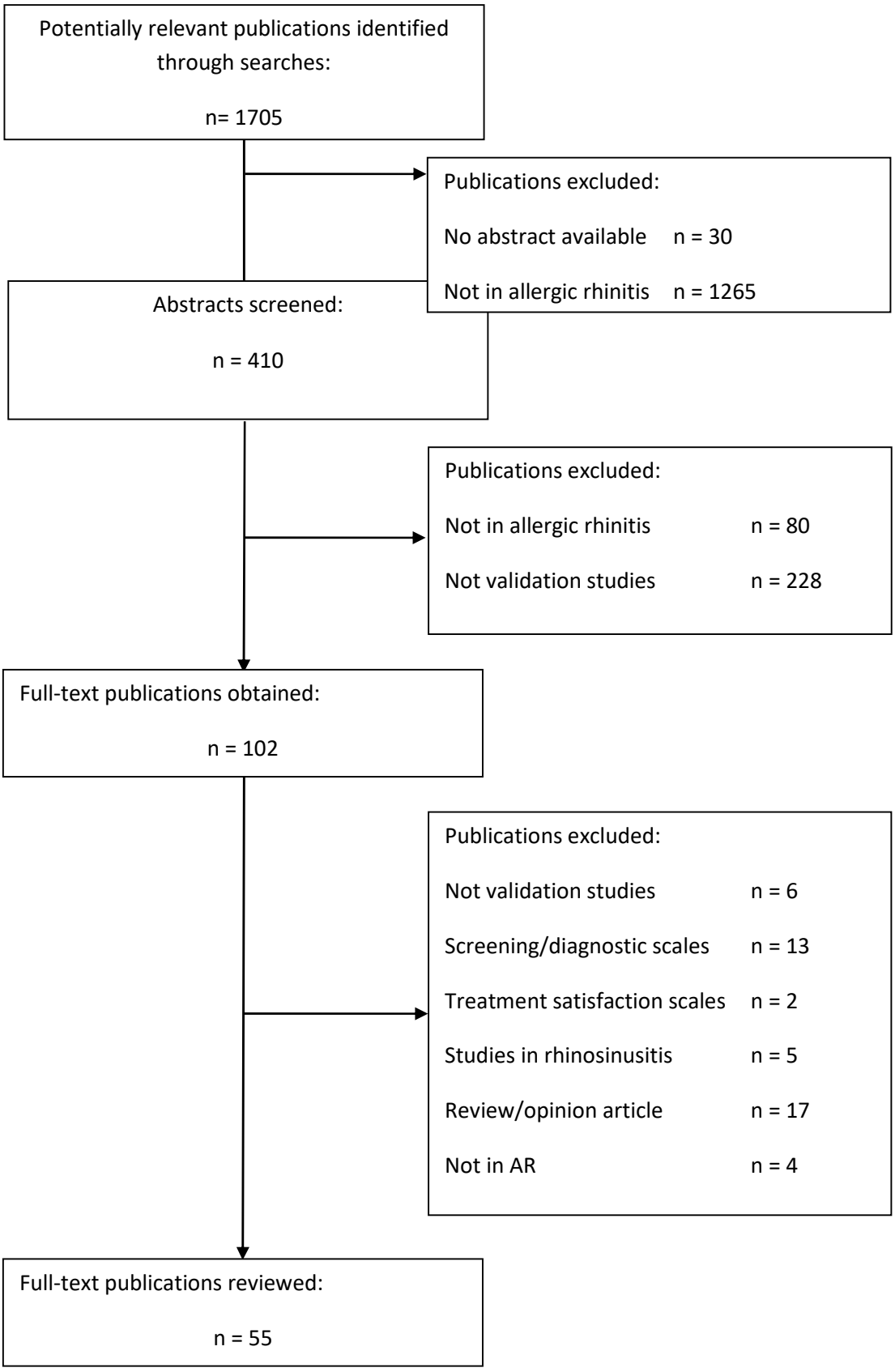

Table E1. Summary of publications on the validation of symptom scores, disease control scores and QoL scores in AR.

