Development, Sensibility, and Reliability of the Toronto Axial Spondyloarthritis Questionnaire in Inflammatory Bowel Disease

Khalid A. Alnaqbi, Zahi Touma, Laura Passalent, Sindhu R. Johnson, George A. Tomlinson, Adele Carty, and Robert D. Inman

ABSTRACT **Objective.** There is an unacceptable delay in the diagnosis of axial spondyloarthritis (axSpA) in its early stages among patients at high risk, in particular those with inflammatory bowel disease (IBD). Our objectives were to develop a sensible and reliable questionnaire to identify undetected axSpA among patients with IBD.

Methods. Literature was reviewed for item generation in the Toronto axSpA Questionnaire on IBD (TASQ-IBD). Sensibility of the questionnaire was assessed among healthcare professionals and patients. This assessment was related to purpose and framework (clinical function, clinical justification, and clinical applicability), face validity, comprehensiveness [oligo-variability (limiting the questionnaire to important items) and transparency], replicability, content validity, and feasibility. The test-retest reliability study was administered to 77 patients with established IBD and axSpA. Kappa agreement coefficients and absolute agreement were calculated for items.

Results. Three domains included IBD, inflammatory back symptoms, and extraaxial features. The entry criterion required a patient to have IBD and back pain or stiffness that ever persisted for ≥ 3 months. Iterative sensibility assessment involved 16 items and a diagram of the back. Kappa coefficients ranged from 0.81–1.00 for each item. Absolute agreement across all items ranged from 91% to 100%.

Conclusion. TASQ-IBD is a newly developed, sensible, and reliable case-finding questionnaire to be administered to patients with IBD who have ever had chronic back pain or stiffness persisting for ≥ 3 months. It should facilitate identification and timely referral of patients with IBD to rheumatologists and minimize the delay in diagnosis of axSpA. Consequently, it should assess the prevalence of axSpA in IBD. (J Rheumatol First Release Sept 1 2013; doi:10.3899/jrheum.130048)

Key Indexing Terms:
SPONDYLOARTHRITIS   INFLAMMATORY BOWEL DISEASE   QUESTIONNAIRE
CLINIMETRICS   PSYCHOMETRICS   SENSIBILITY

Ankylosing spondylitis (AS) is considered the prototype of spondyloarthritis (SpA). The newly coined term "axial SpA (axSpA)" comprises predominantly axial symptoms, either without radiographic sacroiliitis (nonradiographic axSpA) or with radiographic sacroiliitis (i.e., AS). The diagnosis of AS is usually made 8 to 11 years after the onset of symptoms, mainly owing to poor recognition of the disease in its early stages[1]. In fact, definitive radiographic sacroiliitis may not appear until about 8 years after symptom onset, which limits the usefulness of the modified New York criteria in classifying early disease[1,2]. The Assessment of SpondyloArthritis International Society (ASAS) has developed and validated new classification criteria for axSpA to address this[3].

Patients with inflammatory bowel disease (IBD) are at high risk of developing axial symptoms of inflammatory back pain (IBP). The prevalence of IBD-associated AS (i.e., AS diagnosed in patients with IBD) using the modified New York criteria is 4% to 10%[4]. Symptoms of axSpA can precede IBD symptoms in 31% to 50% of patients, while symptoms of IBD and axSpA can occur simultaneously in 15% to 40% of patients[5,6,7,8]. The literature does not support

that patients with IBD have a different presentation of AS. As in primary AS, IBD-associated AS has the same age of onset, delay in diagnosis, work disability, reduced quality of life, and disease activity[1,5,9,10].

Because the predominant feature of axSpA is IBP, attempts have been made to refine IBP features. Three English self-reported questionnaires for identifying IBP exist. The first was developed by Calin, et al[11]. It comprises 5 items for IBP (excluding neck pain), which were validated in patients with AS and mechanical or nonspecific back pain. Although the sensitivity of fulfilling 4 out of 5 criteria was 95% and the specificity was 85%, the diagnostic performance was found to be lower in subsequent validation studies (sensitivity 23% to 38% and specificity 75%)[12]. The second questionnaire was a 12-item case ascertainment questionnaire (CAQ) that was developed and validated to identify patients with radiographic AS[13]. The sensitivity and specificity of the CAQ in the validation sample was about 67% and 95%, respectively. However, it did not target patients with early axSpA because the mean disease duration among patients with AS was 21.8 years. It also did not address patients with IBD-associated AS. A third screening questionnaire for IBP consisted of 6 items[14]. The questionnaire was administered to patients with established AS and mechanical back pain. In multivariate logistic regression analysis, diurnal variation had the strongest association with IBP, with a specificity of 92% but a sensitivity of 49%. The main focus of these questionnaires was to refine the definition of IBP for earlier identification of AS, but this was found to increase posttest probability from 5% to 14%[15]. Further, the interpretation of IBP requires clinical experience. Therefore, IBP is problematic if used as the sole referral criterion. The recommendation is to combine IBP with other features of axSpA to achieve a high posttest probability of axSpA[15].

Attention needs to be directed toward those patients at heightened risk of developing axSpA, as is the case in patients with IBD. One approach to identify those IBD patients with undetected axSpA is to screen them when they present to their gastroenterologists for followup of their IBD. Such screening is defined as case finding. A good case-finding questionnaire requires items that are accurate, highly sensitive, reasonably specific, and that have good predictive value[16]. Further, in this modern era, there is an expectation that instruments should be evaluated for their measurement properties. The clinimetric method supports the use of clinical judgment (by physicians and/or patients) during development of an instrument measuring a complex clinical phenomenon such as axSpA in patients with IBD. Important facets of the clinimetric approach include sensibility (usefulness of an instrument), reliability, validity, and responsiveness. The aim of our study was to develop a simple, sensible, and reliable self-reported questionnaire that allows earlier case finding of axSpA in patients with IBD.

