Development, Sensibility and Reliability of a New Case-finding Questionnaire: the Toronto Axial Spondyloarthritis Questionnaire (TASQ) in Inflammatory Bowel Disease

by

Khalid Abdalla Ali Bin Yarouf Alnaqbi
MBBS, CCD, FRCPC, FACP, FACR

A thesis submitted in conformity with the requirements for the degree of Master of Science

Graduate Department of The Institute of Medical Science University of Toronto

©Copyright by Khalid Abdalla Ali Bin Yarouf Alnaqbi, 2012
Development, Sensibility and Reliability of a New Case-finding Questionnaire: the Toronto Axial Spondyloarthritis Questionnaire (TASQ) in Inflammatory Bowel Disease

Khalid Abdalla Ali Bin Yarouf Alnaqbi

Master of Science
Institute of Medical Science
University of Toronto
2012

Abstract

**Background:** There is an unacceptable delay in diagnosis of axial Spondyloarthritis (axSpA) especially in its early stages among patients with inflammatory bowel disease (IBD).

**Objective:** to develop a sensible and reliable questionnaire to identify undetected axSpA among IBD patients.

**Methods:** Candidate items for the questionnaire were selected on 3 domains (IBD, inflammatory back symptoms, and extra-axial features). Sensibility of the Toronto axSpA Questionnaire (TASQ) was assessed leading to drafting 18 items. Test-retest reliability study was conducted among 77 patients with established IBD and axSpA and kappa agreement coefficients were calculated for items.

**Results:** The TASQ was developed using multiple steps of sensibility assessment resulting in 16 items. Kappa coefficients ranged from 0.81 to 1.00 for all items indicating almost perfect agreement.
Conclusion: TASQ is a newly developed, sensible and reliable questionnaire that should facilitate identification and referral of IBD patients to rheumatologists and should avoid delay in diagnosis of axSpA.
Acknowledgments

I am indebted to my parents, Mariam and Abdalla Alnaqbi, for their love and support since I was born. I am also deeply indebted to my wife, Tamador Alnaqbi, for her continuous support especially during the completion of this thesis. Caring for our lovely 3 children (Mahra, Mayed, and Bader) in a faraway country from home is a great testimony that behind every great work is a woman!

I would like to express my sincere gratitude to my supervisor, Prof. Robert Inman, whose enthusiasm and commitment made a great impact in the achievement of this project. I am blessed to have him as a world expert and a mentor who takes pride in the achievements of his co-workers. This was partly why I pursued a Clinical Research Fellowship in the Spondyloarthritis Program at the University of Toronto.

A debt of my gratitude must go to the members of my thesis committee, Profs. Dafna Gladman, and Brian Feldman. They kindly guided me while my knowledge in research was growing. Thank you for being proud of me. Special thanks go to Dr. George Tomlinson for his guidance on statistical analysis of this project.

I am grateful for the Spondyloarthritis team for their valuable feedback including Laura Passalent (advanced physiotherapist), Joan Blair (clinical trial manager), Donna Young (infusion nurse coordinator), Phyllis McGee (biologic nurse coordinator), and Adele Carty (research analyst). I am also grateful for Drs. Zahi Touma and Sindhu Johnson for their valuable contributions. A great appreciation must be given to my oral defense examiners, Drs. Muhammad Asim Khan and Cheryl Rosen for their insightful comments.

Last but not least, I would like to thank the patients of the Spondylitis Clinic who took time to participate in this project and provided their constructive feedback.
# Table of Contents

Acknowledgments ........................................................................................................ iv  
Table of Contents ....................................................................................................... v  
List of Abbreviations ................................................................................................. ix  
List of Tables ............................................................................................................... x  
List of Figures .............................................................................................................. xi  
Chapter 1 Introduction and Literature Review ............................................................. 1  
  1  Spondyloarthritis ..................................................................................................... 2  
      1.1 Epidemiology and Classification Criteria for AS ............................................. 2  
      1.2 Clinical Manifestations of AS ....................................................................... 4  
      1.3 Outcome Measures in AS ............................................................................ 6  
          1.3.1 Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) .......... 6  
          1.3.2 Bath Ankylosing Spondylitis Functional Index (BASFI) ................. 7  
          1.3.3 Patient Global Assessment of Disease Activity ............................... 7  
          1.3.4 Ankylosing Spondylitis Disease Activity Score (ASDAS) ............... 7  
          1.3.5 Total Back Pain Score ....................................................................... 8  
          1.3.6 Nocturnal Back Pain Score .................................................................. 8  
          1.3.7 Ankylosing Spondylitis Quality of Life (ASQoL) ............................. 8  
          1.3.8 Duration of Morning Stiffness of the Back over Last Week ............. 8  
          1.3.9 Fatigue ................................................................................................. 8  
          1.3.10 Outcome Measures Requiring Physical Examination ..................... 9  
          1.3.11 Acute Phase Reactants (ESR or CRP) ............................................... 9  
      1.4 Primary AS and secondary AS ....................................................................... 9  
      1.5 Axial Spondyloarthritis (axSpA) ................................................................... 9  
      1.6 Management of axSpA ............................................................................... 10  
  2  axSpA in Inflammatory Bowel Disease (IBD) ......................................................... 11  
      2.1 Epidemiology of IBD ................................................................................... 11  
      2.2 History and Epidemiology of IBD-Associated Arthritis ............................ 12  
      2.3 Clinical Manifestations of IBD-Associated Arthritis ............................... 13  
      2.4 Management of IBD-Associated Arthritis ................................................... 16
3 Literature Review about Previous Attempts for Earlier Detection of axSpA

3.1 Early Descriptive Studies of Inflammatory Back pain

3.2 Development of Rome Classification Criteria of AS (1961)

3.3 New York Criteria (1966)

3.4 Studies of Calin and Colleagues (1977)

3.5 Modification of New York Criteria (1984)

3.6 Mau Criteria for Early AS (1985)

3.7 Proposal for Diagnosis of AS by Cats and Colleagues (1987)

3.8 Amor Criteria and ESSG Criteria for Classification of SpA (1990-1991)

3.9 Studies of Rudwaleit and Colleagues (2004-2006)

3.10 Development of the ASAS Criteria for axSpA (2009)

3.11 Development of Case Ascertainment Questionnaire (2010)

3.12 Summary

Chapter 2 Rationale, Hypothesis and Aims

1 Rationale

2 Hypothesis

3 Aims

Chapter 3 Methods

1 Development of Questionnaire Items

1.1 Phase 1: Conceptualization

1.2 Phase 2: Item Pool Generation

1.3 Phase 3: Scaling Responses and Instrument Format

1.4 Phase 4: Selection of Items

1.4.1 Sensibility Assessment of Questionnaire Items

1.5 Phase 5: Pilot Study

1.6 Phase 6: Sensibility Assessment of Post-Piloted TASQ Completed by the Committee Members

2 Reliability study

2.1 Concept

2.2 Population of Interest and Sampling Method

2.3 Sample Size Calculations for Test-Retest Reliability Study

3 Statistical Analyses
3.1 Statistical Analysis for the Sensibility Assessment ........................................ 58
3.2 Statistical Analysis for Patients in the Pilot Study ...................................... 58
3.3 Statistical Analyses for Responders and Non-Responders .......................... 59
3.4 Coding the Questionnaire’s Responses ................................................. 59
3.5 Statistical Analysis for Test-Retest Reliability ......................................... 60

Chapter 4 Results ............................................................................................. 61
1 Selection of Items Using Sensibility Assessment ........................................ 62
2 Pilot study .................................................................................................... 65
   2.1 Stage 1 .................................................................................................. 67
   2.2 Stage 2 .................................................................................................. 68
   2.3 Stage 3 .................................................................................................. 70
3 Second Sensibility Assessment Completed by the Committee Members .......... 70
4 Version 4 of the Questionnaire ................................................................... 72
5 Description of the Items in the Final Questionnaire .................................... 72
   5.1 Inflammatory Bowel Disease (IBD) ...................................................... 74
   5.2 Inflammatory Back Symptoms ........................................................... 74
   5.3 Extra-axial manifestations .................................................................. 77
6 Reliability study ............................................................................................ 78
   6.1 Response Rate of Reliability Questionnaires ....................................... 78
   6.2 Missing Answers on the Questionnaire .............................................. 78
   6.3 Comparison between Responders and Non-Responders ...................... 79
   6.4 Descriptive Statistics of the Items ....................................................... 81
   6.5 Test-Retest (Intra-rater) Reliability ...................................................... 86

Chapter 5 Discussion ....................................................................................... 88
1 Questionnaire Development ....................................................................... 89
2 Clinimetric Measures in the Literature ..................................................... 90
3 Sensibility Assessment of TASQ .................................................................. 92
   3.1 Comprehensibility (Transparency) ....................................................... 92
   3.2 Content Validity .................................................................................. 93
   3.3 Feasibility ............................................................................................ 94
      3.3.1 Acceptability ............................................................................... 94
      3.3.2 Readability .................................................................................. 94
Chapter 4 Reliability .......................................................................................................................... 97
  4.1 Rating Responses of the Questionnaire Items ........................................................................ 97
  4.2 Test-Retest Reliability ............................................................................................................. 98
Chapter 5 Recent Questionnaire for Back Pain (April 2012) .......................................................... 99
Chapter 6 Conclusions ..................................................................................................................... 101
Chapter 7 Limitations and Future Directions ................................................................................ 103
  1 Development ............................................................................................................................. 104
  2 Future Validation Studies .......................................................................................................... 105
Bibliography ...................................................................................................................................... 107
Appendix A: Consent to participate in a research study ................................................................. 117
List of Abbreviations

AS: Ankylosing Spondylitis
ASAS: Assessment of SpondyloArthritis international Society
axSpA: axial SpondyloArthritis
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Disease Functional Index
CI: Confidence Interval
CRP: C-Reactive Protein
ESR: Erythrocyte Sedimentation rate
ESSG: European Spondyloarthropathy Study Group
IBD: Inflammatory Bowel Disease
IBP: Inflammatory Back Pain
LR: Likelihood Ratio
MLBP: Mechanical Low Back Pain
MRI: Magnetic Resonance Imaging
NSAIDs: Non-Steroidal Anti-Inflammatory Drugs
OR: Odds Ratio
SpA: Spondyloarthritis
SPARCC: Spondyloarthritis Research Consortium of Canada Registry for Spondyloarthritis
TASQ: Toronto Axial Spondyloarthritis Questionnaire
TNF: Tumor Necrosis Factor
TWH: Toronto Western Hospital
List of Tables

Table 1. Summary of the characteristics of articular involvement in IBD patients............... 15
Table 2. Mau criteria for early diagnosis of AS (1985). .................................................. 26
Table 3. Proposal for diagnostic criteria of AS (1987) .................................................... 27
Table 4. Summary of diagnostic tests for individual parameters used for the diagnosis of early axSpA in patients with chronic back pain lasting ≥ 3 months................................. 33
Table 5. Frequencies and positive likelihood ratios of IBP features in AS and MLBP patients .. 35
Table 6. Checklist of sensibility assessment with definitions............................................. 51
Table 7. Questionnaire for sensibility assessment ............................................................ 53
Table 8. Sensibility assessment for the pre-piloted TASQ completed by the committee members ................................................................. 63
Table 9. Stages of sensibility assessment of TASQ by patients ........................................ 66
Table 10. Comparison of sensibility assessments completed by the committee members for the pre-piloted and post-piloted TASQ ................................................................. 71
Table 11. Demographic and clinical characteristics of responders and non-responders. Values are means (SD) unless otherwise indicated................................................................. 80
Table 12. Summary of the responses for features of axSpA among patients with axSpA and IBD ......................................................................................................................... 82
Table 13. Results of the test-retest reliability analysis ......................................................... 87
List of Figures

Figure 1. Rome criteria for AS (1961)........................................................................................................... 19
Figure 2. New York criteria for AS (1966).................................................................................................. 21
Figure 3. Modified New York criteria for AS (1984).................................................................................. 24
Figure 4. Amor classification criteria for spondyloarthritis (1990)............................................................ 29
Figure 5. The European Spondyloarthropathy Study Group (ESSG) classification criteria for spondyloarthritis (1991) ................................................................................................................... 31
Figure 6. ASAS classification criteria for axial Spondyloarthritis (axSpA)................................................. 38
Figure 7. Stages of development of the TASQ ............................................................................................ 46
Figure 8. Description of the domains of the TASQ .................................................................................... 73
Figure 9. The Toronto Axial Spondyloarthritis Questionnaire (TASQ)....................................................... 85
Chapter 1

Introduction and Literature Review
1 Spondyloarthritis

Spondyloarthritis (SpA) encompasses a heterogeneous group of idiopathic inflammatory disorders that includes ankylosing spondylitis (AS), psoriatic arthritis, reactive arthritis, inflammatory bowel disease-associated arthritis (or enteropathic arthritis) and undifferentiated spondyloarthritis (SpA). They share common features including age of symptom onset in young or middle age, characteristic clinical features (inflammatory back pain, peripheral arthritis, enthesitis), familial aggregation, sero-negativity (negative rheumatoid factor), and genetic predisposition (such as HLA-B27). While the disease pathogenesis remains unknown, the inflammatory processes in these disorders likely involve tumor necrosis factor (TNF)-α in a central role. The prevalence of SpA as a group as defined by the European Spondyloarthropathy Study Group (ESSG) criteria has been reported in the ranges of 0.3 – 1.4% in Europe and the United States (1, 2).

