VALUE IN HEALTH 16 (2013) 987–992

Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

The Development and Validation of a Multidimensional Sum-Scaling Questionnaire to Measure Patient-Reported Outcomes in Acute Respiratory Tract Infections in Primary Care: The Acute Respiratory Tract Infection Questionnaire

Rune Aabenhus, MD, GP, Hanne Thorsen, MD, PhD, Volkert Siersma, PhD, John Brodersen, MD, GP, PhD*

The Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Objective: Patient-reported outcomes are seldom validated measures in clinical trials of acute respiratory tract infections (ARTIs) in primary care. We developed and validated a patient-reported outcome sum-scaling measure to assess the severity and functional impacts of ARTIs. **Methods:** Qualitative interviews and field testing among adults with an ARTI were conducted to ascertain a high degree of face and content validity of the questionnaire. Subsequently, a draft version of the Acute Respiratory Tract Infection Questionnaire (ARTIQ) was statistically validated by using the partial credit Rasch model to test dimensionality, objectivity, and reliability of items. Test of known groups' validity was conducted by comparing participants with and without an ARTI. **Results:** The final version of the ARTIQ consisted of 38 items covering five dimensions (Physical-upper, Physical-lower, Psychological, Sleep, and Medicine) and five single items. All final dimensions were confirmed to fit the Rasch model, thus enabling sum-scaling of responses. The ARTIQ scores in participants with an ARTI were significantly higher than in those without ARTI (known groups' validity). **Conclusion:** A self-administered, multidimensional, sum-scaling questionnaire with high face and content validity and adequate psychometric properties for assessing severity and functional impacts from ARTIs in adults is available to clinical trials and audits in primary care.

Keywords: acute respiratory tract infections, patient-reported outcome, questionnaire, Rasch analysis.

Introduction

Acute respiratory tract infections (ARTIs), which are often divided into upper– and lower–respiratory tract infections, are among the most frequent diseases seen by general practitioners (GPs) [1,2]. The substantial symptomatic and functional impairment caused by these infections constitutes a major public health problem [2].

Symptoms in ARTIs are diverse and often contain elements from the entire respiratory tract. Even individuals infected with the same viral strain display a striking variance in symptoms and the presentation is linked to illness duration [3]. Patient symptoms and clinical signs are often not sensitive enough to discriminate between the different types of ARTIs (such as acute bronchitis from pneumonia). Because of this diagnostic uncertainty, diagnoses in primary care may not always reflect the explicit pathophysiological criteria commonly applied in medical science [4–6]. Hence, the terms rhinitis, sinusitis, and bronchitis, among others, at best indicate the anatomic site most affected at the time of consultation. Accordingly, in a primary care setting, a precise diagnosis is often not possible and symptom-based criteria (e.g., acute cough) are now frequently used as inclusion criteria in pragmatic clinical trials [7,8].

In clinical conditions with no accepted "gold standard," the assessment of new treatment and diagnostic modalities may be better addressed by using a patient-oriented approach [6,9], bypassing the medically defined "gold standard." When assessing treatment effects, cost, and so on of ARTIs in a primary care setting, the vast majority of patients will recover uneventfully with no or few hard end points such as mortality, highlighting the need to measure the direct impact of disease on a patient's daily life. Furthermore, a change in a medical parameter, such as auscultatory abnormalities or the normalization of a C-reactive protein value, may only modestly reflect the patient's own experience of the illness and does not encompass any associated symptoms (e.g., a cough that impacts daily activities or sleep) or

Conflict of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Research Unit for General Practice, University of Copenhagen, Øster Farimagsgade 5, Box 2099, Copenhagen DK -1014 CPH, Denmark.

E-mail: runeaa@sund.ku.dk.

1098-3015/$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2013.06.011

psychological components (e.g., that individual patients experience different effects on health status despite equal physiological limitations) [10]. This is in line with the recognition that patient-reported outcomes (PROs) are a key component in evaluating health outcomes [11–14].