MATERIALS AND METHODS

Design outline. Figure 1 illustrates the phases of development of the Toronto axSpA questionnaire on IBD (TASQ-IBD).

Item generation. A thorough literature review was done using Medline (from 1941 to April 2011), PubMed, and EMBASE (inception to April 2011) to identify potential items for the development of the questionnaire. Papers were restricted to the English language. Keywords included various combinations of inflammatory back pain, back pain, ankylosing spondylitis, spondyloarthritis, spondyloarthropathy, symptoms, features, manifestations, risk factors, questionnaire, screening, case finding, and diagnosis. Articles were hand-searched to identify additional relevant articles.

Cardinal articles in this field (Hart, et al[17]; Calin, et al[11]; Moller, et al[18]; Gran[19]; and Rudwaleit, et al[12,15]) and previously developed classification criteria for AS and SpA (Rome criteria for AS, 1961; New York criteria for AS, 1968; modified New York criteria, 1984; Mau criteria for identifying early AS, 1985; Amor criteria for SpA, 1990; European Society of Spondyloarthropathy Group criteria for SpA, 1991; and ASAS classification criteria for early axial and peripheral SpA, 2009, 2010)[3,20,21,22,23,24] were also reviewed with a focus on clinical history, particularly features of IBP.

Response options and instrument format. The majority of items asked closed-ended questions with binary responses, with 2 questions providing continuous data. Some branching questions (subquestions) were included. Three major headings were determined *a priori*, representing the domains of the questionnaire: IBD, inflammatory back symptoms, and extraaxial features.

Item reduction using sensibility assessment. Sensibility assessment refers to the usefulness of an instrument. It includes a statement of purpose and framework (clinical function, clinical justification, and clinical applicability), overt format (comprehensibility and replicability), face and content validity, and feasibility (Table 1)[25]. The sensibility of TASQ-IBD was evaluated using a sensibility instrument that was adapted from Rowe and Oxman[26]. This instrument has been successfully used in the development of other recent instruments, such as the Pediatric Cardiopulmonary Physiotherapy Discharge Tool[27] and the Human Immunodeficiency Virus Disability Questionnaire[28]. The sensibility of TASQ-IBD was evaluated by a purposive sample of academic healthcare professionals and researchers involved with SpA. This sample comprised 2 rheumatologists with expertise in SpA, an SpA physiotherapist, a clinical trial manager, 2 SpA nurses, an SpA research analyst, and 2 general rheumatologists.

Pilot study. This phase aimed at testing the questionnaire among a convenience sample (n = 19) of patients with SpA attending the Spondylitis Clinic at the Toronto Western Hospital. This was done to evaluate the sensibility and qualitative analysis from patients' perspective. The clinic is one of the largest academic centers for the care of patients with axSpA in Canada. All patients, ≥ 18 years old and fluent in English, gave their informed consent[29].

Patient characteristics were collected including age, sex, the level of education (below grade 8, high school incomplete, high school graduate, college, university), the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), and the Bath Ankylosing Spondylitis Functional Index (BASFI) scores of the last clinic visit[30,31]. The sensibility assessment instrument was administered to patients and ended with a free text comment area. Also, patients were interviewed by an investigator (KAA) to ascertain their comments on the TASQ-IBD.

Repeat sensibility assessment. Revisions were made to the TASQ-IBD based on feedback from healthcare professionals and patients. The sensibility of TASQ-IBD was then reassessed by the healthcare professionals.

Coding the responses on the questionnaire. Dichotomous response options were coded as 1 for Yes and 0 for No. Question 4 and subquestion 6, which contained continuous variables, were converted into dichotomous variables. The first question asks about the age of onset of back pain, which
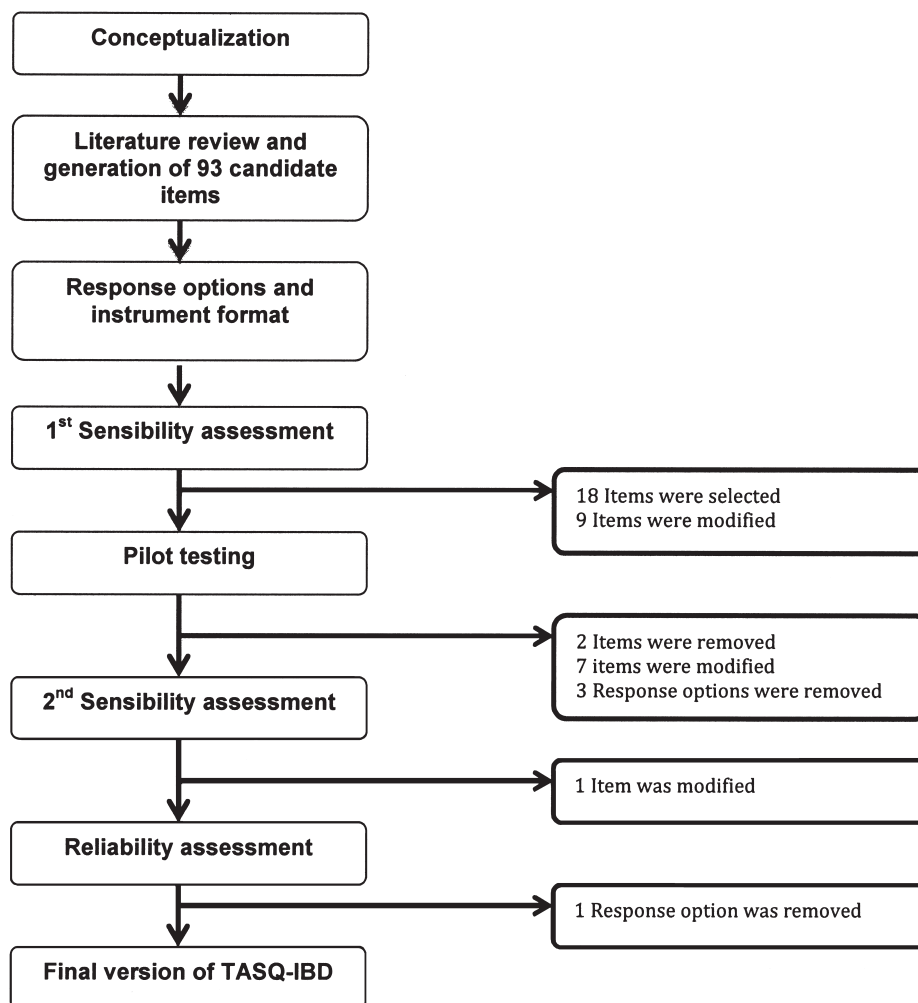
Figure 1. Stages of development of the Toronto Axial Spondyloarthritis Questionnaire on inflammatory bowel disease (TASQ-IBD).