1.1 Epidemiology and Classification Criteria for AS

AS is the most common disease among the SpA clinical subsets. It is characterized by progressive structural damage of the spine (spondylitis, spondylodiscitis, and facet joints arthritis) and the sacroiliac joints. Establishing the true prevalence of AS is challenging due to the use of different criteria to define the disease in many studies, the characteristics of the target population, and the variability in the prevalence of HLA-B27 amongst different racial groups. In addition, differing methodologies in various studies contributes to the difficulty in making cross-study comparisons. The most widely used classification criteria in the recent literature for establishing definitive AS are the modified New York 1984 criteria (3).

Classic AS often starts in the third decade of life, with a male predominance (male to female ratio approximating 2:1). Late-onset AS has been described, but is uncommon. It has been found that 5% of patients who present to primary care physicians with chronic back pain have inflammatory back pain (IBP) secondary to SpA (4). It is estimated that
3.8% of patients with IBP who present to chiropractors have AS (5). The higher prevalence generally reflects populations with a higher prevalence of HLA-B27. In AS patients, HLA-B27 can be found in 80-95%, but is observed less often than other SpA diseases. The prevalence of HLA-B27 ranges from 8% in Caucasians of Western European extraction to 10-16% of Scandinavian and Eastern European countries (1). The prevalence of AS, as defined by the modified New York criteria, ranges from 0.24% to 0.55% (1, 6-9). The highest prevalence of HLA-B27 has been found in 53% of Pawai tribe in the highlands of Papua New Guinea, in 50% of Canadian Haida Indians of the Queen Charlotte in British Columbia, and in 40% of the Chukotka Eskimos in Siberia (Eastern Russia) (10-12). Population studies on the incidence of AS are uncommon. A Norwegian study found an annual incidence of primary AS (as defined by the modified New York criteria) of 7.26 per 100,000, while an age-adjusted mean annual incidence rate of northwest Greece was 1.5 per 100,000 (13, 14). The frequency of AS is very low in sub-Saharan Africans, Arabs, East Asians (e.g. Japan and China), and Australian unmixed aboriginals reflecting the rarity of HLA-B27 in these populations (1).

There is an unacceptable delay in diagnosis of AS of 8 to 11 years (15) and this is partly due to poor recognition of the disease in its early stages. In fact, definitive radiographic sacroiliitis may not appear until approximately 8 years after symptom onset, and this may represent a late finding in chronic relapsing or continuous inflammatory phase of the disease (15, 16). Risk factors for radiographic progression are male gender, smoking, active sacroiliitis on magnetic resonance imaging (MRI), and high C-reactive protein (CRP) level (17). New classification criteria (from the Assessment of SpondyloArthritis international Society or ASAS) have been recently established and validated to identify patients with early axial SpA (axSpA).

Mortality in AS was recently shown to be increased in men with chronic disease, and the most common cause of death is cardiovascular disease which, in ascending order, includes coronary disease, valvular disease, thoracic aortic aneurysm. Other causes of mortality were equally distributed among the following: chronic renal failure, cardiomyopathy, stroke, atrioventricular conduction abnormality, pseudomembranous colitis, and chronic alcoholism (18).
1.2 Clinical Manifestations of AS

The most common manifestation of AS is inflammatory back pain (IBP) especially before the age of 40. Details on characteristics of IBP will be discussed in Section 1.3. The pain can be accompanied by variable degrees of other features such as peripheral arthritis, dactylitis, and enthesitis.

Some expert rheumatologists consider the hip and shoulder included in the axial joints. Hip involvement can be the presenting symptoms in children and adolescents. It is usually a marker of severe disease that may be a harbinger of a future arthroplasty. **Peripheral arthritis** is generally uncommon in early-onset AS, but is more common in women, late-onset AS (> 50 years) and juvenile-onset AS (< 16 years) (19-21). This type of arthritis is often seronegative (negative rheumatoid factor), asymmetrical with predominant lower limb involvement, non-erosive and non-deforming. However, it can rarely be chronic and erosive.

**Dactylitis** is a characteristic finding in SpA. It occurs secondary to flexor tenosynovitis of the finger or toe which is usually associated with joint inflammation. It often appears like a “sausage-like” swelling which can be painful.

**Enthesitis** is inflammation at the site of insertion of a tendon, fascia or ligament to bone. It is considered another hallmark feature of SpA. It can affect any enthesis site but most commonly in the lower limbs. The most common site is the heel where Achilles tendon and plantar fascia are inserted to the calcaneus. Other affected sites can be the insertion site of patellar tendon into the tibial tubercle, insertion of quadriceps tendon into the superior pole of the patella, insertions on the ischial tuberosities, the femoral greater trochanteric areas, the iliac crests, the humeral epicondyles, and supraspinatus insertion (22).
Extra-articular manifestations can affect different organs including the eye (uveitis), bowel (macroscopic and microscopic colitis), skin (psoriasis), lung, heart, kidney, and bone.

**Uveitis** is inflammation of the middle layer of the eye (uveal tract). It is estimated that 15% of all types of uveitis are associated with AS, and this likelihood increases to 30-50% in acute anterior uveitis or iritis. If the patient is HLA-B27 positive, this increases the likelihood even further to 84–90% (23). Uveitis, particularly iritis, is considered to be the most common extra-axial manifestation of AS. A recent systematic review found a prevalence of uveitis in AS to be 33.2% for mean disease duration of 17.7 years. It should be noted that all the included AS patients that study were diagnosed by the Rome 1961 criteria, the New York 1966 criteria, or the modified New York 1984 criteria (24). Iritis is usually acute, anterior, unilateral, and often recurrent especially in HLA-B27 positive AS patients. Patients often complain of painful red eye with photophobia, tearing, and blurred vision. If treated immediately, the episode generally resolves over 2 to 3 months without major complications.

**Subclinical gut inflammation** as confirmed previously by ileocolonoscopy demonstrating inflammation on biopsy, was detected in 24-49% of AS patients, in whom 50-60% had evidence of abnormal histological analysis. There is a correlation between gut inflammation and peripheral arthritis, which is typically an oligoarthritis. Subclinical gut inflammation, especially in chronic lesions, can progress to overt inflammatory bowel disease (IBD) in 6.5% of patients with SpA despite absence of bowel symptoms at the initial ileocolonoscopy. The prevalence of IBD, as an extra-axial manifestation of AS, is 5-10% (25). Therefore, there is a low threshold for screening ileocolonoscopy in AS patients complaining of bowel symptoms.

**Psoriasis** can occur concurrently in 10-25% of AS patients. **Cardiac involvement** including aortitis, aortic regurgitation and conduction disturbances occurs in up to 9% of long-standing AS disease. It is largely due to post-inflammatory fibrotic reaction of the aortic wall and myocardial wall. The incidence of **renal abnormalities** has been reported in 10-35% of AS patients with amyloidosis being the most common feature. Lung
abnormalities are uncommon with an incidence of 1% in AS patients. These can manifest as progressive bilateral apical fibrosis, restrictive lung disease secondary to rigid chest wall, bronchiectasis, interstitial changes, and ground-glass opacities. **Diffuse osteoporosis and vertebral fractures** are the most common bone manifestations of AS (26, 27).

A multi-center Spanish national cohort recently found a 3.5% prevalence of late-onset AS (≥ 50 years) which did not differ from early-onset AS in terms of sex distribution, family history of SpA, HLA-B27 positive status or back pain. Late-onset AS patients had more involvement of cervical spine and arthritis of upper and lower limbs at onset of the disease. In terms of extra-axial manifestations, those patients were significantly different with less uveitis but more cardiac complications compared to early-onset AS patients (19).

Structural damage in AS is defined as new bone formation which can progress to complete or incomplete ankylosis (fusion) of the joints, and can best be visualized on X-ray or computed tomography (CT) scan. Late-stage AS can manifest with severe restriction in spinal mobility, postural changes as a result of spinal deformity, and functional disability.

### 1.3 Outcome Measures in AS

#### 1.3.1 Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

BASDAI is a validated self-administered questionnaire that consists of 6 items on a 10-cm visual analog scale for each item. It assesses level of fatigue, spinal and peripheral joint pain, localized tenderness, level and duration of morning stiffness. An average score of all the items is calculated, and generally the higher the score, the more severe the disease is (28). A score ≥ 4 has been arbitrarily recognized in clinical trials as an indicator of active disease.
1.3.2 Bath Ankylosing Spondylitis Functional Index (BASFI)

BASFI is a validated self-administered questionnaire that consists of 10 items that yields a composite score ranging from 0 to 10 using 10-cm visual analog scale for each question. It assesses the degree of mobility and functional ability. A higher BASFI score correlates with reduced functional ability (29).

1.3.3 Patient Global Assessment of Disease Activity

The patient global assessment of disease activity is the patients’ assessment of how active their spondylitis was on average during last week. It is marked on a 10-cm visual analog scale with 0 representing “not active” on the left-hand box and 10 representing “very active” on the right-hand box (30).

1.3.4 Ankylosing Spondylitis Disease Activity Score (ASDAS)

ASDAS is a newly developed and validated outcome measure in axSpA trials (31, 32). It combines patient-reported symptoms (back pain, duration of morning stiffness during last week, patient global disease activity during last week, and peripheral pain/swelling on a visual analog scale from 0 to 10 cm or on a 10-point numerical rating scale) with an inflammatory serum marker i.e. CRP or Erythrocyte Sedimentation rate (ESR). ASDAS-ESR is a continuous measure with a score range between 0 and 6.9. Inactive disease is defined as a score of < 1.3, a moderate disease activity score is 1.3 – 2.0, a high disease activity score is 2.1 – 3.5, and very high disease activity is > 3.5. ASDAS-CRP is also a continuous measure with a score range between 0 and 6.7. The definition of disease activity in ASDAS-CRP is similar to the ASDAS-ESR.
1.3.5 Total Back Pain Score

Total back pain score is the participant’s assessment of how much pain they have in their back they have due to AS at any time. It is marked on a 10-cm visual analog scale with 0 representing “no pain” on the left-hand box and 10 representing “most severe pain” on the right-hand box (30).

1.3.6 Nocturnal Back Pain Score

It is the subject’s assessment of how much pain they have in their back they have due to AS at night. It is marked on a 10-cm visual analog scale with 0 representing “no pain” on the left-hand box and 10 representing “most severe pain” on the right-hand box (30).

1.3.7 Ankylosing Spondylitis Quality of Life (ASQoL)

The ASQoL is a validated questionnaire to measure the impact of AS on the quality of life. Its score ranges from 0 to 18, and the higher the overall score, the greater impact on QoL (30).

1.3.8 Duration of Morning Stiffness of the Back over Last Week

1.3.9 Fatigue

Fatigue can be marked on a 10-cm visual analog scale or numerical rating scale with 0 representing “no fatigue” on the left-hand side and 10 representing “worse fatigue” on the right-hand side (33).
1.3.10 Outcome Measures Requiring Physical Examination

These outcome measures include measurement of spinal mobility (chest expansion, modified Schober’s test, occiput-to-wall distance, and lumbar lateral flexion), swollen joint count (44 joints) and enthesitis score (e.g. the Maastricht AS Enthesitis Score, Berlin Enthesitis Score, and San Francisco Enthesitis Score) (33).

1.3.11 Acute Phase Reactants (ESR or CRP)

1.4 Primary AS and secondary AS

AS can occur in isolation (primary or idiopathic AS), or associated with other types of SpA (secondary AS) including reactive arthritis, psoriatic arthritis, and enteropathic arthritis. It has been estimated that 12 - 26% of patients with reactive arthritis develop AS depending on the triggering infection and the length of follow up (27). Axial arthritis occurs in 25 - 70% of psoriatic arthritis depending on the definition used, with severe peripheral arthritis and HLA-B27 being risk factors (34, 35). A recent large, multi-center multi-ethnic study found that axial psoriatic arthritis was significantly associated with arthritis of upper and lower limbs, dactylitis, and enthesitis (36). Detailed discussion on IBD-associated AS will follow shortly.