Several PRO measures have been developed to assess disease-specific conditions such as asthma, pneumonia, chronic obstructive pulmonary disease, and the common cold [15–18] and are valuable in controlled trials with access to further diagnostic workup to ensure certainty in the diagnosis. The instruments available, however, use a single sum score of all parameters: a global score. It can be argued that when assessing different aspects of a disease as experienced by a patient, that is, a biopsychosocial model, the name itself indicates a number of different components that are not directly comparable. Accordingly, a multidimensional approach with several scales may be more suitable. To our knowledge, pragmatic trials intended to reflect everyday routine in primary care treatment of unspecific ARTIs do not have a qualitatively developed, psychometrically validated sum-scaling instrument to measure symptomatic severity and functional impacts in ARTIs *en bloc*.

The aim of this study was to develop and validate a PRO questionnaire with high face and content validity and adequate psychometric properties to measure the severity and functional impact of an ARTI in adults to determine the impact of the infection itself and to evaluate the success or failure of therapeutic trials in primary care.

Methods

The study was composed of 1) a literature and qualitative study to ensure face and content validity of the PRO (phase 1 and 2, respectively) and 2) a test of the PROs' psychometric properties (phase 3).

Phase 1: Item Generation

A list of all domains (physical, social, psychological, sleep, etc.) that might be affected during an ARTI and their corresponding symptoms was generated by reviewing other questionnaires covering specific ARTI diseases, looking at guidelines, and consulting colleagues. A first draft version of the questionnaire was constructed.

Phase 2: Face and Content Validity

Interviews

The qualitative study aimed to expand our knowledge of the symptomatic and functional impact of an ARTI. The participating GPs identified patients who agreed to participate and scheduled an interview no later than 48 hours after inclusion. The participants were assigned to focus group or single interviews to ensure adequate concept elicitation and cognitive assessment. The goals of this phase were 1) to elicit participants' experiences with ARTIs for item generation and 2) to verify the comprehensiveness and patient understanding of the questionnaire. We choose to conduct both single and focus group interviews to secure different views and dynamics in generating input on the questionnaire. All interviews took place at the Research Unit for General Practice. No payment was given to interviewees.

Single interviews

The single interviews lasted approximately 30 minutes. The purpose of these interviews was to understand which dimensions of the subjects' lives were affected and how this was experienced. During the interview, the subject completed the draft version of the questionnaire and was interviewed item by item to ascertain relevance and comprehension.

Focus group interviews

The group interviews lasted approximately 2 hours and were audio-recorded. The focus group first engaged in an open-ended discussion about symptoms and functional impairments in an ARTI. Then, the participants completed the draft version of the questionnaire and discussed the instructions, item wording, and ease of completion.

Timeline of qualitative study

We had a focus group session followed by single interviews. This procedure was then repeated. The first half of the qualitative study was mainly on concept elicitation; the second part focused more on cognitive assessment of the questionnaire, the corresponding wording of items, and response categories.

Analysis of interviews (focus and single)

R.A. conducted all interviews. H.T. was a comoderator in the focus groups. After each interview, R.A. and H.T. analyzed data and the draft version of the questionnaire was changed according to relevant input. The draft version of the questionnaire was successively used in the next interviews until further interviews gave no additional information, and the participants completed the questionnaire without problems or misunderstandings of the wording or meaning [19]. Items were worded according to participants' suggestions.

Phase 3: Psychometric Properties

Recruitment and participants

The psychometric properties of the final version of the questionnaire were tested in subjects with and without an ARTI who consulted their GP. Fifteen GPs were invited to enroll 20 subjects with and 20 subjects without an ARTI (scheduled for consultations and without complaints of fever or ARTI symptoms). All ARTI diagnoses were made solely on clinical grounds. Subjects were excluded if they were younger than 18 years or unable to communicate in Danish.