was dichotomized into age ≤ 45 years and age > 45 years in accordance with the ASAS classification criteria for axSpA. The second question asks about duration of morning stiffness, which was dichotomized into ≥ 30 min and < 30 min. Subquestion 7 and questions 11 and 12 have 3 response options: Yes, and 2 negative detailed response options. The negative options were collapsed into 1 category (No) for the purpose of analysis.

Reliability assessment. Reliability evaluates the reproducibility of an instrument that measures an attribute or a construct. Test-retest (intrarater) reliability assesses the stability of the item scores on different occasions to the population of interest. Test-retest reliability was evaluated within a 2-week period between administration times to reduce the risk of recall bias and intercurrent flare of axSpA[32].

Sampling method. As of July 1, 2011, the database of the Spondylitis Clinic had 636 patients with axSpA. Of those patients, 77 (12.1%) had concurrent IBD that was diagnosed by gastroenterologists based on clinical, endoscopic, and histological evaluations. All patients had chronic back pain or stiffness that persisted for ≥ 3 months.

TASQ-IBD was mailed to all 77 patients. The tailored design method of Dillman was adapted to maximize the response rate[33]. Each patient was mailed 2 copies of the TASQ-IBD questionnaire accompanied by a personalized covering letter, which explained the purpose of the reliability study

and provided clear instructions about the completion of each questionnaire on 2 occasions 1 to 2 weeks apart. Patients were blinded to knowing that their responses would be compared among them to minimize the possibility of Hawthorne effect, which refers to responders modifying their responses because of awareness that they were being studied[34]. This information was collected from the database: descriptions of the characteristics between responders and nonresponders, demographics, the highest level of education at the last clinic visit, type of IBD, age of onset of IBP, duration of axSpA disease, BASDAI and BASFI scores of the last clinic visit, erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP) levels at the last clinic visit. Nunnally and Bernstein have suggested a reliability coefficient (R) of 0.70 for a measuring instrument for research purposes, and 0.90 for clinical purposes[35]. A minimally desired reliability of 0.80 was chosen. Therefore, the calculated sample size with hypothesized R = 0.90, standard error 0.05, and 2 observations for each patient is 24 patients[36].

Statistical analyses. Descriptive statistics were used for the data. Appropriate parametric and nonparametric tests were used to compare the variables between responders and nonresponders. For test-retest reliability, Cohen's kappa (κ) statistics were determined and were interpreted using the recommendations of Landis and Koch as follows: κ < 0.00 = poor agreement, κ between 0.00–0.20 = slight agreement, κ 0.21-0.40 = fair

Table 1. Checklist of sensibility assessment with definitions and sensibility assessment[25].

Concept	Definition	Sensibility Assessment
Purpose and framework		
Clinical function	Purpose of the questionnaire	The questionnaire is intended to ask questions related to the presence of axial spondyloarthritis in patients with inflammatory bowel disease. Do you think this questionnaire achieved this goal?
Clinical justification	Rationale behind questionnaire development	
Clinical applicability	Application of the questionnaire to a specific group of patients in a specific setting	
Face validity	Accurate measurement of what the questionnaire is trying to measure	Do you feel that the questionnaire accurately measures what we are trying to measure?
Comprehensibility	Ability of understanding the questionnaire	
Oligo-variability	Limiting the questionnaire to important items	Do you feel that the questionnaire included the necessary questions?
Transparency	Ability to see through the contribution of each scored item and its response options to the final score	Do you think the response options were appropriate (for example, number of options)? Assuming that each question is weighted equally with a score of 1, do you think that the scoring is simple?
Replicability	Clarity of instructions	Do you think the instructions were clear?
Content validity*	Comprehensiveness of the questionnaire	
Important omissions	Exclusion of important items	Do you think an important question was missing?
Inappropriate inclusions	Inclusion of inappropriate items	Do you think an inappropriate question was included?
Feasibility	Ease of use of the questionnaire	
Time to completion	Amount of time taken to complete the questionnaire	Please indicate the amount of time you took to complete reading the questionnaire? _____ minutes
Acceptability	Approval of the questionnaire by patients and health care workers by excluding sensitive or potentially offensive items	Were there any questions that were unacceptable (too personal or offensive)?
Readability	Ease with which the words can be understood	
Clarity of questions	Ability to understand the questions easily	Do you think the questions were clear?
Flow of questions	Transitioning of questions in a logical manner	Was the flow of questions smooth?
Readability tests	Flesch Readability Ease Scale[†] and Flesch-Kincaide grade level	Were the sentences easy to read?
Typographical errors	Presence of typographical errors	Were there any typographical errors? If so, please indicate where.
Font size	Appropriateness of the font size used in the questionnaire format	Was the font size appropriate?
Illustrations	Clarity and usefulness of diagrams and pictures used in the questionnaire	Were the pictures and diagrams useful? If not, please explain.