1.5 Axial Spondyloarthritis (axSpA)

AS is considered the prototype of SpA. The newly coined and more inclusive term “axial Spondyloarthritis (axSpA)” comprises non-radiographic axSpA with predominant axial involvement but lacking the diagnostic changes in the sacroiliac joints or spine seen in classic AS. axSpA captures patients with early diagnosis that can be aided with positive
HLA-B27 and/or sacroiliitis on MRI. It should be emphasized that not all patients with non-radiographic axSpA develop end-stage AS with joint ankylosis, although they may experience the same symptomatic burden of disease as their counterparts who have classic AS. In fact, a recent observational prospective study in 769 patients with AS found complete spinal fusion occurred in 28% of patients after more than 30 years and in 43% of patients after more than 40 years (37). Using the new ASAS classification criteria, axSpA has an estimated point prevalence of 21.5% in primary care patients aged 19-45 years who present with chronic low back pain (38). This emphasizes the importance of the new term axSpA.

### 1.6 Management of axSpA

The 2010 ASAS/ EULAR (European League against Rheumatism) guidelines for the management of axSpA recommend an initial non-pharmacological approach with patient education and exercise. The recommended pharmacological approach considers full doses of non-steroidal anti-inflammatory drugs (NSAIDs) as the first-line therapy for pain and stiffness. These drugs (especially celecoxib) can improve the symptoms and reduce the structural changes of the spine when taken regularly. There is no evidence that NSAIDs accelerate the progression of subclinical gut inflammation found in AS patients to overt IBD (25). A recent study in axSpA patients did not find a correlation between NSAIDs use and bowel symptoms or elevation of fecal calprotectin (a marker in intestinal inflammation) (39).

The literature does not support the use of systemic glucocorticosteroids and disease-modifying antirheumatic drugs (DMARDs) for axial symptoms. The latter, especially sulfasalazine and methotrexate, can be used for peripheral arthritis.

TNF inhibitors are novel biological therapies that have revolutionized the management of many autoimmune inflammatory diseases including axSpA and also IBD. They should be used in patients who fail or have contraindications to NSAIDs and have persistently
active disease. The most widely used TNF inhibitors are infliximab (Remicade™), etanercept (Enbrel™), adalimumab (Humira™), and golimumab (Simponi™). Switching to a second TNF inhibitor due to primary non-response, secondary loss of efficacy, and intolerance has been shown to be beneficial. Fortunately, patients who fail 3 TNF inhibitors are very rare (40-43). Although these recommendations were focused on classic AS, randomized controlled trials have demonstrated the efficacy of TNF inhibitors in non-radiographic axSpA. However, the ability to retard structural damage in the spine has yet to be proven (44-46).

2 axSpA in Inflammatory Bowel Disease (IBD)

2.1 Epidemiology of IBD

IBD encompasses a group of idiopathic inflammatory diseases that affect the gastrointestinal (GI) system and comprises predominantly ulcerative colitis and Crohn’s disease, and to a lesser extent, indeterminate colitis (47). Crohn’s disease (also known as terminal ileitis or regional enteritis) can affect any part of the GI tract “from mouth to perianal area”, predominantly the ileum and colon. Ulcerative colitis often affects the rectum and any part of the colon. Unclassified (indeterminate) colitis has a prevalence of 10-15% of all IBD depending on the definition used. It may evolve to ulcerative colitis or Crohn’s disease, or may remain unclassified (48). In a population-based Canadian study, the peak incidence of Crohn’s disease occurred between 20-29 years of age, with an overall incidence rate of 13.4 per 100,000. This incidence rate is amongst the highest in the world. The peak age of incidence of ulcerative colitis was different across the provinces but appeared to be bimodal, with an overall incidence rate of 11.8 per 100,000 (49). The prevalence of HLA-B27 in IBD patients is similar to that of general population (50, 51). However, the association of HLA-B27 with AS-associated IBD is still higher than the IBD population without AS, and therefore, HLA-B27 in IBD patients seems to increase the risk of AS development (52).
2.2 History and Epidemiology of IBD-Associated Arthritis

The first description of arthritis in IBD was published in 1929. Prior to 1950s, AS, psoriatic arthritis, reactive arthritis, and IBD-associated AS were lumped together into a unifying disease entity called “rheumatoid spondylitis” as they were considered part of the spectrum of rheumatoid arthritis. In the 1950s, the peripheral arthritis of IBD was finally separated from rheumatoid arthritis, and in 1960, the concept of IBD-associated AS was born (53, 54).

Articular involvement is the most common extra-intestinal manifestation of IBD with a prevalence rate of 16-33%. It generally occurs more commonly in patients with colonic inflammation (55, 56). Articular involvement in IBD encompasses arthralgias and arthritis (peripheral arthritis, IBP, asymptomatic radiographic sacroiliitis, and axSpA). Arthralgias, as defined by pain in any joint without swelling, occur in 8 to 30% of IBD patients (55).

Peripheral arthritis is more common and has been reported in 2.8 – 30.6% of IBD patients (57). The prevalence of IBP as defined by the Calin criteria (for detailed description, see section 3.4 Studies of Calin and Colleagues (1977) in IBD ranges between 5.2 to 30% (50, 58-61). Isolated asymptomatic sacroiliitis in IBD has been reported in 2–18% on pelvic X-ray (50, 58, 62, 63), and in 13.6–33.3% on CT scan especially in Crohn’s disease (63-66). This unique type of axial abnormality is weakly associated with HLA-B27. However, it is strongly associated with CARD15 polymorphism which is a susceptibility gene associated with Crohn’s patients (51). Gastroenterologists regard AS as a significant extra-intestinal manifestation of IBD. The prevalence of IBD-associated AS (i.e. AS diagnosed in IBD patients) using the modified New York criteria ranges from 3.1% to 10% depending on whether the study was population- or hospital-based (50, 58, 60, 67). The 4-year prevalence of IBD-associated AS in the Spondylo-Arthritis Research Consortium of Canada (SPARCC) registry is 6.7% by the modified New York criteria
The association between HLA-B27 and IBD-associated AS is not as strong as in primary AS patients (60% and 85-90% respectively) (26, 69).

The age of onset of IBD-associated AS is similar to that of primary AS (14, 15). In contrast, a large multi-center population-based study has recently found an earlier age of onset in patients with primary AS compared to IBD-associated AS (36). Population-based studies have found no association between AS and the type of IBD. Studies confirmed the male predominance in this disease (50, 60, 63). These findings were also demonstrated in a population-based Canadian study of patients with IBD for more than 10 years who made at least 5 clinic visits, when the diagnosis of AS was established using administrative data based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) code although the diagnostic criteria were not mentioned in this study (70). The delay in diagnosis of IBD-associated AS is similar to that of primary AS (14, 50). In both diseases, patients equally suffered with respect to work disability, reduced quality of life, and presence of active disease (36).

2.3 Clinical Manifestations of IBD-Associated Arthritis

Axial symptoms in IBD-associated arthritis are not different from those in primary AS with IBP being the predominant feature. These symptoms generally do not run in parallel to bowel activity.

Symptoms of axSpA can precede IBD symptoms in 31 to 50% while symptoms of IBD and axSpA can occur simultaneously in 15 to 40% (50, 59, 71, 72). In a preliminary IBD cohort study conducted at University Health Network in Toronto, 39% of IBD patients experienced IBD symptoms prior to the onset of axSpA symptoms, 52% experienced bowel symptoms after the onset of AS symptoms, and 9% had symptoms of IBD and AS concurrently (73).

Extra-axial manifestations of IBD-associated arthritis include peripheral arthritis, enthesitis, dactylitis, uveitis, and psoriasis. Peripheral arthritis occurs more commonly in AS-associated IBD than among IBD patients without AS (50, 51). The Oxford IBD
group led by Orchard has classified the peripheral arthritis without axial involvement into oligo- or pauci-articular (type I) and poly-articular (type II). Both types share common characteristics with the peripheral arthritis of AS in that the arthritis is often sero-negative, non-erosive and non-deforming which progresses in only a minority of cases (10%) to chronic, erosive arthritis. This classification has been circulating in the gastroenterology literature but less in the rheumatology literature (59). Table 1 summarizes the characteristics of the articular involvement in IBD patients.
<table>
<thead>
<tr>
<th></th>
<th>Oligo-articular arthritis (Type I)</th>
<th>Poly-articular arthritis (Type II)</th>
<th>Axial (AS)</th>
<th>Axial (asymptomatic sacroiliitis)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence</strong></td>
<td>4 – 8%, mostly in Crohn’s disease</td>
<td>3%</td>
<td>3.1 – 10%</td>
<td>X-ray: 2 – 18%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT scan: 13.6-33.3%</td>
</tr>
<tr>
<td><strong>Sex distribution</strong></td>
<td>males = females</td>
<td>males = females</td>
<td>males &gt; females</td>
<td>males = females</td>
</tr>
<tr>
<td><strong>HLA genotype</strong></td>
<td>–DRB1<em>03, -B35, and –B</em>27</td>
<td>HLA–B44</td>
<td>–B<em>27 and –DRB1</em>01</td>
<td>–</td>
</tr>
<tr>
<td><strong>Affected joints</strong></td>
<td>&lt; 5 large weight-bearing joints (e.g. knees, ankles)</td>
<td>≥ 5 small joints especially of the hands and feet</td>
<td>Sacroiliac joints and spine</td>
<td>Sacroiliac joints</td>
</tr>
<tr>
<td><strong>Symmetry of joint involvement</strong></td>
<td>Asymmetrical</td>
<td>Symmetrical</td>
<td>Symmetrical or asymmetrical</td>
<td>Symmetrical</td>
</tr>
<tr>
<td><strong>Natural course</strong></td>
<td>Acute, self-limiting (&lt; 10 weeks)</td>
<td>Insidious onset. Symptoms can last for months or years.</td>
<td>Often progressive</td>
<td>Often asymptomatic</td>
</tr>
<tr>
<td><strong>Association with bowel activity</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Association with other extra-intestinal manifestations</strong></td>
<td>Erythema nodosum uveitis</td>
<td>Uveitis</td>
<td>Uveitis, peripheral arthritis, enthesitis</td>
<td>Erythema nodosum, uveitis</td>
</tr>
</tbody>
</table>

Adapted from the following references (27, 50, 51, 55, 58, 59, 64, 74, 75).
Enthesitis is the most frequently affected site of peripheral inflammation in IBD patients, with Achilles tendon presenting as heel pain. Furthermore, enthesitis occurs more commonly in AS-associated IBD than amongst IBD patients without AS (50). The association between AS and uveitis has been also observed in IBD patients. Uveitis occurs more commonly in AS-associated IBD than among IBD patients without AS (50, 51). In addition, in a population-based Canadian study, iritis was the most prevalent extra-intestinal manifestation in IBD over a 10-year period, especially in women with ulcerative colitis although this study utilized exclusively an administrative database and excluded patients with diagnosis of peripheral arthritis (70).

2.4 Management of IBD-Associated Arthritis

Some studies have found that NSAIDs use may exacerbate IBD symptoms and may result in ulceration of the GI tract (76-78). However, their use is not absolutely contraindicated in IBD patients. This was evident from 2 large randomized controlled trials which concluded that cyclooxygenase-2 (COX-2) inhibitors use might actually be safe in some patients with IBD (79, 80). A recent large retrospective observational study of 629 IBD patients, who were followed for 1315 clinic visits, has found no correlation between low-dose NSAIDs and reactivation of IBD. Although high-dose NSAID use was associated with a higher disease activity index score among Crohn’s disease with colonic inflammation, this was not reflected by a significant increase in disease activity (81). In experimentally-induced colitis, COX-2 inhibitors improved the severity of colitis (82). In summary, a subset of IBD patients may tolerate these drugs and perhaps even benefit from these drugs. The general recommendation is to use caution when prescribing NSAIDs in IBD with close monitoring for possible GI adverse events. DMARDs are used mainly for managing IBD symptoms and peripheral arthritis, but not for axial symptoms.

The main biologic agents used in IBD are infliximab (Remicade™) and adalimumab (Humira™). They have been found to induce and maintain remission in moderate to severely active Crohn’s disease and ulcerative colitis (83-85). They have been also used
in treating both active axSpA and IBD, primarily infliximab (44, 86, 87). Although etanercept (Enbrel™) can treat IBD-associated arthritis, its efficacy is marginal in treating IBD. Furthermore, etanercept-related exacerbation of IBD has been increasingly reported. Therefore, the general recommendation is to avoid etanercept in the management of IBD. If IBD patients have undiagnosed axSpA which has been masked by the use of these agents, axSpA symptoms can be expected to be significantly improved or even resolve.