Statistical Analyses

Rasch analysis

Many PROs consist of ordinal response categories measuring the severity of the latent health trait (e.g., functional ability, pain, or happiness) as perceived by the respondent. When summing raw scores of items in a scale, however, an assumption of unidimensionality is made, that is, that the items in the questionnaire describe different aspects of the same construct (dimension), allowing raw scores of the items to be added together [20,21]. Rasch models are tests for unidimensionality in the sense that they investigate whether responses from persons can be listed according to item difficulty, from lowest to highest, in a probabilistic Guttmann pattern [22,23]. Items with poor fit to a Rasch model may be excluded in an iterative process, with one item removed at a time and model fit reestimated accordingly. When items fit a Rasch model, objective (invariant) measurement is obtained. Moreover, Rasch models can also test whether one or more items possess uniform or nonuniform differential item functioning (DIF), indicating that an item has unequal function in different subgroups, for example, sex, age, and diagnostic groups, and thereby disturbs the invariant measurement [23]. They also estimate whether response categories function as intended [24,25], that is, that an ordinal ranking of the response categories is obtained.

The pairwise estimation procedure implemented in the software program RUMM2030 was used to estimate item parameters in the partial credit Rasch model for dichotomous and polytomous items [26]. The overall fit of the model and scales was assessed by

using Wright-Panchapakesan chi-square statistics [27], while item fit statistics were used to identify poorly fitting items. To adjust for multiple testing, the significance level of the Wright-Panchapakesan statistics was set to 0.01 when determining unidimensionality. Reliability and capacity of the scale to differentiate between patients were assessed by using Cronbach's alpha and the Person Separation Index, respectively [28].

Known groups' validity

It provides credibility to construct validity when the PRO can discriminate between a group of individuals known to have a particular trait and a group of individuals who do not have that trait.

Analyses are presented as median (interquartile range) and n (%) as appropriate. Differences in subject characteristics and ARTIQ scores of single items and ARTIQ sum-scores of dimensions between the ARTI group and the noninfected control group (known groups' validity) were assessed by using a nonparametric Wilcoxon signed-rank test. A P value of <0.05 was considered significant.

The study was approved by the Danish Data Protection Agency.

Results

Phase 1: Item Generation

A first draft of the questionnaire encompassed 48 items: 38 that graded severity in four response categories and 10 dichotomous items.

Phase 2: Face and Content Validity

Two focus group sessions (each with three participants) and 12 single interviews formed the basis of the qualitative study (Table 1). Based on the information provided, 46 items were identified in six domains: Physical (23 items), Psychological (5 items), Energy (3 items), Sleep (4 items), Medicine (9 items), and Social (2 items) (Table 2). The domain Energy, as identified by participants, included questions that in medical terms could be regarded as physical symptoms, such as item 26 "Felt dizzy." The lay word "dizzy" in this context, however, should not be mistaken for the medical term "vertigo," but rather a vague description of feeling unwell. Response categories in the ARTIQ were readjusted to a three-point scale phrased as follows: *No, Yes—some,* and *Yes—a lot*. Ten dichotomized items ("Yes", "No") (items 36–46) were maintained.

Phase 3: Psychometric Properties

Recruitment and participants

GPs each recruited 5 to 40 participants; 274 participants completed the final draft version of the questionnaire; 13 (5%) were excluded because of missing data and 2 in accordance with exclusion criteria (age < 18 years), leaving 259 participants (Table 1).

Rasch analysis

Four domains (Physical, Psychological, Sleep, and Medicine) were subjected to Rasch modeling; the remaining two domains (Energy and Social) were not considered because of the low number of items.