* Agreement on each item was assessed by consensus. † The Flesch Reading Ease is a 100-point scale interpreted as follows: 90–100 = very easy, 80–89 = easy, 70–79 = fairly easy, 60–69 = standard, 50–59 = fairly difficult, 30–49 = difficult, and 0–29 = very confusing[50].

agreement, κ 0.41–0.60 = moderate agreement, κ 0.61–0.80 = substantial agreement, and κ > 0.80 = almost perfect agreement[37]. The percentage of agreement for the items was also determined. Statistical analyses were done using SAS (version 9.3) and R (version 2.14.2; The R Foundation for Statistical Computing). Statistical analyses were 2-sided, and the statistical significance was defined by p < 0.05.

The institutional review ethical board approved our study.

RESULTS

Item generation. Ninety-three potential items were generated. Sensitive and/or specific items were chosen in keeping with the properties of a case-finding instrument. Pictures of peripheral arthritis of hand joints, dactylitis of a finger and a toe, and acute anterior uveitis were included along with a diagram of the back.

Item reduction. An entry criterion was added to reduce referrals for mechanical back pain. To qualify for the questionnaire, a patient must have IBD and back pain or stiffness that ever persisted for ≥ 3 months. Iterative sensi-

bility assessment of TASQ-IBD resulted in selection of 18 items with some modifications for shorter and clearer sentences (Table 2). This version included a diagram of the back and one picture of toe dactylitis.

Pilot study. In all the 3 pilot stages (n = 4 patients in stage I, n = 9 in stage II, and n = 6 in stage III), the total number of patients was 19 with a median age of 36 years (range 24–61 yrs) and 68% were men. The median BASDAI score was 3.0 (range 0–8.8), and the median BASFI score was 5.0 (range 0–7.7). Forty-two percent of patients were high school graduates, 42% were university graduates, and 16% were college graduates. Table 3 summarizes the sensibility assessment among patients with axSpA. During each successive stage, patients who participated in the previous stage did not participate in the next one.

Throughout all 3 pilot stages, items underwent either further modifications or removal if they were still unclear. For example, the item asking about the mode of onset of

Table 2. Comparison of sensibility assessments completed by healthcare professionals for the pre-piloted and post-piloted Toronto Axial Spondyloarthritis Questionnaire on inflammatory bowel disease (TASQ-IBD).

Sensibility Assessment	Pre-pilot Testing, n = 9 (%)*	Post-pilot Testing, n = 9 (%)*
Clinical function	9 (100)	9 (100)
Comprehensibility		
Oligo-variability	7 (78)	9 (100)
Transparency		
Appropriateness of the number of response options	9 (100)	9 (100)
Weighting each item	0 (0)	8 (89)
Replicability	7 (78)	9 (100)
Face validity	6 (67)	9 (100)
Content validity		
Important omissions	2 (22)	0 (0)
Inappropriate inclusions	3 (33)	0 (0)
Feasibility		
Time to completion, median (range)†	4 (3–5) min	4 (2–5) min
Acceptability	8 (89)	9 (100)
Readability		
Clarity of all questions	4 (44)	9 (100)
Flow of questions	7 (78)	9 (100)
Flesch reading ease scale†	72.2%	74.7%
Flesch-Kincaid grade level†	5.6	5.3
Absence of typographical errors	8 (89)	9 (100)
Appropriate font size	7 (78)	9 (100)
Usefulness of illustration(s)	5 (56)	9 (100)

* All the numbers in the table represent number of healthcare professionals who agreed on the corresponding principles (percentages in parentheses). † This was determined independently from the healthcare professionals.

Table 3. Iterative stages of sensibility assessment of Toronto Axial Spondyloarthritis (axSpA) Questionnaire on inflammatory bowel disease (TASQ-IBD) by patients with axSpA.

Sensibility Assessment	Stage I, n = 4 (%)*	Stage II, n = 9 (%)*	Stage III, n = 6 (%)*
Comprehensibility			
Transparency†	2 (50)	9 (100)	6 (100)
Replicability	4 (100)	9 (100)	6 (100)
Feasibility			
Completion time, median (range)	5 (3–5) min	4 (3–5) min	4 (3–5) min
Acceptability	4 (100)	9 (100)	6 (100)
Readability			
Clarity of all questions	2 (50)	7 (78)	6 (100)
Flow of questions	4 (100)	9 (100)	6 (100)
Ease of readability	3 (75)	7 (78)	6 (100)
Absence of typographical errors	4 (100)	8 (89)	6 (100)
Appropriate font size	4 (100)	9 (100)	6 (100)
Usefulness of picture(s) and/or diagram	4 (100)	6 (67)	6 (100)
Actions following each stage	Removal of 4 items and modifications of 4 items	Removal of 2 items, removal of dactylitis picture, and modification of 3 items	Minor modifications of few items

* All the numbers in the table represent number of patients who agreed on the corresponding principles (percentages in parentheses). † Appropriate response options for each item.

back pain underwent iterative remedy (modification of stem question and response options) until the final version was reached. Some patients suggested the addition of response options to make the questions more practical. For instance, they suggested adding a third response option (I do not get up) for the item asking about improvement of nocturnal back pain when getting up and moving. The questions about awakening because of back pain during the second half of

the night and alternating buttock pain were removed. Despite their importance as characteristics of IBP, patients found them difficult to understand. Some patients suggested adding "hip pain" to the diagram of the back although they drew a line pointing toward the sacroiliac joints. This item was dropped during item selection because a proper definition was lacking. These modifications led to a revised questionnaire with 16 items and 1 diagram. We retained a subquestion on prior diagnosis of AS to confirm the diagnosis in patients during the reliability study.