3 Literature Review about Previous Attempts for Earlier Detection of axSpA

3.1 Early Descriptive Studies of Inflammatory Back pain

It proves a challenge to separate previous attempts to define the characteristics of IBP from the history of the classification criteria for AS and SpA. This section will provide essential historical background in order to understand the inherent challenge in the proper definition of IBP.

The earliest description of the clinical features of AS was reported in 1691 by an Irish physician, Dr. Bernard Conner. I am not aware of an English translation of his MD thesis. In the 19th century, publications in the medical literature enhanced knowledge of AS with more detailed clinicopathologic description of AS from such authors as Charles Fagge, A. Strumpell, Vladimir Bechterew, Pierre Marie, and Bernard Sachs. Prior to 1950s, AS was thought to be a variant of rheumatoid arthritis, hence it was called “rheumatoid spondylitis”. This term encompassed AS, psoriatic arthritis, reactive arthritis and IBD-associated AS. With the discovery of rheumatoid factor, the realization that AS has specific clinical features that distinguished it from rheumatoid arthritis, and the utilization of X-rays, it was evident that AS was not related to rheumatoid arthritis. The American Association of Rheumatism changed the nomenclature to AS in 1963 (88).
There have been several attempts to increase the sensitivity of questions for IBP in order to capture patients with early axSpA disease. Hart and colleagues provided the earliest descriptive features of IBP in 1949: pain first occurring most commonly in the lumbo-sacral area, and less commonly in mid-lumbar, mid-thoracic, cervical spine, or buttocks. This was associated with morning stiffness that improved during the day and/or with frequent exercise but was aggravated with immobility. In contrast, some patients perceived rest as an alleviating factor for their back pain. Some had to wake at night to exercise their back to avoid morning stiffness. Some patients complained of chest tightness that resulted in dyspnea. A young patient used to practice playing bagpipes to overcome his thoracic stiffness. These symptoms led many patients to suffer from chronic anxiety. The diagnosis up to this point in time was made based on clinical background and radiological criteria (sacroiliitis on X-ray) without meeting specific criteria (89, 90).

3.2 Development of Rome Classification Criteria of AS (1961)

As the literature about AS started to grow, a group of rheumatologists convened in Rome in 1961 and agreed to formulate new criteria for AS, the Rome criteria, which were based on clinical data that relied on history (low back pain and stiffness for more than 3 months not relieved by rest, pain and stiffness in the thoracic region, history or evidence of iritis or its sequelae), physical examination (limited motion in lumbar spine and limited chest expansion), and radiographic sacroiliitis (without grading). The diagnosis of definitive AS was based on bilateral sacroiliitis and any clinical criteria, or, fulfillment of 4 out of 5 clinical criteria (Figure 1). Some historical items were removed from the final criteria including buttock pain that awakens patients at night, and good response to phenylbutazone (an NSAID) which was found to improve IBP (91).
**Figure 1. Rome criteria for AS (1961)**

<table>
<thead>
<tr>
<th>Clinical:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Low back pain and stiffness &gt; 3 months not relieved by rest.</td>
</tr>
<tr>
<td>2. Pain and stiffness in thoracic region</td>
</tr>
<tr>
<td>3. Limited motion in lumbar spine</td>
</tr>
<tr>
<td>4. Limited chest expansion</td>
</tr>
<tr>
<td>5. History or evidence of iritis or its sequelae.</td>
</tr>
<tr>
<td>6. Radiograph showing bilateral sacroiliac changes characteristic of AS (this would exclude bilateral osteoarthrosis of sacroiliac joints)</td>
</tr>
</tbody>
</table>

Definitive AS = Bilateral sacroiliitis (radiograph) and ≥ 1 clinical criteria, *or*, 4/5 clinical criteria.

The Rome criteria were validated in Pima Indians with definitive AS, but some items have been found to have poor diagnostic utility. For example, the item of historical iritis was found to be infrequent with low sensitivity (4%) and difficult to be elicited with certainty on history. Another historical item “low back pain lasting 3 months or more and not relieved by rest” was difficult to apply and had low sensitivity (30%). Similarly, the item on “pain and stiffness in the thoracic region” had low sensitivity (15%).

3.3 New York Criteria (1966)

The group of rheumatologists convened again in New York in 1966 and modified these criteria (Figure 2). The iritis was removed, the description of back pain was changed to (a history or the presence of pain at the dorsolumbar junction or in the lumbar spine), items on physical examination were modified, and radiographical description of sacroiliitis was clarified and graded. This has led to formulation of the New York criteria for AS (1966).

In 1973, Moll and Wright validated the New York criteria but found focusing only on the item of historical back pain to be too sensitive and too non-specific, with the concern that this might lead to including more false positive patients (92).
**Figure 2. New York criteria for AS (1966)**

*Clinical criteria for AS:*

1. A history or the presence of pain at the dorsolumbar junction or in the lumbar spine.
2. Limitation of motion of the lumbar spine in all 3 planes: anterior flexion, lateral flexion, and extension.
3. Limitation of chest expansion to 1 inch (2.5 cm) or less, measured at the level of the 4th intercostal space.

*X-ray grading:*

0 = normal.

1 = suspicious.

2 = abnormal with erosions or sclerosis.

3 = unequivocal abnormal, moderate or advanced sacroiliitis showing one or more: erosions, sclerosis, widening, narrowing, and partial ankylosis.

4 = total ankylosis.

*Application of these criteria:*

**Definitive AS:**

- Grade 3-4 bilateral sacroiliitis AND 1 or more clinical criteria, OR
- Grade 3-4 unilateral sacroiliitis (or grade 2 bilateral sacroiliitis) AND clinical criterion 1 (or both clinical criteria 2 and 3).

**Probable AS:** grade 3 or 4 bilateral sacroiliitis without a clinical criterion.

* When sacroiliitis is present, the following variants of AS should be designated individually: rheumatoid arthritis, psoriasis, ulcerative colitis or regional ileitis, Reiter’s syndrome, and juvenile RA.

* Individuals with conditions e.g. fluorosis, hypophosphatemic osteomalacia, brucellosis, and familial Mediterranean fever, which may confuse the picture of AS, should be identified in tabulations for AS and should be listed separately.

3.4 Studies of Calin and Colleagues (1977)

Calin and colleagues pursued the quest in 1977 to identify AS patients earlier and developed the first screening questionnaire which was administered to patients with known AS, orthopedic patients (with normal sacroiliac joints and negative HLA-B27), and healthy individuals. The aim of this symptom-based questionnaire was to identify characteristics of IBP in AS. It was developed in English and had 17 items with binary response options except for one item asking about the age of onset of back pain. The final questionnaire had 5 items which were thought to represent the characteristics of IBP (excluding neck pain) in AS patients compared to mechanical or non-specific back pain: 1) age of onset of back pain occurring less than 40 years, 2) insidious onset of pain, 3) persistent pain for 3 months or more, 4) association with morning stiffness, and 5) improvement with exercise. In this study, the item “no improvement with rest” was significantly associated with AS compared to non-specific back pain, but was not included in the final Calin criteria. Similarly, the item “awakening at night for back pain” was significantly associated with AS compared to controls of healthy people, but was not included in the final criteria. The specificity of fulfilling 4 out of 5 criteria was 85% and the sensitivity was 95%. However, certain limitations deserve to be mentioned.

First, the optimal diagnostic parameters could not be validated in 2 subsequent studies and the specificity was shown to be moderate (75%) with the sensitivity low (23-38%). In one of these studies, a self-administered questionnaire (in German) contained 11 items in which 9 items were related to back pain, and 2 items were related to pain in the interscapular area and anterior chest. Some of the items were translated from Calin criteria. Second, the term “insidious onset” remained ambiguous. In fact, a recent study found that this term differentiates AS patients from those with mechanical back pain in only one fourth of those with AS if the term indicates an onset of back pain within a year (93, 94). Third, the Calin criteria alone cannot be used to classify or diagnose patients with radiographic or non-radiographic axSpA. In fact, a population-based study conducted by Calin’s group validated their questionnaire in 10,150 patients with back pain who worked in an industrial complex. Of the 1880 respondents, 367 (20%) fulfilled Calin criteria of whom only 16 (4.4%) had AS.
In 1983, Møller and colleagues found that 73% of AS patients had insidious onset of back pain, 15% had sudden onset of back pain (minutes – hours), and 12% started with back pain developing over a few days to weeks (95). Another study published in the same year showed that the presence of morning stiffness of the back had a sensitivity of 93% and a specificity of 43% for early AS (96).

### 3.5 Modification of New York Criteria (1984)

In 1984, a modification of the New York criteria for AS was published (Figure 3). The IBP was defined as low back pain and stiffness for more than 3 months, which improves with exercise but not with rest. The items on physical examination were modified again. Diagnosis of definitive AS requires at least 1 clinical feature and 1 radiographic abnormality of the anteroposterior pelvic X-ray. The clinical features include IBP, restriction in lumbar spine motion, and restriction in chest expansion. The radiological criteria include either bilateral sacroiliitis grade ≥ 2 or unilateral sacroiliitis grade ≥ 3 (3).
Figure 3. Modified New York criteria for AS (1984)

1. Clinical criteria:
   a. Low back pain and stiffness for more than 3 months which improves with exercise, but is not relieved by rest.
   b. Limitation of motion of the lumbar spine in both the sagittal and frontal planes.
   c. Limitation of chest expansion relative to normal values correlated for age and sex.

2. Radiological criterion:
   Sacroilitis grade $\geq 2$ bilaterally or grade 3-4 unilaterally

Definite ankylosing spondylitis if the radiological criterion is associated with at least 1 clinical criterion.


From the ASAS slide collection, http://www.asas-group.org, open access.
In 1985, Gran proposed addition of new characteristics to further refine IBP: back pain that awakens at night (specificity 53%, sensitivity 71%) with getting out of bed (specificity 79%, sensitivity 65%), not relieved by lying down (specificity 49%, sensitivity 80%), and associated with morning stiffness for 30 minutes or more (specificity 59%, sensitivity 64%) (97).

3.6 Mau Criteria for Early AS (1985)

In the French literature, the criteria of Sèze and Lequesne were formulated in 1961. Baudoin and Landureau modified these criteria in 1979. A third modification was proposed for early diagnostic criteria for AS and was published in English in 1985 (Table 2). The aim was to establish more reliable criteria for early diagnosis of AS in patients without sacroiliitis. The criteria took into account historical variables, laboratory (ESR) and radiological signs (spinal abnormalities of AS without sacroiliitis). Historical criteria include IBP (as defined by Calin), spontaneous low back pain radiating to the buttocks or back of the thighs, spontaneous thoracic pain (related to deep inspiration or coughing), peripheral arthritis or heel pain, and anterior uveitis. The performance of these criteria for diagnosis of early AS on a 5- and 10-year follow-up showed sensitivity of 82% and 88% respectively, and specificity of 68% and 59% respectively. However, the use of these criteria fell out of favor, probably due to lack of specificity (98, 99).
### Table 2. Mau criteria for early diagnosis of AS (1985).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genetic</strong></td>
<td></td>
</tr>
<tr>
<td>HLA-B27 positive</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
</tr>
<tr>
<td>Spinal pain (inflammatory type as per Calin criteria)</td>
<td>1</td>
</tr>
<tr>
<td>Lower back pain, radiating to the buttocks or the back of the thighs, spontaneous or elicited by stress tests of the sacroiliac joints.</td>
<td>1</td>
</tr>
<tr>
<td>Thoracic pain, spontaneous or produced by compression, or limited chest expansion (≤ 2.5 cm)</td>
<td>1</td>
</tr>
<tr>
<td>Peripheral arthritis or heel pain</td>
<td>1</td>
</tr>
<tr>
<td>Anterior uveitis</td>
<td>1</td>
</tr>
<tr>
<td>Limited motion of cervical or lumbar spine in all planes</td>
<td>1</td>
</tr>
<tr>
<td><strong>Laboratory</strong></td>
<td></td>
</tr>
<tr>
<td>Elevated ESR:</td>
<td></td>
</tr>
<tr>
<td>Age &lt; 50: male &gt; 15 mm/hr, female &gt; 25 mm/hr.</td>
<td>1</td>
</tr>
<tr>
<td>Age ≥ 50: male &gt; 20 mm/hr, female &gt; 30 mm/hr.</td>
<td>1</td>
</tr>
<tr>
<td><strong>Radiological</strong></td>
<td></td>
</tr>
<tr>
<td>Spinal signs: syndesmophytes / squaring phenomenon / barrel-shaped vertebrae / involvement of apophyseal or costovertebral joints</td>
<td>1</td>
</tr>
</tbody>
</table>

A total count of ≥ 3.5 points indicates early AS.