Physical Items. A score summing all the 23 Physical domain items did not fit the Rasch model. Prespecified subgroup analysis of the Physical domain identified two subdomains, namely, Physical-Upper respiratory tract (PhysUP) and Physical-Lower respiratory tract (PhysLOW) (Table 2). Of the 15 items covering PhysUP, 13 items fitted the Rasch model (Tables 2 and 3). Three PhysUP items possessed uniform DIF in relation to participants with and without an ARTI. Because of the different directions and magnitude of the DIF in these three items, the DIF leveled out. Among the two poorly fitting items (item 15 "Muscle pains" and item 16 "Joint pains"), the face validity was low for "Joint pains" and hence this item was excluded (Tables 2 and 4). Item 15 "Muscle pains" was retained as a single item because of its high content validity and its ability to discriminate between those with and without an ARTI (Tables 2 and 4). The eight PhysLOW items fitted a unidimensional model, and none possessed DIF (Tables 2 and 3).

Sleep. All four items in this dimension fitted the Rasch model, and none possessed DIF (Tables 2 and 3).

Psychological. This five-item domain was unidimensional (Tables 2 and 3). Item 28 "Not feeling yourself" possessed uniform DIF relative to an ARTI. Nevertheless, the item was maintained because of its content validity and reliability.

Medicine. The nine items did not initially exhibit unidimensionality. The poor fit was due to item 37 "Taken painkillers." Because paracetamol, nonsteroidal anti-inflammatory drugs, and acetylsalicylic acid are over-the-counter drugs in Denmark, and because these drugs have both analgesic and fever-reducing properties, it may be difficult to distinguish between item 36 "Taken medicine to reduce fever" and item 37 "Taken painkillers." Although the face validity was high, item 37 was excluded because of this possible ambiguity. The remaining eight items fitted the Rasch model, and no items possessed DIF (Tables 2 and 3).

No DIF was identified relative to gender in any of the 46 items.

In four PhysUP items (items 7, 14, 18, and 19) and item 11 in PhysLOW, the response categories did not function as intended. Items with inadequately functioning response categories were rescored by collapsing the responses "Yes —some" with "Yes—a lot." After rescoring, the overall fit of the PhysUP domain and the PhysLOW domain to the Rasch model increased.

The overall fit statistics of the five ARTIQ domains are presented in Table 3, including adequate values of reliability: all Wright-Panchapakesan statistics were insignificant; Chronbach's alpha was >0.8 in all dimensions except medicine; and Person

Table 1 – Demographic data of participants.

Data collection	No.	ARTI +	Age (y), mean (range)	Sex, female/male
Cognitive interviews	12	12	45 (23–76)	4/8
Focus groups	6	6	47 (27–73)	3/3
Validation study	259	122	41 (18–77)	166/91

ARTI +, acute respiratory tract infection according to treating general practitioner.

Table 2 – Characteristics of items included or excluded in the ARTIQ.