Repeat sensibility assessment. There was a significant improvement in the sensibility of the final questionnaire compared to the initial sensibility assessment. This time, all healthcare professionals but 1 agreed on the simplicity of the scoring system assuming that each question was weighted equally by a score of 1 (Table 2).

Reliability study. Of the 77 mailed questionnaires, 34 were returned by the cutoff date. Six questionnaires were returned owing to incorrect addresses. The response rate during the summer of 2011 was 44.2%. Questionnaires from 5 out of 34 patients had ≥ 15% missing answers. Because calculating κ statistics using our small sample size requires each item to be answered, we contacted 4 of the nonresponders within a week of receiving their returned envelopes. The unanswered questions were read for them over the phone as written without providing any further explanation to avoid inter-viewer bias. The overall percentage of missing items was 16.6%. The most common items not completed by the 5 responders were the type of IBD, time to develop back pain or stiffness, responsiveness of nonsteroidal antiinflam-matory drugs, family history of AS, and history of peripheral arthritis. Data from 33 out of 34 patients were usable for the descriptive analysis, and 34 pairs of question-naires were usable for test-retest reliability. We could not identify 1 male patient because of a missing date of birth on both questionnaires and inability to crosscheck the code on his questionnaire with the spondylitis database (clerical error), but we used his questionnaires for test-retest relia-bility. Because we required a sample size of 24 to assess test-retest reliability, the final number of (n = 33) responders was sufficient.

Comparison between responders and nonresponders. There was no statistically significant difference between the groups in terms of age, sex, the highest level of education, types of IBD, age of onset of back pain, duration and activity of axSpA, functional activities, ESR, or CRP levels (Appendix 1).

All responders confirmed at test and retest times that they had back pain or stiffness that persisted for ≥ 3 months and were diagnosed with AS.

At this stage, the healthcare professionals agreed on removing a response option "Prior diagnosis of psoriatic arthritis" from item 13 for 2 reasons. First, this item was only endorsed by 1 patient. Second, during the validation phase of the CAQ, this item did not differentiate AS from chronic back pain[13]. We considered psoriasis adequate for this questionnaire. Further, psoriasis was recently shown to be more associated with nonradiographic axSpA (OR 3.6) compared to classic AS[38]. The final version of TASQ-IBD (dated May 2012) is shown in Figure 2.

Test-retest reliability. The κ coefficients of each item ranged between 0.81 and 1.00, which indicates almost perfect agreement for test-retest reliability (Table 4). The absolute percentage of agreement across all items ranged from 91% to 100%.

The range of κ coefficients for each domain was as follows: 0.84–1.00 for the IBD, 0.85–1.00 for inflammatory back symptoms, and 0.81–1.00 for extraaxial features. The ranges of absolute agreement between different domains were as follows: 91% to 100% for the IBD, 94% to 100% for inflammatory back symptoms, and 91% to 100% for extraaxial features.

DISCUSSION

The self-administered TASQ-IBD was designed to serve as a case-finding instrument that can facilitate referrals of IBD patients with suspected axSpA to rheumatologists for further evaluation. Early diagnosis allows for earlier intervention, an important concept because tumor necrosis factor inhibitors have been found to be efficacious in early disease[39]. Previous questionnaires focused on refining the characteristics of IBP and did not specifically target IBD patients or use an illustration. TASQ-IBD is inexpensive and easy to administer at a single point in time to patients with IBD who have ever had chronic back pain or stiffness that has persisted for ≥ 3 months. Completion of the question-naire takes ≤ 5 min, which is in accordance with the recom-mended completion time of between 5 and 15 min[40].

In the design of our questionnaire, we adopted multiple steps to ensure an appropriate methodology and to yield a sensible and reliable instrument. The items on TASQ-IBD represent discriminatory features that increase the likelihood of axSpA.

Psychometric scales start with a large pool of candidate items, which was used in the development of TASQ-IBD. Psychometrics relies heavily on statistical analyses in item reduction with less detailed assessment of face and content validity[41]. A combination of psychometric and clinimetric approaches was used in the development of TASQ-IBD. The sensibility assessment proved to be a valuable and compre-hensive approach during item selection and pilot testing that enabled an analysis of different facets of the questionnaire and addressed any weaknesses. The clinimetric approach has been used in the literature to create instruments, such as the New York Heart Association Functional Classification and the Pittsburgh Sleep Quality Index[25,42].

Assessing internal consistency is of less relevance in

DATE (dd/mm/yyyy)	Date of Birth (dd/mm/yyyy)	SEX	STUDY #
__/__/____	__/__/____	☐ Male ☐ Female	

Toronto Axial Spondyloarthritis Questionnaire (TASQ) for Patients with Inflammatory Bowel Disease

1. Has there been a time when you had pain or stiffness in your back for 3 months or more? Please check (✓) the correct response.
   ☐ Yes ☐ No

IF YES TO QUESTION 1, PLEASE ANSWER ALL QUESTIONS BELOW BY CHECKING (✓) THE CORRECT RESPONSE AND FILL IN THE BLANK WHEN REQUIRED ON BOTH PAGES

Inflammatory bowel disease

2. What type of inflammatory bowel disease do you have? **Check ONE answer**
   ☐ Crohn's disease   ☐ Ulcerative colitis   ☐ Unclassified (indeterminate) colitis

3. Has your inflammatory bowel disease ever been treated with a biologic drug such as Remicade or Humira or other biologic drugs?
   ☐ Yes ☐ No