3.7 Proposal for Diagnosis of AS by Cats and Colleagues (1987)

Cats and colleagues proposed diagnostic criteria of AS based 4 clinical and 1 radiological criteria (100). Table 3 summarized these criteria. The proposal also emphasized stating associated diseases including psoriasis, IBD, Reiter’s disease, reactive arthritis, and juvenile chronic arthritis.

Table 3. Proposal for diagnostic criteria of AS (1987)

<table>
<thead>
<tr>
<th>Clinical</th>
<th>1. Insidious onset of low back pain lasting ≥ 3 months, associated with morning stiffness, improves with exercise, with age of onset ≤ 45 years.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Existence of any of the following in a relative (1st or 2nd degree) of an AS patient, or in HLA-B27 patient, and age of onset ≤ 45 years:</td>
</tr>
<tr>
<td></td>
<td>a. Recurrent chest pain and thoracic stiffness of unknown cause.</td>
</tr>
<tr>
<td></td>
<td>b. Chronic or recurrent enthesopathy and/or oligo-arthritis with negative rheumatoid factor and antinuclear antibody.</td>
</tr>
<tr>
<td></td>
<td>c. Unilateral acute anterior uveitis.</td>
</tr>
<tr>
<td></td>
<td>3. Restricted mobility of the lumbar spine in 2 planes (frontal and sagittal) excluding other causes (e.g. infectious spondylitis, neurological diseases, herniated or degenerative disc diseases, diffuse idiopathic skeletal hyperostosis)</td>
</tr>
<tr>
<td></td>
<td>4. Restricted chest expansion relative to normal values corresponding to age and sex, excluding other causes (e.g. severe juvenile or adolescent deformity of thoracic spine, congenital heart disease, or severe pulmonary diseases).</td>
</tr>
<tr>
<td>Radiological</td>
<td>5. Bilateral sacroiliitis of grade ≥ 2 on pelvic X-rays, or unilateral grade ≥ 3, excluding other causes of sacroiliac diseases (e.g. Paget’s disease of the bone, hypophosphatemia, hyperparathyroidism, fluorosis, infection with tuberculosis or brucellosis, familial Mediterranean fever, severe neurological disorders)</td>
</tr>
</tbody>
</table>

**Definitive AS:**
- 3 out of 4 clinical criteria
- Radiological criteria AND at least 1 clinical criteria.

**Possible AS:**
If a patient fulfills any of the above criteria with age of onset ≤ 45 years.

3.8 Amor Criteria and ESSG Criteria for Classification of SpA (1990-1991)

In early 1990s, 2 sets of classification criteria for the entire group of SpA were established; the Amor criteria and the ESSG criteria. Both criteria included undifferentiated SpA for the first time, in addition to the classic members of the group (AS, psoriatic arthritis, reactive arthritis, enteropathic arthritis). Yet both sets of criteria may have low sensitivity in establishing early diagnosis of SpA.

The Amor criteria (Figure 5) were developed with multiple entry criteria which make them feasible and may actually facilitate a diagnosis in an individual patient. The 12 parameters incorporate historical, laboratory (HLA-B27), and radiographical (sacroiliitis) features. Historical data ask about nocturnal back pain and/or morning stiffness, buttock pain (more points if alternating), a good response to a full-dose of NSAIDs within 48 hours, and common extra-articular manifestations. These were the first criteria to recognize response to NSAIDs as a characteristic of IBP. A recent Cochrane review of a 65 trials showed that short-term use of NSAIDs (including COX-2 inhibitors) in patients with acute and chronic mechanical back pain without sciatica provided a marginal relief. This endorses the concept that NSAIDs have a better response in SpA (101). One of the advantages of the Amor criteria is that they can be applied to patients who present with extra-articular symptoms such as uveitis or enthesitis without having IBP. These criteria performed slightly better than ESSG criteria to classification of SpA (sensitivity 85%, specificity 90%) (102).
**Figure 4. Amor classification criteria for spondyloarthritis (1990)**

<table>
<thead>
<tr>
<th>A. Clinical Symptoms/History</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pain at night (spine) or morning stiffness</td>
<td>1</td>
</tr>
<tr>
<td>2. Asymmetrical oligoarthritis</td>
<td>2</td>
</tr>
<tr>
<td>3. Gluteal (buttock) pain (any) or alternating gluteal pain</td>
<td>1</td>
</tr>
<tr>
<td>4. Sausage like digit or toe (dactylitis)</td>
<td>2</td>
</tr>
<tr>
<td>5. Enthesitis (heel)</td>
<td>2</td>
</tr>
<tr>
<td>6. Uveitis</td>
<td>2</td>
</tr>
<tr>
<td>7. Urethritus/Cervicitis within 1 month before onset of arthritis</td>
<td>1</td>
</tr>
<tr>
<td>8. Diarrhoea within 1 month before onset of arthritis</td>
<td>1</td>
</tr>
<tr>
<td>9. Psoriasis, balanitis or inflammatory bowel disease</td>
<td>2</td>
</tr>
<tr>
<td><strong>B. X-rays</strong></td>
<td></td>
</tr>
<tr>
<td>10. Sacroiliitis (grade 2 bilaterally or grad 3 unilaterally)</td>
<td>3</td>
</tr>
<tr>
<td><strong>C. Genetical background</strong></td>
<td></td>
</tr>
<tr>
<td>11. HLA-B27 positive or positive family history for AS, ReA, uveitis, psoriasis or inflammatory bowel disease</td>
<td>2</td>
</tr>
<tr>
<td><strong>D. Good response to NSAIDs</strong></td>
<td></td>
</tr>
<tr>
<td>12. NSAIDs show a good response within 48 hours, or relapse within 48 hours after NSAID are stopped</td>
<td>2</td>
</tr>
</tbody>
</table>

At least 6 points are necessary


From the ASAS slide collection, [http://www.asas-group.org](http://www.asas-group.org), open access.
The ESSG criteria (Figure 6) rely on historical data and the entry criterion is presentation with IBP (as defined by Calin criteria) and/or synovitis predominantly of the lower limbs. The presence of radiographic sacroiliitis is not mandatory. There is no item on response of back pain to NSAIDs or HLA-B27 status. The latter was removed during the development stage in anticipation that these criteria would be applied in population studies without the need for an expensive laboratory test. The specificity and sensitivity are 87% and 86% respectively which were also confirmed in other studies (103).
Figure 5. The European Spondyloarthropathy Study Group (ESSG) classification criteria for spondyloarthritis (1991)

Inflammatory Back Pain  or  Synovitis

- asymmetric or
- predominantly in the lower limbs

plus one of the following:

- enthesitis (heel)
- positive family history
- psoriasis
- Crohn's disease, Colitis ulcerosa
- urethritis / cervicitis or acute diarrhea within one month before arthritis
- buttock pain (alternating between right and left gluteal areas)
- sacroiliitis


From the ASAS slide collection, http://www.asas-group.org, open access.
3.9 **Studies of Rudwaleit and Colleagues (2004-2006)**

Rudwaleit and colleagues took the lead to further study the characteristics of IBP in order to facilitate early referral to rheumatologists. They previously found that IBP, as defined by the Calin criteria, increased pre-test probability of axSpA from 5% in patients who present to their primary care physicians with chronic back pain to 14%. The positive likelihood ratio (LR) for IBP was 3.1, while the highest positive LR was the combination of a positive HLA-B27 and a positive MRI (+LR of 9.0 for each). LR is unlikely to be influenced by the disease prevalence. Table 4 summarizes the diagnostic properties for clinical, laboratory and imaging parameters. They also showed that post-test probability dramatically increased upon using different combinations of historical parameters of the ESSG criteria (range 35-95%). Further, this probability also increased to 90% when clinical and laboratory features of axSpA were combined. Based on the positive LRs of individual parameters, the authors developed a diagnostic algorithm (later known as Berlin criteria) for early axSpA in patients who have chronic low back pain for more than 3 months. Definitive diagnosis of axSpA was made when the probability was 90% or more, while probable diagnosis was made when the probability was 89 – 90%. This new algorithm works well in the individual patient (104).
Table 4. Summary of diagnostic tests for individual parameters used for the diagnosis of early axSpA in patients with chronic back pain lasting ≥ 3 months

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sensitivity % (range *)</th>
<th>Specificity % (range*)</th>
<th>Positive LR</th>
<th>Post-test probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory back pain</td>
<td>75 (38 – 95)</td>
<td>76 (76 – 100)</td>
<td>3.1</td>
<td>14</td>
</tr>
<tr>
<td>Alternate buttock pain</td>
<td>40 (20–43)</td>
<td>90 (88–100)</td>
<td>4.0</td>
<td>17</td>
</tr>
<tr>
<td>Enthesitis (heel pain)</td>
<td>37 (16–52)</td>
<td>89 (89–96)</td>
<td>3.4</td>
<td>15</td>
</tr>
<tr>
<td>Dactylitis</td>
<td>18 (12–27)</td>
<td>96 (96–99)</td>
<td>4.5</td>
<td>19</td>
</tr>
<tr>
<td>Anterior uveitis</td>
<td>22 (10–22)</td>
<td>97 (97–100)</td>
<td>7.4</td>
<td>28</td>
</tr>
<tr>
<td>Family history of SpA‡</td>
<td>32 (7–36)</td>
<td>95 (93–100)</td>
<td>6.4</td>
<td>25</td>
</tr>
<tr>
<td>Peripheral arthritis</td>
<td>41 (26–62)</td>
<td>87 (91–100)</td>
<td>4.0</td>
<td>17</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>10 (1–17)</td>
<td>96</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
<td>4 (2–7)</td>
<td>99</td>
<td>4.0</td>
<td>17</td>
</tr>
<tr>
<td>Good response to NSAIDs</td>
<td>77 (61–77)</td>
<td>85 (75–85)</td>
<td>5.1</td>
<td>21</td>
</tr>
<tr>
<td>Elevated CRP</td>
<td>50 (38–69)</td>
<td>80 (67–100)</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>HLA-B27 positive (with axial involvement)</td>
<td>88 (83–96)</td>
<td>90 (91–96)</td>
<td>9.0</td>
<td>32</td>
</tr>
<tr>
<td>Positive MRI (STIR sequence)</td>
<td>90 (54–93)</td>
<td>90 (83–100)</td>
<td>9.0</td>
<td>32</td>
</tr>
</tbody>
</table>

* Ranges were extracted from different studies. ‡ Family history of SpA includes AS, reactive arthritis, IBD, psoriasis, and anterior uveitis.

**Abbreviations:** NSAIDs, Non-steroidal Anti-inflammatory Drugs; CRP, C-Reactive Protein; MRI, Magnetic Resonance Imaging; STIR, Short Tau Inverse Recovery; positive LR [Likelihood Ratio = sensitivity / (100 - specificity)]. Adapted with permission from the BMJ Group © Rudwaleit, M. et al. Ann Rheum Dis. 63,535-543 (2004).
In 2006, German investigators led by Rudwaleit published modified criteria for IBP. They developed a standardized questionnaire (in German) which was administered face-to-face by an examiner in the rheumatology, orthopedic surgery, and neurosurgery clinics of hospital-based departments and private practices where patients ≤ 50 years with chronic low back pain for at least 3 months were recruited. Questionnaire items were related to age of onset of back pain, disease duration, time period of the onset of back pain (within 1 hour, 1 day, 1 week, between 1 and 4 weeks, or within 1 year), current and any previous episodes of morning stiffness, duration of morning stiffness, alleviation of back pain with rest, presence of buttock pain, pain radiating to the leg, awakening nocturnal pain, time of nocturnal awakening, preceding events (e.g. trauma, mental or emotional stress, or infection), time of first consultation with a doctor, number of consulted doctors, duration of sick leaves in days within the last 4 weeks, current intensity of back pain, and impact of back pain on well-being. Patients were classified to either AS, based on the modified New York criteria, or mechanical low back pain (MLBP). Blinding of the examiner from the diagnosis was not achieved completely, which may have influenced how the answers were obtained.