No. Item	Initial domain	Chi-square	Probability of fit to Rasch model	Item location	Final dimensions
1. Dry cough	Physical	6.329	0.096665	−0.096	Included PhysUP
2. Coughing up mucus	Physical	3.953	0.266624	−0.247	Included PhysUP
3. Painful pressure in ears	Physical	4.059	0.255207	0.441	Included PhysUP
4. Blocked nose	Physical	1.055	0.787859	−0.107	Included PhysUP
5. Runny nose	Physical	3.314	0.34575	−0.03	Included PhysUP
6. Sneezing—no ARTI	Physical	3.929	0.269277	−0.606	Included PhysUP, DIF
6. Sneezing—ARTI +	Physical	2.008	0.57069	0.23	Included PhysUP, DIF
7. Watery eyes—no ARTI	Physical	2.745	0.432674	−0.456	Included PhysUP
7. Watery eyes—ARTI +	Physical	3.17	0.3662	0.737	Included PhysUP, DIF
8. Hoarseness	Physical	7.975	0.046524	−0.041	Included PhysUP
9. Feeling feverish	Physical	6.156	0.104251	−0.429	Included PhysLOW
10. Sweats	Physical	2.038	0.564656	−0.704	Included PhysLOW
11. Chills	Physical	3.769	0.287563	0.183	Included PhysLOW
12. Headache	Physical	9.684	0.021455	−0.515	Included PhysUP
13. Tickles in the throat	Physical	6.561	0.087294	−0.473	Included PhysUP
14. Sore throat—no ARTI	Physical	1.894	0.594612	0.583	Included PhysUP
14. Sore throat—ARTI +	Physical	0.762	0.858581	−0.575	Included PhysUP, DIF
15. Muscle pains	Physical	20.842	0.000115	−0.104	Included single Item
16. Joint pains	Physical	30.301	0.000001	0.021	Excluded
17. Chest pain	Physical	2.885	0.409619	0.416	Included PhysLOW
18. Painful sinuses	Physical	8.342	0.039458	0.727	Included PhysUP
19. Swollen glands	Physical	7.298	0.06298	0.428	Included PhysUP
20. Loss of appetite	Physical	0.632	0.889047	−0.069	Included PhysLOW
21. Problems breathing	Physical	10.75	0.01316	−0.125	Included PhysLOW
22. Wheezing	Physical	1.403	0.704885	0.219	Included PhysLOW
23. Shortness of breath	Physical	4.221	0.23855	0.51	Included PhysLOW
24. Difficulty in thinking clearly?	Psychological	3.505	0.320075	0.952	Included Psych
25. Difficulty in going about your daily business	Psychological	2.581	0.460838	−0.151	Included Psych
26. Felt dizzy	Energy				Excluded
27. Felt tired	Energy				Included single item
28. Not feeling yourself—no ARTI	Psychological	1.884	0.59687	0.069	Included Psych, DIF
28. Not feeling yourself—ARTI +	Psychological	2.805	0.422647	−1.053	Included Psych, DIF
29. Being so unwell you had to stay in bed	Energy				Included single item
30. Poor quality of sleep	Sleep	5.064	0.167197	−0.757	Included Sleep
31. Waking up several times at night	Sleep	2.68	0.443613	−1.019	Included Sleep
32. Difficulty falling asleep	Sleep	10.015	0.018438	0.528	Included Sleep
33. Been awake most of the night	Sleep	4.601	0.203426	1.248	Included Sleep
34. Been in a bad mood	Psychological	2.844	0.416254	0.082	Included Psych
35. Been irritable	Psychological	0.271	0.965458	0.1	Included Psych
36. Taken medicine against fever	Medicine	17.152	0.000658	−0.588	Included Medicine
37. Taken painkillers	Medicine	6.162	0.045915	−2.173	Excluded
38. Taken antibiotics (such as penicillin)	Medicine	4.506	0.105078	−0.134	Included Medicine
39. Taken asthma/COPD spray/ inhalator	Medicine	1.025	0.599084	1.436	Included Medicine
40. Taken asthma/COPD tablets	Medicine	3.457	0.177543	−0.748	Included Medicine
41. Taken antitussives	Medicine	3.078	0.214545	1.387	Included Medicine
42. Taken eye drops	Medicine	7.27	0.02638	−0.903	Included Medicine
43. Taken nasal sprays	Medicine	2.691	0.260466	1.224	Included Medicine
44. Taken herbal medicine	Medicine	3.604	0.164932	−0.089	Included Medicine
45. Signed in sick or cancelled work-related activities	Social				Included single item
46. Cancelled leisure activities	Social				Included single item

Notes. Items 15, 27, 29, 37, 45, and 46 exhibited high face validity. Items 16 and 26 had low face validity. Items 1 to 35 had three degrees of freedom. Items 36 to 46 had two degrees of freedom.ARTI +, acute respiratory tract infection according to treating general practitioner; ARTIQ, Acute Respiratory Tract Infection Questionnaire; COPD, chronic obstructive pulmonary disease; DIF, differential item functioning; Phys, Physical; PhysLOW, Physical-Lower respiratory tract; PhysUP, Physical-Upper respiratory tract.

Table 3 – Wright-Panchapekesan (WP) fit statistics, Person separation index (PSI), and Cronbach's alpha of the five dimensions in the ARTIQ.