   If YES, please answer the following (**if NO → skip to Question 4**):

   Specify the biologic drug:

	Currently	Previously
Remicade	☐	☐
Humira	☐	☐
Other _____	☐	☐
_____	☐	☐

   Did your back pain or stiffness, that lasted for 3 months or more, improve **after** taking ANY biologic drug?
   ☐ Yes ☐ No

Back pain or stiffness

4. At what age did your back pain or stiffness begin? _____ years of age

5. When your back pain first began, did it develop: (**check ONE answer**)
   ☐ Quickly (over hours or days)   ☐ Slowly (over weeks or months)

6. Is your back stiff in the morning?
   ☐ Yes ☐ No

   If YES, for how long? _____ number of hour(s) _____ number of minute(s)
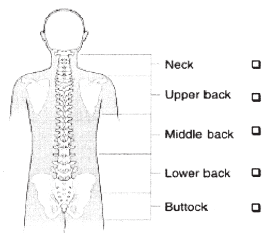
7. Does your back pain currently wake you up at night?
   ☐ Yes ☐ No

   If YES → Does it improve during the night upon getting up, walking or stretching?
   ☐ Yes   ☐ No, this would not improve my back pain   ☐ No, I do not get up

8. Where is your back pain or stiffness usually located? **Check all that apply.**

   Neck ☐
   Upper back ☐
   Middle back ☐
   Lower back ☐
   Buttock ☐

9. Does your back pain or stiffness improve **at all** with daily physical activities?
   ☐ Yes ☐ No

10. Does your back pain or stiffness improve **at all** with rest?
    ☐ Yes ☐ No

11. Does your back pain or stiffness improve after taking anti-inflammatory drugs (such as Motrin™, Advil™, Voltaren™, Naprosyn™, Aleve™, Indocin™, Mobic™, or Celebrex™)?
    ☐ Yes ☐ No, these do not improve my symptoms ☐ No, I do not take anti-inflammatory drugs

12. Do you have a parent or sibling who has been diagnosed with Ankylosing Spondylitis?
    ☐ Yes ☐ No ☐ Unknown

Extra-axial features:

13. Has a doctor ever diagnosed you with **any** of the following conditions? **Check all that apply.**

☐ Ankylosing Spondylitis	☐ Iritis or uveitis	☐ Psoriasis

14. Have you ever had pain **and** swelling in any joint that was **not** related to injury?
    ☐ Yes ☐ No

15. Have you ever had pain **and** swelling of your entire finger or toe that was **not** related to injury?
    ☐ Yes ☐ No

16. Have you ever had heel pain that was **not** related to injury?
    ☐ Yes ☐ No

Figure 2. The Toronto Axial Spondyloarthritis Questionnaire on inflammatory bowel disease (TASQ-IBD).

clinimetric measures because the aim of developing a measuring instrument is to identify different aspects (e.g., characteristics of IBP and presence of extraaxial features) of a complex construct (axSpA in IBD). Therefore, the internal consistency is likely to be low, although this is not always the case[43]. This concept has been adopted in developing quality criteria for measurement properties of health questionnaires[44]. Most psychometricians, on the other hand, tend to focus on constructing unidimensional scales with homogenous items, as this is one of the axioms of classic test theory. However, psychometric indices do not always require internal consistency[45]. This was probably why the most recent Guidelines for Reporting Reliability and Agreement Studies (GRRAS) did not endorse assessment of internal consistency when constructing new measuring instruments[34]. Based on this reasoning, we did not report the internal consistency of TASQ-IBD.

We provided detailed information on the reliability and agreement using coefficients of reliability (along with statistical uncertainty) and percentage agreement as per the recent recommendations of the GRRAS. The degree of impression for some items may be due to the small sample size.

Few studies have compared psychometric and clinimetric methods during the development of 2 similar versions of health-related scales. For example, during item reduction for the development of the Quality of Life After Myocardial Infarction and the QuickDASH, the clinimetric method slightly outperformed the psychometric method when both versions of each questionnaire were tested for concurrent validity and responsiveness[46,47]. Further, during the development of a health measurement scale, it was found that using the clinimetric method for constructing a heterogeneous scale measuring a complex phenomenon satisfied the criteria of psychometric method for constructing a homogeneous scale. The conclusion was that these measurement techniques complemented each other[43].

Constructing a clear questionnaire proved to be a challenge. Patients may not necessarily understand what healthcare workers perceive as a straightforward question, because patients come from diverse linguistic and cultural backgrounds. Some items were dropped because they were unclear to patients or were susceptible to misinterpretation, even though they were relevant to axSpA. Examples include sleep disturbance at the second half of the night, alternating buttock pain, and hip pain or stiffness. A strength of the TASQ-IBD compared to previous questionnaires is that it

Table 4. Test-retest reliability.

Questions	Kappa Coefficient (95% CI)	Percentage of Agreement
IBD type	0.84 (0.66–1.00)	91
Biologic agents for IBD	1.00	100
Any type of biologic agent	1.00	100
Improvement of back pain or stiffness after taking biologic agents	1.00	100
Age of onset of back pain ≤ 45 yrs	1.00	100
Time to develop	1.00	100
Presence of morning stiffness	1.00	100
Morning stiffness lasting ≥ 30 min	1.00	100
Current back pain awakening at night	0.94 (0.82–1.00)	97
Current back pain awakening at night improves upon getting up, walking, or stretching	0.95 (0.86–1.00)	97
Locations of back pain		
Neck	0.91 (0.75–1.00)	97
Upper back	1.00	100
Middle back	0.93 (0.78–1.00)	97
Lower back	1.00	100
Buttock	0.85 (0.65–1.00)	94
Improvement with physical activity	1.00	100
No improvement with rest	0.88 (0.71–1.00)	94
Improvement with NSAID	1.00	100
Family history of AS	1.00	100
Iritis/uveitis	1.00	100
Psoriasis	1.00	100
Psoriatic arthritis	1.00	100
Peripheral arthritis	0.92 (0.77–1.00)	97
Dactylitis	0.81 (0.60-1.00)	91
Heel pain	1.00	100