First author and year of publication	Name and type of score	Age class of study subjects.	Psychometric validity			Reliability *(inter-/intra-rater, test-retest, or internal consistency.)*	Responsiveness *(ability to detect change)*	MCID		Other quality indicators
			Construct validity *(convergent or discriminant, cross-sectional or longitudinal, language)*	Content validity *(comprehensiveness of item selection)*	Criterion validity *(concurrent or predictive, vs. a gold standard or an outcome)*			Anchor-based	SD-based	
Wasserfallen 1997 (88)	RASS (symptom score)	adults	✓	✓	–	✓ Test-rest reliability was 0.8 at 1 week, and internal consistency was 0.8 for individual organs and 0.7 for individual symptoms.	✓	–	–	Prediction of performance after a change in the number of items; appropriateness of the data distribution and the statistical tests applied.
Juniper 1998 (113)	PRQLQ (QoL score)	children	✓	–	–	✓ Overall intraclass correlation coefficient for test-retest reliability. = 0.93	✓	–	–	–

Revicki 1998 (90)	RSUI (symptom score)	adults	✓	–	✓	–	–	–	–	–
Juniper 1999 (111)	RQLQ(S) (QoL score)	adults	✓ vs. diary symptoms, SF-36, "feeling thermometer" patient preference, "standard gamble" utility	–	–	✓ Intraclass correlation coefficient for test-retest reliability = 0.97	✓ $p<0.001$	✓ MID = 0.48 ± 0.93, against a GRCS	–	–
Kemker 1999 (75)	AOS (hybrid disease control, symptom, outcome score)	not stated	✓	–	–	✓	✓	–	–	–
Leong 1999 (105)	RQLQ (QoL score) *administered to English-speaking patients in Singapore*	adults	✓	✓	–	–	✓	–	–	Identification of candidates for scale-shortening
Juniper 2000 (66)	mini-RQLQ (QoL score)	adults	✓	✓	✓ against the RQLQ	✓	✓	✓ MID = 0.70 ± 1.15, against a GRCS	–	–

Nascimento Silva 2001 (116)	modified RQLQ (QoL score)	children and teens	✓	✓	–	✓	✓	–	–	–
Baiardini 2003 (74)	RHINASTHMA (QoL score)	adults	✓	–	✓	✓ Test–retest: Pearson coefficient between 0.70 and 0.92 for 25 items and between 0.61 and 0.69 for five items.	✓	–	–	–
Juniper 2003 (112)	NRQLQ (QoL score)	adults	✓	–	–	✓	✓	–	–	–
Roberts 2003 (72)	PADQLQ (QoL score)	children and teens	✓	–	–	✓	✓	✓ MID [95%CI] = 0.33 [0.11–0.54], against a VAS score	–	–
Johnson 2004 (93)	SoAR (symptom score)	adults	✓	–	–	✓	–	–	–	–
Stewart 2004 (53)	NOSE (symptom score) *in adults* (has also been validated in Spanish (117),	adults	✓	✓	✓	✓	✓	–	–	–

	Dutch (118), French (119), Italian, Turkish (120, 121), Portuguese (122) and Arabic (123))									
Leong 2005 (106)	RQLQ *(vs. SF-36 in patients with persistent AR)*	adults	✓	–	–	✓	✓	–	–	–
Okuda 2005 (107)	RQLQ *in Japanese* and JRQLQ (QoL scores)	adults	✓	–	–	✓	–	–	–	–
Lo 2006 (91)	RSUI *in Chinese patients* (symptom score)	adults	✓	–	–	✓	–	–	–	–
Santanello 2006 (89)	NtSS (symptom score)	teens and adults	✓	–	–	✓	✓	–	–	–
Esprint Study Investigators 2007 (73)	ESPRINT (QoL score)	adults	✓	✓	–	✓	✓	–	–	–
Stull 2007 (54)	CQ7 (symptom score)	adults	✓	✓	–	✓	✓	–	–	–

AbuRuz 2009 (71)	AAQQ (QoL score)	adults	✓	✓	-	✓	✓	✓ MID ± SD = 1.1 ± 1.0 against a GRCS score	-	-
Nogueira-Silva 2009 (60)	CARAT, 17 questions (disease control score)	adults	✓	✓	-	-	-	-	-	-
Yüksel 2009 (108)	RQLQ and mini-RQLQ *in Turkish* (QoL scores)	adults	✓	-	-	✓	-	-	-	-
Yüksel 2009 (114)	PRQLQ in Turkish (QoL score)	children and teens	✓	-	-	✓	-	-	-	-
Barnes 2010 (36)	mini-RQLQ (QoL score), PNIF (physical measurement) and T4NSS (symptom score).	adults	-	-	-	-	-	✓ Direct anchor-based: 0.15 [-0.13 – 0.44]; Direct anchor-based by regression: 0.23 [0.03 – 0.44]: indirect anchor-based by regression: 0.42 units [0.30 – 0.51]	✓ Distribution -based: 0.24 [0.21 – 0.27];	-