The best diagnostic results were obtained from a set of 4 combined parameters: morning stiffness for more than 30 minutes, improvement with exercise but not with rest, back pain that awakens at the second half of the night, and alternating buttock pain. If 2 out of 4 the criteria are fulfilled, the sensitivity are 70%, the specificity is 81%, and the positive LR is 3.7. Any single parameter could not discriminate between AS and MLBP. None of the parameters was able to rule out AS. Furthermore, some items were found in both disease entities. For example, low back pain frequently awakened patients at night in both diseases (75% and 63% respectively). Low back pain lasting 30 minutes or more was also prevalent in 64% of AS patients and in 25% of MLBP. The term “insidious onset” was problematic and only discriminated one-fourth of AS from MLBP when it meant an onset of back pain within a year. Table 5 summarizes the frequencies and positive LRs of features of IBP among patients with AS and MLBP.
Table 5. Frequencies and positive likelihood ratios of IBP features in AS and MLBP patients

<table>
<thead>
<tr>
<th>Feature</th>
<th>% of AS patients (n = 101)</th>
<th>% of MLBP patients (n = 112)</th>
<th>Positive LR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of onset of back pain &lt; 40 years</td>
<td>96.0</td>
<td>86.0</td>
<td>1.1 (1.0 – 1.2)</td>
</tr>
<tr>
<td>Morning stiffness for more than 30 minutes</td>
<td>64.3</td>
<td>24.7</td>
<td>2.7 (1.9 – 4.0)</td>
</tr>
<tr>
<td>Improvement with exercise</td>
<td>78.2</td>
<td>50.0</td>
<td>1.6 (1.3 – 1.9)</td>
</tr>
<tr>
<td>Improvement with rest</td>
<td>31.7</td>
<td>58.0</td>
<td>0.5 (0.4 – 0.8)</td>
</tr>
<tr>
<td>Improvement with exercise but not with rest</td>
<td>55.4</td>
<td>21.4</td>
<td>2.6 (1.8 – 4.1)</td>
</tr>
<tr>
<td>Any buttock pain</td>
<td>63.4</td>
<td>57.1</td>
<td>1.1 (0.9 – 1.4)</td>
</tr>
<tr>
<td>Alternating buttock pain</td>
<td>36.6</td>
<td>11.6</td>
<td>3.2 (1.8 – 5.6)</td>
</tr>
<tr>
<td>Awakening back pain any time</td>
<td>75.0</td>
<td>62.7</td>
<td>1.2 (1.0 – 1.4)</td>
</tr>
<tr>
<td>Awakening back pain at the second half of the night</td>
<td>44.0</td>
<td>22.7</td>
<td>2.0 (1.3 – 2.9)</td>
</tr>
</tbody>
</table>

3.10 Development of the ASAS Criteria for axSpA (2009)

The modified New York criteria perform well in establishing a definitive diagnosis of a patient with radiographic sacroiliitis and at least one clinical feature. Because these strict criteria do not take into account patients who have IBP and normal sacroiliac joints i.e. non-radiographic axSpA, earlier detection of those patients is often missed. The ESSG and Amor criteria classify the whole SpA group without focus on axSpA. MRI of the sacroiliac joints was shown to demonstrate active inflammation in patients who do not have radiographic sacroiliitis.

To overcome these inherent limitations of these classification criteria, new classification criteria (ASAS) for axSpA have been recently established and validated offering additional clinical, laboratory (HLA-B27 status, ESR, and CRP) and radiological features to facilitate earlier diagnosis (Figure 7). The entry criteria are chronic back pain lasting 3 months or more and occurring before the age of 45. During the development of the ASAS classification criteria for axSpA, there was an increased sensitivity of the criteria when selecting (chronic back pain) as an entry criterion as opposed to the newly developed IBP criteria by Rudwaleit et al (93).

Patients then should either have positive HLA-B27 (clinical arm) or sacroiliitis (imaging arm) detected either by pelvic X-ray as per the modified New York grading criteria or by MRI. Once patients enter either diagnostic arm, they should have one of two more parameters. The definition of IBP, as one of the SpA features, was based on expert opinion and is to be elicited by the expert rheumatologist. The 4 items includes age of onset < 40 years, insidious onset, improvement with exercise, no improvement with rest, and pain at night with improvement upon getting up. Because of lack of statistical significance on the item “morning stiffness”, it was removed from the IBP definition. If 4 criteria out of 5 are fulfilled, the sensitivity is 80% and specificity is 72% in the validation cohort. The new ASAS criteria have a sensitivity and specificity of approximately 83% and 84% respectively for the entire set. These criteria showed a superior performance over the ESSG and Amor criteria in the validation study, with sensitivity and specificity of 83% and 84% respectively. Therefore, they are reliable in
classifying patients for research and in diagnosing patients in the rheumatology practice (105, 106).
Figure 6. ASAS classification criteria for axial Spondyloarthritis (axSpA)

In patients with ≥3 months back pain and age at onset <45 years

**Sacroiliitis on imaging**
- plus
- ≥1 SpA feature

**OR**

**HLA-B27**
- plus
- ≥2 other SpA features

*SpA features
- inflammatory back pain
- arthritis
- enthesitis (heel)
- uveitis
- dactylitis
- psoriasis
- Crohn's/collitis
- good response to NSAIDs
- family history for SpA
- HLA-B27
- elevated CRP

*Sacroiliitis on imaging
- active (acute) inflammation on MRI highly suggestive of sacroiliitis associated with SpA
- definite radiographic sacroiliitis according to mod NY criteria


From the ASAS slide collection, [http://www.asas-group.org](http://www.asas-group.org), open access.
These criteria carry certain limitations. First, although the estimated prevalence of AS after the age of 45 is 5%, some patients will be missed at the entry criteria. Second, the current definition of IBD includes Crohn’s disease and ulcerative colitis but not indeterminate (unclassified) colitis. In this case, patients with the latter category may fulfill the ASAS classification criteria for axSpA if they meet the entry criteria, a clinical and/or imaging criteria, and have other supporting parameters such as family history of SpA or peripheral arthritis. Third, elevated CRP or ESR levels can occur in IBD patients independent of co-existing axSpA.

3.11 Development of Case Ascertainment Questionnaire (2010)

A recent case ascertainment questionnaire (107) has been developed and validated to capture patients with radiographic AS.

In phase 1, a literature review using PubMed was performed generating 118 candidate items. Members of an advisory board assessed each item by using a 9-point scale. Items were selected only if their average score was ≥ 6. This led to retention of 42 items.

In phase 2, the 42 items were cognitively tested among 3 groups of patients with established AS (diagnosed by the modified New York criteria): patients with disease duration ≤ 10 years, patients with disease duration of > 10 years, and patients with mixed disease duration. The purpose of this phase was to allow patients to evaluate the items in terms of comprehension, clarity of questions, flow of questions, font size and typographical errors. This led to re-wording of some items, splitting some items into more than one or vice versa, changing the order of some questions, and addition of 9 items on fatigue when it was frequently identified by patients. This phase led to drafting 52 items.

Phase 3 was a feasibility study with the purpose of piloting the items among 50 patients with AS and 50 patients with chronic back pain for ≥ 3 months. Patients were required to be ≥ 18 years, fluent in English, and non-participants in phase 1. Internal consistency was
measured, although not reported, by Cronbach’s alpha, the average correlation between each pair of items (inter-item correlation), and the correlation of each item to the total sum score of items (item-total correlation). The investigators used the result of internal consistency, bivariate and multivariate analyses to guide them into further item reduction and modification.

The goals of phase 4 were to further refine the items, develop a scoring system for the questionnaire and validate it in a case-control study (patients with AS and chronic back pain). Similar to the requirements of phase 3, patients could not have participated in phase 3. The first step of this phase involved measuring bivariate and multivariate analyses in addition to internal consistency. The bivariate analysis of all items demonstrated a statistical difference between both groups for 24 out of 43 items. The multiple regression analysis further identified 12 items that were highly associated with classic AS patients compared to patients with chronic back pain. These 12 items were related to 3 main features: gender, characteristics of back pain, and iritis. IBP questions are as follows: pain or stiffness for more than 3 months, locations of pain or stiffness (neck, hip, other regions), age of onset of back pain, duration of back pain or stiffness, influence of exercise on the pain or stiffness of the lower back or buttocks, influence of daily physical activity on pain or stiffness on the lower back or buttocks, and NSAIDs use and whether they improve symptoms within 48 hours. There is one question which asked about numbness or tingling radiating to the legs and another one asking if the pain or stiffness was a result of trauma. A picture or diagram of the back is not used in this questionnaire. The sensitivity and specificity of the case ascertainment questionnaire in the development sample was approximately 70% and 99% respectively. The second step of this phase is the validation of the questionnaire which involved selection of a random sample of patients with AS and chronic back pain. The sensitivity and specificity of the case ascertainment questionnaire in the validation sample was approximately 67% and 95% respectively. About 54% of the AS patients in this study were on biologic drugs. However, the investigators demonstrated that the answers of the questionnaire did not differ in patients who were on these drugs compared to those who were not.
There are some limitations of this questionnaire. First, it did not target patients with early axSpA as the mean disease duration among AS patients was 21.8 years. Second, it did not address patients with IBD-associated AS. Thirdly, while the scoring system was driven by data, it is complicated and not easily applied. The scoring was designed to have a predicted probability of at least 66.86%. Finally, it was not comprehensive in terms of asking about different features of axSpA.

3.12 Summary

In the past century, there has been extensive work attempting earlier identification of axSpA in order to overcome the delay in diagnosis for these patients. Different classification criteria were developed with a special focus on refining the features of IBP. Up to April 2011, the literature has only 2 self-reported questionnaires in English: the 5-item Calin questionnaire and the case ascertainment questionnaire. Neither questionnaires included clinical pictures or diagrams of the back nor did they target cases of early axSpA in IBD patients. Recent studies have recommended combining IBP features with extra-axial manifestations to achieve a higher post-test probability of axSpA.

Attention needs to be directed at those patients who are at heightened risk of developing axSpA, as is the case in IBD patients. One approach to identify those IBD patients with undetected axSpA is to screen them when they present to their gastroenterologist for follow-up of their IBD. This type of screening is recommended by Sackett and Holland in this specific scenario and is called “case finding” which is best administered to patients who present to their physicians with different complaints so that there is an opportunity to ask them about symptoms of the disease of interest. A good case finding questionnaire requires items that are accurate, highly sensitive (commonly present in the target population), reasonably specific, and that have good predictive value (108).
Chapter 2

Rationale, Hypothesis and Aims
1 Rationale

There is an agreement among different studies that in order to detect axSpA earlier, it is imperative to obtain an appropriate clinical history using various questions to address different aspects of axSpA. A health care worker (usually a rheumatologist) or a self-reported questionnaire can extract the clinical history. The only 2 questionnaires (in English) available in the literature, as of April 2011, were aimed to identify patients with classic AS by asking specific questions which were mainly focused on characteristics of IBP.

Patients with IBD are at risk of developing axSpA which remains undetected in many patients for years. Once those patients are identified, a timely initiation of appropriate treatments can in turn lead to better pain control, improved functional outcomes and enhanced quality of life. This would minimize unnecessary and expensive diagnostic and therapeutic interventions.

2 Hypothesis

I hypothesized that a simple, sensible and reliable self-reported questionnaire could be developed to allow earlier case finding of axSpA in IBD patients (populations at risk) who are being seen in hospital-based gastroenterology clinics.

3 Aims

1) To describe the development including sensibility of the Toronto Axial Spondyloarthritis Questionnaire (TASQ).

2) To evaluate the test-retest (intra-rater) reliability of the TASQ.
Chapter 3

Methods
1 Development of Questionnaire Items

Development of questionnaire items required multiple steps including: conceptualization, literature review to identify characteristics of the construct and select candidate items, choosing appropriate scaling response and format, sensibility assessment among the SpA team and 2 general rheumatologists, 3 stages of pilot testing and creation of the final version. Figure 8 depicts the stages for the development of the Toronto Axial Spondyloarthritis Questionnaire (TASQ).
Figure 7. Stages of development of the TASQ
1.1 Phase 1: Conceptualization

As mentioned in the Rationale section, the specific aim of the project is case finding to facilitate early referral of IBD patients who potentially may have undiagnosed axSpA. The construct in this study is axSpA in IBD.

1.2 Phase 2: Item Pool Generation

A thorough literature review was performed using Medline (from 1941 to April 2011) and PubMed to identify potential items in preparation to the development of questionnaire. All the reviewed papers were restricted to the English language. Keywords included “inflammatory back pain”, “back pain”, “symptoms”, “features”, “manifestations”, “ankylosing spondylitis”, “spondyloarthritis”, “spondyloarthropathy”, “risk factors”, “questionnaire”, “screening”, “case finding” and “diagnosis”. If an article cited other relevant article(s), the latter was reviewed to identify potential items.

The main items for IBP symptoms were identified and reviewed from previous descriptions of the work of Hart et al, Calin et al, Møller et al, Gran, and Rudwaleit et al (89, 90, 93-95, 97). In addition, a thorough review of previously reported characteristics of IBP associated with IBD was conducted. Previously developed classification criteria for AS and SpA were reviewed with the focus on clinical items. These include the 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), ESSG criteria for SpA (1991), and the most recent ASAS classification criteria for early axial and peripheral SpA (2009 and 2010 respectively) (3, 92, 98, 102, 103, 105, 109).