IBD: inflammatory bowel disease; NSAID: nonsteroidal antiinflammatory drugs; AS: ankylosing spondylitis.

includes a diagram of the back to allow patients to identify the location(s) of their back pain or stiffness. Illustrations were incorporated in recent screening questionnaires, such as the Toronto Psoriatic Arthritis Screening (ToPAS)[48], which was found to perform slightly better than the Psoriatic Arthritis Screening and Evaluation questionnaire partly because ToPAS used illustrations[49].

Assigning appropriate weighting to each item of the questionnaire is another challenge. There are many methods for scoring questionnaire items based on clinical judgment or statistical methods. Scoring of the TASQ-IBD will be explored in the validation phase, ideally when TASQ-IBD is administered to patients with IBD presenting to gastro-enterology clinics. If they have chronic back pain persisting for ≥ 3 months, they will be referred to a rheumatology clinic for further assessment. Then they will be divided into those who have a diagnosis of axSpA and those who do not. A cutoff score is to be explored in light of an appropriate sensitivity and specificity.

While this study achieved its primary goals, there are some limitations. The pilot study was conducted at the Spondylitis Clinic in a tertiary center using a convenience sample of patients with axSpA. Ideally, a sample of the target patients should be used, i.e., patients with IBD and axSpA. Although our sample size was small, all of the patients in our study represented different levels of disease severity and levels of education with a wide spectrum of axSpA disease activity. The questionnaire is limited to patients who have completed a fifth-grade education. It could potentially be too sophisticated for patients who have lower educational levels. One solution would be to use more illustrations. This questionnaire is only applicable to English-speaking and literate patients. Therefore, future research is needed for cross-cultural adaptation and validation to different languages.

The TASQ-IBD is a newly developed, self-reported questionnaire to be administered to patients with IBD who have ever had chronic back pain or stiffness that has persisted for ≥ 3 months. We used the principles of measurement science to produce a sensible and reliable CAQ that should facilitate early patient referral to rheuma-tologists and avoid delay in diagnosis of axSpA. Consequently, TASQ-IBD should be useful in assessing the prevalence of axSpA in IBD.

REFERENCES

1. Feldtkeller E, Khan MA, van der Heijde D, van der Linden S, Braun J. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. Rheumatol Int 2003;23:61-6.
2. Bennett AN, McGonagle D, O'Connor P, Hensor EM, Sivera F, Coates LC, et al. Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-B27 status in early inflammatory back pain predict radiographically evident ankylosing

APPENDIX 1. Demographic and clinical characteristics of responders and nonresponders in the reliability study. Values are means (SD) unless otherwise indicated.

	Responders, n = 33*	Nonresponders, n = 43
Age, yrs	45.7 ± 13.5	40.5 ± 11.7
Male, n (%)	27 (79.4)	31 (72.1)
Education, median †	5	4
Crohn disease, n (%)	21 (63.6)	26 (60.5)
Ulcerative colitis, n (%)	11 (33.3)	18 (41.9)
Unclassified colitis, n (%)	1 (3)	0
Age of onset of back pain, yrs	23.8 ± 9.0	25.1 ± 9.7
Disease duration, median yrs	17	13
BASDAI††	3.7 ± 2.6	4.1 ± 2.6
BASFI††	3.7 ± 3.0	3.5 ± 3.1
ESR, mm/h, median	6.5	13.0
CRP mg/dl, median	8	7

* The response of 1 patient was dropped because of incomplete information. † On a scale of 1 to 5: 1 = below grade 8; 2 = high school incomplete; 3 = high school graduate; 4 = college; 5 = university. †† On a 10-point scale. All p values are > 0.05. BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein.

spondylitis at eight years. Arthritis Rheum 2008;58:3413-8.

3. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis International Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777-83.

4. Rudwaleit M, Baeten D. Ankylosing spondylitis and bowel disease. Best Pract Res Clin Rheumatol 2006;20:451-71.

5. Palm O, Moum B, Ongre A, Gran JT. Prevalence of ankylosing spondylitis and other spondyloarthropathies among patients with inflammatory bowel disease: a population study (the IBSEN study). J Rheumatol 2002;29:511-5.

6. Scarpa R, del Puente A, D'Arienzo A, di Girolamo C, della Valle G, Panarese A, et al. The arthritis of ulcerative colitis: clinical and genetic aspects. J Rheumatol 1992;19:373-7.

7. Dekker-Saeys BJ, Meuwissen SG, Van Den Berg-Loonen EM, De Haas WH, Meijers KA, Tytgat GN. Ankylosing spondylitis and inflammatory bowel disease. III. Clinical characteristics and results of histocompatibility typing (HLA B27) in 50 patients with both ankylosing spondylitis and inflammatory bowel disease. Ann Rheum Dis 1978;37:36-41.

8. Orchard TR, Wordsworth BP, Jewell DP. Peripheral arthropathies in inflammatory bowel disease: their articular distribution and natural history. Gut 1998;42:387-91.

9. Bakland G, Nossent HC, Gran JT. Incidence and prevalence of ankylosing spondylitis in Northern Norway. Arthritis Rheum 2005;53:850-5.

10. Perez Alamino R, Maldonado Cocco JA, Citera G, Arturi P, Vazquez-Mellado J, Sampaio-Barros PD, et al. Differential features between primary ankylosing spondylitis and spondylitis associated with psoriasis and inflammatory bowel disease. J Rheumatol 2011;38:1656-60.