Fonseca 2010 (61)	CARAT10 *to assess rhinitis and asthma in patients with asthma* (disease control score)	adults	-	-	✓ *stated as concurrent validity, vs. physician's assessment, VAS of symptoms; and ACQ5 score. Probably construct validity, rather than criterion validity.*	-	-	-	-	-
Schatz 2010 (100)	RCAT (disease control score)	adults	✓	-	-	✓	-	-	-	-
Stull 2010 (55)	CQ5 (symptom score)	adults	✓	✓	-	✓	-	-	-	Determination of the cut-off for detecting congestion.
Demoly 2011 (62)	ARCT (disease control score)	teens and adults	✓	✓	-	✓	✓	-	-	Determination of the cut-off for detecting control. Sp, Se, NPV and PPV.
Franzke 2011 (92)	PBI-AR (symptom score)	adults	✓	-	-	✓	-	-	-	Acceptability and feasibility in clinical routine

Grouin 2011 (94)	AdSS (symptom score)	children, teens and adults	✓ *vs. RTSS and a combined score*	–	–	–	–	–	–	–
Hafner 2011 (56)	ACS (a CSMS)	adults	✓	–	–	✓	–	–	–	Feasibility
Kiotseridis 2011 (110)	PADQLQ (QoL score)	children and teens	✓	–	–	✓	✓	–	✓ 0.5 SD = 0.42	–
Valero 2011 (87)	CQ7 *in Spanish* (symptom score)	adults	✓	–	–	✓	✓	–	–	Sp, Se, cut-off
Braido 2012 (109)	RAPP (QoL score)	teens and adults	✓	✓	✓ *against RHINASTHMA*	✓	✓	–	✓ *cut-off of 2*	–
Fonseca 2012 (95)	CARAT10 (disease control score)	adults	✓	–	–	✓	✓	–	–	–

Hafner 2012 (57)	ACS for the eyes and nose only (RC-ACS, a CSMS)	adults	✓	–	–	✓	–	–	–	–
Schatz 2012 (101)	mini-RQLQ (QoL score)	adults	–	✓	–	✓	–	–	–	–
	RCAT (disease control score)	adults	–	✓	–	✓	–	✓ 2.6 for the RCAT	–	–
Demoly 2013 (104)	VAS for the burden of AR in primary care (disease control score)	adults	✓	–	–	–	✓	–	–	–
Higaki 2013 (52)	a T5SS *in Japanese cedar/cypress pollinosis* (symptom score)	adults	–	–	–	–	–	✓ 1.426 in 2009 and 1.441 in 2010	–	–
	a T6SS *in Japanese cedar/cypress pollinosis*	adults	–	–	–	–	–	✓ 4.115 in 2009	–	–

	(symptom score)							and 3.183 in 2010		
	JRQLQ, 17-item, *in Japanese cedar/cypress pollinosis* (QoL score)	adults	-	-	-	-	-	✓ 10.469 in 2009 and 6.026 in 2010	-	-
Kanatani 2013 (124)	ACS *in Japanese* (a CSMS)	teens and adults	-	✓	-	-	-	-	-	Accessibility and acceptability
Meltzer 2013 (102)	RCAT *in rhinitis* (disease control score)	teens and adults	✓	-	-	✓	-	✓ Mean of 14 measurements: 2.4	✓ SEM = 2.1, 0.5 SD = 2.2	Screening accuracy
Borrego 2014 (97)	CARATkids (disease control score)	children	-	✓	-	-	-	-	-	Cognitive testing
Devillier 2014 (50)	RTSS (symptom score)	children, teens and adults	-	-	-	-	-	✓ 1.24 ± 0.17 (vs. GRCS) and 1.12 ± 0.14 (vs. RQLQ) in children, 1.33 ± 0.14 and 1.20 ± 0.13 in adolescents;	✓ 0.33 SD = from 1.09 to 1.13 ; 0.5 SD = from 1.22 to 1.40	-

								and 1.13 ± 0.14 and 0.89 ± 0.12 in adults		
Linhares 2014 (47)	CARATkids (disease control score)	children	✓	✓	-	✓	✓	-	-	-
Wang 2014 (64)	ARCT *in Chinese* (disease control score)	adults	✓	-	-	-	-	-	-	-
Demoly 2015 (67)	PBI-AR (symptom score)	adults	✓	-	✓ against a discomfort VAS and the miniRQLQ	✓	-	-	✓ 0.5 SD = 0.4	-
van der Leeuw 2015 (96)	CARAT10 *in Dutch* (disease control score)	adults	✓	-	-	✓	-	✓ 3.50 against the GRCS	✓ SEM = 2.83	-
Zicari 2015 (86)	NOSE *in children* (symptom score)	children	✓	-	-	-	-	-	-	-