Dimensions (no. of items)	WP	Degrees of freedom	P*	PSI	Cronbach's alpha
Physical—upper airways (13)	73.08	48	0.01	0.75	0.85
Physical—lower airways (8)	31.85	24	0.13	0.51	0.83
Psychological (5)	13.89	18	0.74	0.47	0.84
Sleep (4)	22.36	12	0.03	0.72	0.89
Medicine (8)	31.79	16	0.01	0.68	0.53

ARTIQ, Acute Respiratory Tract Infection Questionnaire.
* Significance level at 0.01 to adjust for multiple testing.

Table 4 – ARTIQ scores of domains and single items of participants with and without ARTI.

	Range	Healthy (n = 137)			Ill (n = 122)			P*
		Median	IQR	Max–Min	Median	IQR	Max–Min	
Rasch model domains (no. of items)								
Physical—upper airways (13)	0–26	2	0–3	0–16	10	7–14	0–19	<0.001
Physical—lower airways (8)	0–16	0	0–1	0–5	6	3–8	0–13	<0.001
Psychological (5)	0–10	0	0–1	0–10	3	2–5	0–9	<0.001
Sleep (4)	0–8	1	0–3	0–8	3	2–7	0–8	<0.001
Medicine (8)	0–8	0	0–0	0–5	1	0–2	0–5	<0.001
Separate items								
Muscle pain (item 15)	0–2	0	0–1	0–2	1	0–1	0–2	<0.001
Joint pains (item 16)	0–2	0	0–1	0–2	0	0–1	0–2	0.127
Felt dizzy (item 26)	0–2	0	0–0	0–2	0	0–1	0–2	0.088
Felt tired (item 27)	0–2	1	0–1	0–2	2	1–2	0–2	<0.001
Being so unwell … (item 29)	0–2	0	0–0	0–1	1	0–1	0–2	<0,001
Signed in sick or … (item 45)	0–1	0	0–0	0–1	1	0–1	0–1	<0.001
Cancelled leisure activities (item 46)	0–1	0	0–0	0–1	1	0–1	0–1	<0.001

ARTI, acute respiratory tract infection; IQR, interquartile range.
* P value from a Wilcoxon signed-rank test.

Separation Index was at 0.7 in three of the five domains. An example of the questionnaire layout is presented in Figure 1.

Discriminative ability/known groups' validity

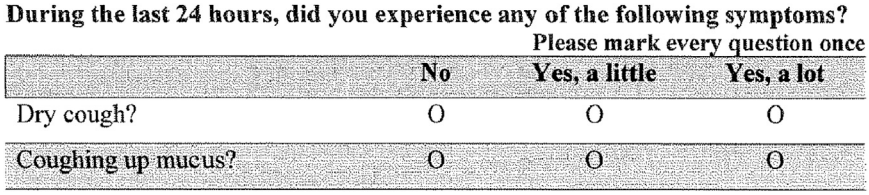
Based on the item scores *No* = 0, *Yes—some* = 1, and *Yes—a lot* = 2 for polytomous items and the item scores *No* = 0 and *Yes* = 1 for dichotomous items, the scores from participants with and without an ARTI were analyzed in single items as well as sum-scores of validated domains. All scales and single items (except item 16 "Joint pains" and item 26 "Felt dizzy") exhibited significantly higher scores in participants with an ARTI (Table 4).

Discussion

Respiratory tract infections are common in primary care, and because no gold standard exists, the need for sound measurements of PROs on severity and functional impact is essential. The ARTIQ has been developed to meet this need. The ARTIQ uses both qualitative and quantitative psychometrical methods to ensure high face and content validity and unidimensional invariant measurement with reliable sum-scores for the five scales: PhysUP and PhysLOW symptoms, Psychological, Sleep, and Medicine, plus five single items (Table 2).