11. Calin A, Porta J, Fries JF, Schurman DJ. Clinical history as a screening test for ankylosing spondylitis. JAMA 1977;237:2613-4.

12. Rudwaleit M, Metter A, Listing J, Sieper J, Braun J. Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. Arthritis Rheum 2006;54:569-78.

13. Weisman MH, Chen L, Clegg DO, Davis JC Jr., Dubois RW, Prete PE, et al. Development and validation of a case ascertainment tool for ankylosing spondylitis. Arthritis Care Res 2010;62:19-27.

14. Keeling SO, Majumdar SR, Conner-Spady B, Battie MC, Carroll LJ, Maksymowych WP. Preliminary validation of a self-reported screening questionnaire for inflammatory back pain. J Rheumatol 2012;39:822-9.

15. Rudwaleit M, van der Heijde D, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535-43.

16. Sackett DL, Holland WW. Controversy in the detection of disease. Lancet 1975;2:357-9.

17. Hart FD, Robinson KC, Allchin FM, Maclagan NF. Ankylosing spondylitis. Q J Med 1949;18:217-34.

18. Moller P, Vinje O, Kass E. How does Bechterew's syndrome (ankylosing spondylitis) start? Interrelations between onset of symptoms. Scand J Rheumatol 1983;12:289-98.

19. Gran JT. An epidemiological survey of the signs and symptoms of ankylosing spondylitis. Clin Rheumatol 1985;4:161-9.

20. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361-8.

21. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34:1218-27.

22. Amor B, Dougados M, Mijiyawa M. Criteria of the classification of spondylarthropathies. Rev Rhum Mal Osteoartic 1990;57:85-9.

23. Rudwaleit M, van der Heijde D, Landewe R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Ann Rheum Dis 2011;70:25-31.

24. Zeidler H, Mau R, Mau W, Freyschmidt J, Majewski A, Deicher H. Evaluation of early diagnostic criteria including HLA-B27 for ankylosing spondylitis in a follow-up study. Z Rheumatol 1985;44:249-53.

25. Feinstein AR. Clinimetrics. New Haven: Yale University Press, 1987:141-64.

26. Rowe BH, Oxman AD. An assessment of the sensibility of a quality-of-life instrument. Am J Emerg Med 1993;11:374-80.

27. Ellerton C, Davis A, Brooks D. Preliminary development and validation of a paediatric cardiopulmonary physiotherapy discharge tool. Physiother Can 2011;63:34-44.

28. O'Brien K, Solomon P, Bayoumi AM, Alexander R, King K, Murray J, et al. Assessing the sensibility of a new HIV disability questionnaire (abstract). Can J Infect Dis Med Microbiol 2010;21:71B.

29. Gladman DD, Rahman P, Cook RJ, Shen H, Zummer M, Thomson G, et al. The Spondyloarthritis Research Consortium of Canada registry for spondyloarthritis. J Rheumatol 2011;38:1343-8.

30. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.

31. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.

32. Streiner DL. A checklist for evaluating the usefulness of rating scales. Can J Psychiatry 1993;38:140-8.

33. Dillman DA, Smyth JD, Christian LM. Internet, mail, and mixed-mode surveys: the tailored design method. 3rd ed. Hoboken, New Jersey: John Wiley & Sons Inc.; 2009.

34. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96-106.

35. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill; 1994.

36. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. New York: Oxford University Press; 2008.

37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

38. Braun A, Saracbasi E, Grifka J, Schnitker J, Braun J. Identifying patients with axial spondyloarthritis in primary care: how useful are items indicative of inflammatory back pain? Ann Rheum Dis 2011;70:1782-7.

39. Braun J, van den Berg R, Baraliakos X, Boehm H, Burgos-Vargas R, Collantes-Estevez E, et al. 2010 update of the ASAS/EULAR recommendations for the management of ankylosing spondylitis. Ann Rheum Dis 2011;70:896-904.

40. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293-307.

41. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. Int J Clin Pract 2012;66:11-5.

42. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193-213.

43. Marx RG, Bombardier C, Hogg-Johnson S, Wright JG. Clinimetric

and psychometric strategies for development of a health measurement scale. J Clin Epidemiol 1999;52:105-11.

44. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.

45. Streiner DL. Clinimetrics vs. psychometrics: an unnecessary distinction. J Clin Epidemiol 2003;56:1142-5; discussion 6-9.

46. Ribera A, Permanyer-Miralda G, Alonso J, Cascant P, Soriano N, Brotons C. Is psychometric scoring of the McNew Quality of Life after Myocardial Infarction questionnaire superior to the clinimetric scoring? A comparison of the two approaches. Qual Life Res 2006;15:357-65.

47. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005;87:1038-46.

48. Gladman DD, Schentag CT, Tom BD, Chandran V, Brockbank J, Rosen C, et al. Development and initial validation of a screening questionnaire for psoriatic arthritis: the Toronto Psoriatic Arthritis Screen (ToPAS). Ann Rheum Dis 2009;68:497-501.

49. Diaconu-Popa DA, Driessen RJB, Kerkhof PCM, van Riel PLCM, de Jong EMGJ, Fransen J. Performance of two screening instruments, ToPAS and PASE, for detecting psoriatic arthritis in patients with psoriasis (abstract). Ann Rheum Dis 2009;68:656.

50. Centers for Medicare and Medicaid Services. Toolkit for making written material clear and effective, toolkit part 7: Using readability formulas. [Internet. Accessed June 27, 2013.] Available from: www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterials Toolkit/ToolkitPart07.html

The Journal of Rheumatology 2013; 40:10; doi:10.3899/jrheum.130048