Boulay 2016 (58)	RCSS (disease control score)	adults	✔ against RQLQ(S), VAS and TNSS-TRSS	✔	✔	✔	-	-	-	-
Fernandes 2016 (103)	RCAT *in Brazilian Portuguese* (disease control score)	teens	✔	-	-	✔	-	-	-	Sensitivity and specificity
Mavroudi 2016 (115)	Ped-AR-QoL (QoL score)	children and teens	✔ With the Disabkids paediatric QoL scale	-	-	✔	-	-	-	-
Wang 2016	ARCT *in Chinese* (disease control score)	children, teens and adults	-	-	-	-	✔	-	-	-
Amaral 2017 (98)	CARATKids *in Brazilian Portuguese* (disease control score)	children	✔	-	-	✔	✔	-	-	Sensitivity and specificity

Caimmi 2017 (81)	Four VASs *measuring control of overall, nasal, ocular, and asthma symptoms* (disease control score)	not stated	–	–	–	✓	–	–	–	Sensitivity and acceptability
Emons 2017 (99)	CARATKids *in Dutch* (disease control score)	children and teens	✓	–	–	✓	✓	✓ GRCS: 2.76; ACT: 2.77	✓ 1.96 SEM = 2.80	–
Molinengo 2017 (70)	RAPP (QoL score)	adults	✓	✓	–	–	–	–	–	–
Liedtke 2017 (59)	RCAT (disease control score)	adults	✓	–	–	–	–	–	–	–
Zhu 2018 (65)	ARCT *in Chinese* (disease control score)	adults	✓	–	–	–	✓	–	–	–
Flokstra-de Blok 2018 (125)	CARAT *paper and electronic versions* (disease control score)	adults	✓	–	–	✓	–	–	–	–

Age classes of study subjects are defined as children: younger than 13; teens: aged 13 to 17; adults: aged 18 and over.

Abbreviations (in alphabetical order):

AAQQ: Arabic Allergic Rhinitis Quality of Life Questionnaire

ACS: Allergy-Control-SCORE

AdSS: Adjusted Symptom Score

AOS: Allergy Outcome Survey

ARCT: Allergic Rhinitis Control Test

ARSQOL: Allergic Rhinitis-Specific Quality of Life

CARAT10: Control of Allergic Rhinitis and Asthma Test

CQ5: Congestion Quantifier Five-Item Screener

CQ7: Congestion Quantifier Seven-Item Test

ESPRINT: Cuestionario ESPañol de Calidad de Vida en RINiTis

JRQLQ: Japan Rhinoconjunctivitis Quality of Life Questionnaire

MCID: minimal clinically important difference.

MID: minimal important difference.

mini-RQLQ: Mini Rhinoconjunctivitis Quality of Life Questionnaire

NtSS: Nighttime Symptoms Score

NOSE: Nasal Obstruction Symptom Evaluation

NRQLQ: Nocturnal Rhinoconjunctivitis Quality of Life Questionnaire

PADQLQ: Pediatric Allergic Disease Quality of Life Questionnaire

PBI-AR: Patient Benefit Index - Allergic Rhinitis;

Ped-AR-QoL: Paediatric Allergic Rhinitis Quality of Life Questionnaire

PNIF: peak nasal inspiratory flow

PRQLQ: Paediatric Rhinoconjunctivitis Quality of Life Questionnaire

RAPP: RhinAsthma Patient Perspective

RASS: a rhinoconjunctivitis and asthma symptom score

RCAT: Rhinitis Control Assessment Test

RCSS: Rhinitis Control Scoring System

RHINASTHMA: a disease-specific QoL questionnaire

RQLQ(S): Standardized Rhinoconjunctivitis Quality of Life Questionnaire

RSUI: Rhinitis Symptom Utility Index

RTSS: Rhinoconjunctivitis Total Symptom Score

SoAR: Scale of Allergic Rhinitis

T4NSS: 4-item total nasal symptom score.

T5NSS: 5-item total nasal symptom score.