Strengths and Limitations of the Study

In PROs, great care must be exercised to ensure an adequate and understandable description of the patient's condition [10,14], particularly when no gold standard exists. The qualitative part of the present study, involving future users in the development of questionnaire items, wording, and response categories, ensured a high level of face and content validity of the broad term "severity and functional impact" or "bothersomeness" of an ARTI, and ascertained the acceptability of the questionnaire (proportion of missing data <5%). The recall period was set to 24 hours to ascertain that the symptom or impairment measured was


During the last 24 hours, did you experience any of the following symptoms?
Please mark every question once

	No	Yes, a little	Yes, a lot
Dry cough?	O	O	O
Coughing up mucus?	O	O	O

Fig. 1 – An example of the layout of items and response categories in the Acute Respiratory Tract Infection Questionnaire.

present in the mind and of relevance to the status of the participant at testing but still sensitive to a change in severity of the ARTI in question. Barret et al. [15] used the same recall time in the WURSS questionnaire of the common cold.

The ability of the ARTIQ dimensions to discriminate between participants with and without an ARTI on the basis of sum-scoring of the scales is not of diagnostic use. Nevertheless, if applied in trials testing the participants at different time points, it illustrates the anticipated response to change in a recovery from an ARTI to a normal level and thus determines when subjects can be regarded as cured. We believe that the term sensitivity to change or responsiveness is difficult to assess in an acute respiratory condition with no gold standard or similar questionnaires for comparison because deciding which time points to compare is complex: is the disease gradually or acutely worsening or is the patient recovering?

It is important to acknowledge that the DIF of the psychological dimension must be taken into account when sum-scoring a longitudinal cohort study. In a randomized controlled trial, however, the DIF will not affect the results because of randomization.

The broad inclusion criteria (any ARTI) without application of strict diagnostic measures may raise doubts about the specificity of the illnesses included. Realistic and pragmatic research, however, will benefit from an instrument that is comprehensive and symptom based and covers the full range of illnesses met in primary care. Nevertheless, the current findings should be externally validated in a different ARTI population to better assess the generalizability of the findings.

Findings in Relation to Other Studies

Unpublished questionnaires and diaries have been used to measure PRO in clinical trials despite the lack of rigorous validation of the applied instrument(s); this is not appropriate [14]. Often, the questionnaires were developed without qualitative methods ensuring relevance and comprehensibility. An example to follow could be the disease-specific WURSS questionnaire on the common cold, which has been developed and standardized as a result of the work by Barret et al. [15]. The ARTIQ, however, is the first validated PRO questionnaire developed to measure the functional impairment in an ARTI *en bloc*, which in selected studies may better reflect the diagnostic uncertainties of primary care research in a pragmatic trial. Furthermore, the ARTIQ is the first Rasch-tested ARTI questionnaire. It comprises five independent scales for comparing each dimension affected because we believe that adding different scales into one single questionnaire sum score cannot be justified; in addition, discriminative power will diminish. The number of items (43) is comparable to other research questionnaires [15,18] as is time of completion (5 minutes).

Unanswered Questions and Future Research

The ARTIQ was developed in Danish. Any new language version of the questionnaire must also be tested by using Rasch analysis to ensure its measurement properties. Rasch analysis enables invariant comparison across subjects of different nationalities and cultural or ethnic backgrounds. The ARTIQ will be made freely available for nonprofit research.

Acknowledgments

The authors thank the participating general practitioners and their patients for making this study possible.

Source of funding: Rune Aabenhus received a grant from The General Practitioners Trust for Development (PLU).