Using this review I generated 93 potential items which did not include any items loaded with sensitive issues. A mixture of sensitive and specific items was chosen in keeping with the properties of a case-finding instrument. I added 4 pictures of peripheral arthritis.
of hand joints (2 pictures) and dactylitis (2 pictures of a finger and a toe). I also added a diagram of the back with arrows pointing towards the neck, upper back (between the upper and lower borders of the scapula), middle back, lower back (lower thoracic to end of lumbar areas), and buttock.

1.3 Phase 3: Scaling Responses and Instrument Format

Most of the questionnaire items asked closed-ended questions with binary responses, with fewer questions providing continuous data. In general, questions with binary responses are easy and quick to complete for patients. For clinicians, they are easy to code and analyze. Standardized instructions on how to answer the questionnaire were provided. Some branching questions (sub-questions) were included.

Demographic questions were asked at the beginning of the questionnaire (sex and date of birth). No question identified the patient’s name in order to reassure patients of confidentiality and to comply with privacy guidelines. Three major headings represented the domains of the questionnaire (IBD, inflammatory back symptoms, and extra-axial features).

1.4 Phase 4: Selection of Items

All items were worded neutrally i.e. free of value-loaded items. There was one item which asked about race since the axSpA is more prevalent in Caucasians of European origin. In order to qualify for the questionnaire, a patient must have inflammatory bowel disease and back pain or stiffness. This is an absolute criterion. The first question is general and relevant to the main construct “Has there been a time when you had pain or stiffness in your back for 3 months or more?” This is designed to reduce referrals for mechanical back pain in which the pain commonly lasts less than 3 months. Back pain was considered equivalent to back stiffness, allowing the use of this double-barreled
1.4.1 Sensibility Assessment of Questionnaire Items

Alvan R. Feinstein, who is considered one of the founders of modern clinical epidemiology, published the concept of Clinimetrics in 1982. It advocates for the methodological approach of clinical judgment when developing an instrument that measures a clinical observation, diagnostic criteria, prognostic or therapeutic outcome. Important facets of the clinimetric approach include sensibility, reliability, validity, and responsiveness. Sensibility assessment refers to the usefulness of an instrument. It includes a statement of clinical function (purpose), clinical justifications, clinical applicability (indication of population and setting), overt format (comprehensibility, replicability, and suitability of scale), face and content validity, and feasibility (110).

The clinical function (purpose) and clinical justification (rationale) were discussed in detail in Chapter 2 (Rationale, Hypothesis, and Aims).

In terms of clinical applicability, our questionnaire at this point is only applicable to a subset of patients with seronegative SpA i.e. axSpA in IBD patients. In this study, clinical applicability was evaluated by asking: “The questionnaire is intended to ask questions related to presence of axial spondyloarthritis in patients with inflammatory bowel disease. Do you think this questionnaire achieved this goal?”

Comprehensibility refers to the ability of understanding the questionnaire which is assessed by limiting the questionnaire to important items (oligo-variability) and the appropriateness of the number of response options and clarity of item weighting (transparency). I assessed oligo-variability by asking, “Do you feel that the questionnaire included the necessary questions?” With regards to transparency, I assessed it from 2 perspectives. The first perspective asked about response options; “Do you think the response options were appropriate (for example, number of options)?” The second perspective initially assessed transparency by asking the members about the appropriate way to score the items. I then administered the new version (post-pilot) of TASQ
questionnaire among them and assessed transparency by asking, “Assuming that each question is weighted equally with a score of 1, do you think that the scoring is simple?”

**Replicability** refers to the clarity of instructions.

**Face validity** is an attribute to be evaluated when creating a new measuring instrument to ensure that items or domains appear to evaluate the desired construct. In this study, the construct was axSpA in IBD. I evaluated face validity by asking the following: “Do you feel that this questionnaire accurately measures what we are trying to measure?”

**Content validity** refers to how comprehensive the questionnaire is i.e. whether items address sufficient domains for the construct of the disease. In this study, content validity was assessed by asking 2 questions: “Do you think an important question was missing?” and “Do you think an inappropriate question was included?” Items were only included if there was a 100% agreement on the relevance or irrelevance of each item to the construct of the questionnaire.

**Feasibility** refers to ease of use of an instrument. It was assessed in our study by asking about completion time, acceptability of the questionnaire by patients (i.e. no sensitive or potentially offensive items), and readability (clarity of questions, flow of questions, presence of typographical errors, appropriate font size, usefulness of pictures and diagrams). Readability was evaluated by asking “Were the sentences easy to read?” I also tested readability by calculating the Flesch-Kincaid grade level and the Flesch Reading Ease Scale using Microsoft Word for Mac 2011. The recommended grade for a questionnaire is grade 6 or less. The Flesch Reading Ease is a 100-point scale, and the higher the score; the easier it is to read the questionnaire. The scale is 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 (111). Table 6 shows the checklist of sensibility assessment with their definitions which were adapted from Feinstein’s principles.
### Table 6. Checklist of sensibility assessment with definitions

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose and framework</strong></td>
<td></td>
</tr>
<tr>
<td>Clinical function</td>
<td>Purpose of the questionnaire</td>
</tr>
<tr>
<td>Clinical justification</td>
<td>Rationale behind questionnaire development</td>
</tr>
<tr>
<td>Clinical applicability</td>
<td>Application of the questionnaire to a specific group of patients in a specific setting</td>
</tr>
<tr>
<td>Face validity</td>
<td>Accurately measuring what the questionnaire is trying to measure</td>
</tr>
<tr>
<td><strong>Comprehensibility</strong></td>
<td></td>
</tr>
<tr>
<td>Oligo-variability</td>
<td>Limiting the questionnaire to important items</td>
</tr>
<tr>
<td>Transparency</td>
<td>Ability to see through the contribution of each scored item and its response options to the final score</td>
</tr>
<tr>
<td><strong>Replicability</strong></td>
<td>Clarity of instructions</td>
</tr>
<tr>
<td><strong>Content validity</strong></td>
<td></td>
</tr>
<tr>
<td>Important omissions</td>
<td>Exclusion of important items</td>
</tr>
<tr>
<td>Inappropriate inclusions</td>
<td>Inclusion of inappropriate items</td>
</tr>
<tr>
<td><strong>Feasibility</strong></td>
<td></td>
</tr>
<tr>
<td>Time to completion</td>
<td>Amount of time taken to complete the questionnaire</td>
</tr>
<tr>
<td>Acceptability</td>
<td>Approval of the questionnaire by patients and health care workers by excluding sensitive or potentially offensive items</td>
</tr>
<tr>
<td><strong>Readability</strong></td>
<td></td>
</tr>
<tr>
<td>Clarity of questions</td>
<td>Ability to understanding the questions easily</td>
</tr>
<tr>
<td>Flow of questions</td>
<td>Transitioning of questions in a logical manner</td>
</tr>
<tr>
<td>Readability tests</td>
<td>Flesch Readability Ease Scale and Flesch-Kincaide grade level</td>
</tr>
<tr>
<td>Typographical errors</td>
<td>Presence of typographical errors</td>
</tr>
<tr>
<td>Font size</td>
<td>Appropriateness of the font size used in the questionnaire format</td>
</tr>
<tr>
<td>Illustrations</td>
<td>Clarity and usefulness of diagrams and pictures used in the questionnaire</td>
</tr>
</tbody>
</table>

Adapted from Clinimetrics 1987 (110)
Rowe and Oxman translated the above principles of sensibility into 12 questions using 7-point Likert scales and one last question that allowed patients to add any comment (112). I adapted their questions and added new 5 items for feasibility. I also changed the response options from Likert scales into dichotomous scales and allowed a final open-ended question for any comment. The sensibility questions were asked to the committee members (SpA team and 2 general rheumatologists) and patients (Table 7).
Table 7. Questionnaire for sensibility assessment

<table>
<thead>
<tr>
<th>Concept</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical function</strong></td>
<td>The questionnaire is intended to ask questions related to presence of axial spondyloarthritis in patients with inflammatory bowel disease. Do you think this questionnaire achieved this goal?</td>
</tr>
<tr>
<td><strong>Face validity</strong></td>
<td>Do you feel that it accurately measures what we are trying to measure?</td>
</tr>
<tr>
<td><strong>Comprehensibility</strong></td>
<td></td>
</tr>
<tr>
<td>Oligo-variability</td>
<td>Do you feel that the questionnaire included the necessary questions?</td>
</tr>
</tbody>
</table>
| **Transparency**       | 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          | Do you think the instructions were clear?                                |
| **Content validity**   | Do you think an important question was missing? (important omissions)   

Do you think an inappropriate question was included? (inappropriate inclusions)

| **Feasibility**        | **Time to completion** Please indicate the amount of time you took to complete reading the questionnaire?  

_____ minutes |
| Acceptability         | Were there any questions that were unacceptable question (too personal or offensive)? |
| **Clarity of questions** | Do you think the questions were clear?                        |
| **Flow of questions**  | Was the flow of questions smooth?                                |
| **Ease of reading**    | Were the sentences easy to read?                                 |
| **Typographical errors** | Were there any typographical errors?  

If so, please indicate where. |
| **Font size**          | Was the font size appropriate?                                    |
| **Illustrations**      | Were the pictures and diagram useful?                             |
The questionnaire for sensibility assessment was administered to the SpA team and 2 general rheumatologists (committee members). The former consisted of 2 SpA rheumatologists, an advanced physiotherapist, a clinical trial nurse, an infusion nurse coordinator, a biologic nurse coordinator, and a research analyst. The SpA team met on several occasions to assess the items for sensibility considering the overall purpose of case finding of the questionnaire. Relevant questions of the sensibility assessment were asked during the meetings.

1.5 Phase 5: Pilot Study

The purpose of the pilot study was to pretest the questionnaire among a small, targeted sample in a controlled setting in order to obtain a sensibility assessment and qualitative analysis that could refine the items of the questionnaire if necessary. This approach leads to reducing measurement error and improving the quality of the collected data.

I conducted pilot testing in 3 stages on axSpA patients, who attended the Spondylitis clinic at the Toronto Western Hospital (TWH). The Spondylitis Clinic at TWH is the major referral clinic for the care of patients with axSpA in Southern Ontario. It is part of the Spondyloarthritis Research Consortium of Canada Registry for Spondyloarthritis (SPARCC) which is a trans-disciplinary national research program focusing on genetic, pathogenesis, and outcome measures for patients with SpA (AS, psoriatic arthritis, IBD-related arthritis, undifferentiated SpA, and reactive arthritis). Beside Toronto, the program covers Newmarket, London Ontario, St. John’s, Montreal, Winnipeg, Saskatoon, and Edmonton. All SPARCC patients, who are 18 years or older and are fluent in English, give their informed consent at the beginning of their enrollment. Patients have an initial clinic visit then once a year after that. At each visit, patients follow a standardized protocol including self-administered questionnaires, clinical assessment by a rheumatologist, blood tests and X-ray if necessary (68).
I explained the purpose of the pilot phase to patients and instructed them to read the questionnaire and answer the questions at the end that are part of the sensibility assessment. Patient characteristics were collected including age, sex, the highest level of education (on a scale of 1 to 5: 1 = below grade 8, 2 = high school incomplete, 3 = high school graduate, 4 = college, 5 = university) reflecting the highest educational level obtained at the patient’s last clinic visit, the BASDAI score at the last clinic visit, and the BASFI score of the last clinic visit.

An additional questionnaire for sensibility assessment was administered to patients after filling out the TASQ questionnaires. The assessment ended with a free text comment area and a thank-you note. I also conducted a qualitative analysis by interviewing all patients after completion of the 2 questionnaires.

1.6 Phase 6: Sensibility Assessment of Post-Piloted TASQ
Completed by the Committee Members

I administered the sensibility assessment to the committee members after the modification made in the pilot testing.

2 Reliability study

2.1 Concept

Classical test theory has been adopted to measure the reliability of this new questionnaire in order to obtain the measurement error that is inherent in observed scores. Reliability represents a key concept in this theory and evaluates the reproducibility of an instrument in measuring an attribute or a construct. Test-retest (intra-rater) reliability is one of its types and is directly relevant here. It assesses the stability of the item scores on different
occasions to the population of interest within a 2-week period between administration times in order to reduce the chance of recall and intercurrent flare of axSpA (113).

2.2 Population of Interest and Sampling Method

As of July 1st, 2011, the database of the Spondylitis Clinic at the TWH 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. The patients had chronic back pain or stiffness that lasted 3 months or more and fulfilled the ASAS classification criteria for axSpA. The one exception was a male patient who fulfilled the modified New York criteria for diagnosis of AS based on typical features of IBP, and bilateral sacroiliitis of grade 2 but did not fulfill the ASAS criteria because the age of onset of his back pain was greater than the cut-off age of 45 years. However, we included this patient in our analysis because he conceptually still fulfilled the generic definition of axSpA. An amendment to distribute our new questionnaire was approved by the institutional review ethical board.