REFERENCES

[1] Goossens H, Ferech M, Vander SR, Elseviers M. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet 2005;365:579–87.
[2] Gonzales R, Malone DC, Maselli JH, Sande MA. Excessive antibiotic use for acute respiratory infections in the United States. Clin Infect Dis 2001;33:757–62.
[3] Gwaltney JM Jr, Hendley JO, Patrie JT. Symptom severity patterns in experimental common colds and their usefulness in timing onset of illness in natural colds. Clin Infect Dis 2003;36:714–23.
[4] Stanton N, Francis NA, Butler CC. Reducing uncertainty in managing respiratory tract infections in primary care. Br J Gen Pract 2010;60:e466–75.
[5] Hopstaken RM, Coenen S, Butler CC. Treating patients not diagnoses: challenging assumptions underlying the investigation and management of LRTI in general practice. J Antimicrob Chemother 2005;56:941–3.
[6] Aabenhus R, Jensen JU. Procalcitonin-guided antibiotic treatment of respiratory tract infections in a primary care setting: are we there yet? Prim Care Respir J 2011;20:360–7.
[7] Butler CC, Kelly MJ, Hood K, et al. Antibiotic prescribing for discoloured sputum in acute cough/lower respiratory tract infection. Eur Respir J 2011;38:119–25.
[8] Cals JW, Schot MJ, de Jong SA, et al. Point-of-care C-reactive protein testing and antibiotic prescribing for respiratory tract infections: a randomized controlled trial. Ann Fam Med 2010;8:124–33.
[9] Schuetz P, Christ-Crain M, Muller B. Procalcitonin and other biomarkers to improve assessment and antibiotic stewardship in infections–hope for hype? Swiss Med Wkly 2009;139:318–26.
[10] Jones PF, Miravitlles MF, van der Molen TF, Kulich K. Beyond FEV(1) in COPD: a review of patient-reported outcomes and their measurement. Int J Chron Obstruct Pulmon Dis 2012;7:697–707.
[11] Editorial. Quality of life and clinical trials. Lancet 1995;346:1–2.
[12] Curtis JR, Patrick DL. The assessment of health status among patients with COPD. Eur Respir J 2003;41(Suppl.).36s–45.
[13] Brodersen J, Thorsen H, Kreiner S. Consequences of screening in lung cancer: development and dimensionality of a questionnaire. Value Health 2010;13:601–12.
[14] McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. BMC Med 2011;9:86.
[15] Barrett B, Locken K, Maberry R, et al. The Wisconsin Upper Respiratory Symptom Survey (WURSS): a new research instrument for assessing the common cold. J Fam Pract 2002;51:265.
[16] Juniper EF, O'Byrne PM, Guyatt GH, et al. Development and validation of a questionnaire to measure asthma control. Eur Respir J 1999;14:902–7.
[17] Lamping DL, Schroter S, Marquis P, et al. The community-acquired pneumonia symptom questionnaire: a new, patient-based outcome measure to evaluate symptoms in patients with community-acquired pneumonia. Chest 2002;122:920–9.
[18] Meguro M, Barley EA, Spencer S, Jones PW. Development and validation of an improved, COPD-specific version of the St. George Respiratory Questionnaire*. Chest 2007;132:456–63.
[19] Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2006;18:59–82.
[20] Wright BD. Fundamental measurement for psychology. In: Embertson SHSL, ed., The new rules of measurement: What every psychologist and educator should know. Mahwah, New Jersey and London: Lawrance Erlbaum Associates; 1999.
[21] Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care 2004;42:I7–16.
[22] Rosenbaum P. Criterion-related construct validity. Psychometrika 1989;54:625–33.
[23] Rasch G. An informal report on a theory of objectivity in comparisions. In: Van der Kamp L, Vlek C, eds. Psychological Measurement Theory. Leyden: University of Leyden, 1967.
[24] Holland P. Differential Item Functioning. Hillsdale: Erlbaum, 1993.
[25] Brodersen J, Meads D, Kreiner S, et al. Methodological aspects of differential item functioning in the Rasch model. J Med Econom 2007;10:309–24.
[26] Andrich D, Luo G. Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. J Appl Meas 2003;4:205–21.
[27] Wright BD, Panchapakesan N. A procedure for sample-free item analysis. Educ Psychol Meas 1969;29:23–48.
[28] Andrich D. Rasch models for measurement. Berverly Hills, CA: Sage, 1998.