The questionnaires were mailed to all the 77 patients who have established axSpA and IBD concurrently. These patients reflected the sample of the total population of the IBD and axSpA. I adapted the tailored design method of Dillman to maximize the response rate (114). Each patient was provided with a package of 2 self-addressed and stamped return envelopes in which each had a one-page, double-sided and identical questionnaire accompanied by a personalized covering letter. The letter format was dated and had the TWH logo at the beginning, and the University of Toronto logo at the end, and was signed by my supervisor. This procedure was used to give patients the impression that their opinions are being sought by an academic clinician and leader from well-recognized academic institutions. The letter explained the purpose of the reliability study, the importance of their participation in our study and provided clear instructions about the completion of each questionnaire on two occasions 1-2 weeks apart. An ending statement acknowledged their time and cooperation (Appendix A). Each questionnaire was coded to
allow identification of the patient. The only identifying information on the questionnaire was date of birth. Patients were unaware that their responses would be compared among them in order to minimize the possibility of Hawthorne effect which refers to respondents modifying their responses because of awareness that they were being studied (115).

The return envelopes had a logo of the University Health Network which includes 3 hospitals: Toronto General Hospital, TWH, and Princess Margaret Hospital. The package was sent by a regular mail.

### 2.3 Sample Size Calculations for Test-Retest Reliability Study

Calculation of the sample size was based on 3 estimates – hypothesized reliability (R), number of observations and the confidence interval (CI). For test-retest reliability study, the number of observations was 2 when the questionnaire was to be administered to the same patient within a 2-week interval. With regards to the recommended reliability coefficient, Nunnally and Bernstein suggested a reliability of 0.70 for research purposes, and 0.90 for clinical purposes (individual level of patient care) (116). Streiner recommends a reliability of > 0.50 to define stability of a measuring instrument. I chose a minimally desired reliability of 0.80. The hypothesized reliability (R) of TASQ was estimated as 0.90, and in order to be 95% certain that it is definitely above 0.80, the 95% CI around the reliability is ± 0.10 (0.90 minus 0.80). The standard error (SE) of the reliability coefficient is approximately half this CI, or 0.05. Therefore, the calculated sample size with hypothesized R = 0.90, SE 0.05 and 2 observations is 24 patients (117).

The following calculations demonstrate steps used to reach the sample size for test-retest reliability:

First, I needed to calculate a conservative reliability coefficient (R*) to get a larger sample size than relying only on (R + SE). Second, I then used Fisher log transformation of Intraclass Correlation (ICC). It should be noted that the ICC here is equivalent to the reliability coefficient. The number of observations per subject is 2.
The calculated sample size (n) is rounded up to 24. I elected to mail the questionnaires to all the 77 consecutive patients with concurrent IBD and axSpA to account for an anticipated low response rate during the summer and for incomplete questionnaires. All the responses were entered into a database and were double-checked.

### 3 Statistical Analyses

#### 3.1 Statistical Analysis for the Sensibility Assessment

Qualitative and descriptive analyses were used for the sensibility assessment among the committee members and among patients with axSpA.

#### 3.2 Statistical Analysis for Patients in the Pilot Study

Descriptive analysis was used to describe the clinical and laboratory characteristics of axSpA patients who participated in the pilot study.
3.3 Statistical Analyses for Responders and Non-Responders

For description of the clinical characteristics between patients who responded to our questionnaire and those who did not, the following parameters were collected: type of IBD (Crohn’s disease, ulcerative colitis, indeterminate colitis), the highest level of education, BASDAI and BASFI scores of the last clinic visit, ESR level of the last protocol visit, and CRP level of the last protocol visit. Serum ESR and CRP levels are inflammatory markers that may be elevated with activity of disease in axSpA.

Categorical variables were summarized by frequencies and percentages, while continuous variables were summarized by mean, standard deviation and median. A two-sample Student’s t-test was used to compare the mean difference of age, and scores of BASDAI and BASFI between the groups (respondents and non-respondents). Wilcoxon rank sum test was used to test the medians of the level of education between groups. Chi-square test was used to assess the association between categorical variables (i.e. types of IBD) between groups. Statistical analyses were 2-sided, and the statistical significance is defined by p values of <0.05.

3.4 Coding the Questionnaire’s Responses

Dichotomous response options were coded as 1 for Yes and 0 for No. Two questions, which contained continuous variables, were converted into dichotomous variables. The first question asks about the age of onset of back pain (age ≤ 45 years or age > 45 years) in accordance to the ASAS classification criteria for axSpA. The second question asks about duration of morning stiffness (≥ 30 minutes or < 30 minutes). For the 3 questions with 3 response options (Yes, and 2 negative detailed response options), the negative options were coded as No. Similarly, the question which asked about family history of AS was coded as Yes, and No for No or Unknown.
3.5 Statistical Analysis for Test-Retest Reliability

For test-retest reliability, Cohen’s kappa statistics were obtained to determine \( \kappa \) agreement coefficient. The percentage of agreement for the items was also determined. The \( \kappa \) agreement coefficients were interpreted using the recommendations of Landis and Koch as follows: \( \kappa < 0.00 = \) poor agreement, \( \kappa \) between 0.00 to 0.20 = slight agreement, \( \kappa \) 0.21 to 0.40 = fair agreement, \( \kappa \) 0.41 to 0.60 = moderate agreement, \( \kappa \) 0.61 to 0.80 = substantial agreement, and \( \kappa > 0.80 = \) almost perfect agreement (118).

Statistical analyses were generated using SAS (version 9.3 Institute Inc., Cary, NC, USA) and R (version 2.14.2; The R Foundation for Statistical Computing). Statistical analyses were 2-sided, and the statistical significance is defined by \( p \) values of <0.05 as stated.
Chapter 4

Results
1 Selection of Items Using Sensibility Assessment

Following the first 3 phases (conceptualization, item pool generation, scaling responses and instrument format), phase 4 consisted of selection of items after a thorough assessment of their sensibility. Table 8 summarizes the results of sensibility assessment of this stage.
Table 8. Sensibility assessment for the pre-piloted TASQ completed by the committee members

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

* All the numbers in the table represent number of committee members who agreed on the corresponding principles (percentages in parenthesis).

† This was determined independently from the committee members.
**Clinical function**: The committee members agreed that the purpose of the questionnaire is clear.

**Comprehensibility**: Two members suggested adding some questions related to construct such as addition of response of back pain or stiffness to biologic agents and history of psoriatic arthritis. All members agreed on the appropriate number of response options. However, the concept of weighing each item was felt to be unclear to by the members.

**Replicability**: Two members had some suggestions to clarify the first instruction at the beginning of the questionnaire.

**Face validity**: Three members suggested removal of many questions that did not appear to measure what the questionnaire was supposed to measure. For example, the question regarding the extent of bowel disease involvement was thought not to be relevant to finding cases of axSpA, although it was relevant to peripheral arthritis. Another example is dropping of an item on “fatigue”. While fatigue is one of the features of axSpA, we decided to drop it because it can also be found in IBD patients who may suffer concurrently from anemia which often causes fatigue.

**Content validity**: Two members found that certain important questions were missing such as response of back pain or stiffness to biologic drugs. Questions which were inappropriately included were identified and discussed at length.

**Feasibility**: The median of self-reported time to completion of the questionnaire was 4 minutes (range 3 – 5) among all members. There was one member who identified the demographic question on ethnicity as irrelevant and potentially sensitive. Five members indicated that certain questions required clarifications, sentence shortening and better wording. Two members suggested changing the order of few questions to improve the flow. The Flesch reading ease scale was 72.2% which makes reading the questionnaire fairly easy. The Flesch-Kincaid grade level was 5.6 which is within the recommended grade for a questionnaire. One member identified one typographical error in the questionnaire. Two members found the font size small. Four members found 3 pictures not useful. The 2 pictures of peripheral arthritis could be misinterpreted as dactylitis. One
picture of finger dactylitis could be confused with peripheral arthritis. A picture of iritis could be mistaken as conjunctivitis and the members agreed on substituting this pictorial item with a clear and short question on history of iritis. All members agreed on the usefulness of the picture of toe dactylitis and the diagram of the back.

**Version 1 of TASQ:** Item reduction was made to include all the necessary items with consideration of including missing questions and excluding inappropriate items. For example, the item of ethnicity was removed as it was found unnecessary and potentially sensitive. Revisions of some items were made in order to have simple and short sentences whenever possible with one or 2 clauses. Double-barreled questions were allowed for the terms “back pain or stiffness” and “pain and swelling”. The latter was included to specifically describe peripheral arthritis and dactylitis independently. The pictures of peripheral arthritis, finger dactylitis and iritis were removed. This led to drafting the first version of the questionnaire which included 18 items, a diagram of the back and one picture of toe dactylitis.

## 2 Pilot study

In all the 3 pilot stages (4 patients in stage I, 9 in stage II, and 6 in stage III), the total number of patients was 19 with a mean age of 37 (range 24 to 53) and approximately 68% were males. The mean BASDAI score was 3.5 (range 0 – 8.8), while the mean BASFI score was 4.2 (range 0 – 7.7). The median of the patients’ highest level of education was 4.5 (range 3 – 5). All patients answered the first screening question that there had been a time when they had back pain or stiffness in their back for 3 months or more. Table 9 summarized the evaluation of sensibility among axSpA patients during the pilot testing.
Table 9. Stages of sensibility assessment of TASQ by patients

<table>
<thead>
<tr>
<th>Sensibility assessment</th>
<th>Stage I ( n = 4 ) (%) *</th>
<th>Stage II ( n = 9 ) (%) *</th>
<th>Stage III ( n = 6 ) (%) *</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comprehensibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparency ( ) (appropriate response options for each item)</td>
<td>2 (50)</td>
<td>9 (100)</td>
<td>6 (100)</td>
</tr>
<tr>
<td><strong>Replicability</strong></td>
<td>4 (100)</td>
<td>9 (100)</td>
<td>6 (100)</td>
</tr>
<tr>
<td><strong>Feasibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completion time, median (range) †</td>
<td>5 (3 – 5) minutes</td>
<td>4 (3 – 5) minutes</td>
<td>4 (3 – 5) minutes</td>
</tr>
<tr>
<td>Acceptability</td>
<td>4 (100)</td>
<td>9 (100)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Readability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of all questions</td>
<td>2 (50)</td>
<td>7 (78)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Flow of questions</td>
<td>4 (100)</td>
<td>9 (100)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Ease of readability</td>
<td>3 (75)</td>
<td>7 (78)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Absence of typographical errors</td>
<td>4 (100)</td>
<td>8 (89)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Appropriate font size</td>
<td>4 (100)</td>
<td>9 (100)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Usefulness of picture(s) and/or diagram</td>
<td>4 (100)</td>
<td>6 (67)</td>
<td>6 (100)</td>
</tr>
<tr>
<td><strong>Actions following each stage</strong></td>
<td>Removal of 4 items Modifications of 4 items</td>
<td>Removal of 2 items Removal of dactylitis picture Modification of 3 items</td>
<td>Minor modifications of few items</td>
</tr>
</tbody>
</table>

* All the numbers in the table represent number of patients who agreed on the corresponding principles (percentages in parenthesis).

† This was determined independently from the patients.
2.1 Stage 1

In stage 1, the questionnaires were administered to 4 axSpA patients. All patients found the instructions clear. None of them commented on inclusion of sensitive items. No patient found typographical errors, and all of them found the font size appropriate and the illustrations useful. The median time for completion of the question was 5 minutes (range 3 – 5).

Item Modification

One patient did not know how to answer the question which asked about awakening due to back pain during the second half of the night. The reason was that the cut-off time of sleep disturbance was not well defined. The patient suggested adding a third response option (both). However, we decided to keep the question as originally phrased with 2 response options because it was found to better characterize IBP in Rudwaleit’s study (93). We accepted a suggestion from a patient to add a third response option “Not applicable” to the question, “If you have buttock pain, does it move from one side to the other?” One patient commented on the question, “Does your back pain improve during the night upon getting up, walking or stretching?” and suggested the addition of a third response option “I do not get up” because back pain did awaken him at night but he did not necessarily get up, walk or stretch. This agreed with our clinical experience as some axSpA patients were awakened at night because of back pain but they do not get up and walk. Therefore, we accepted this suggestion.

There were 2 questions that proved difficult to understand. The first question was about the mode of onset of back pain “When your back pain first began, indicate how long it took to develop: ___hour(s), ___day(s), ___week(s), ___month(s), ___years”. This was modified to “Indicate the duration of time it took to develop”. The second question was on buttock pain “Do you have any buttock pain that alternates from side to side?” This was re-worded into the following: “If you have buttock pain, does it move from one side to the other?” Patients found the flow of questions smooth.
**Item and Response Options Reduction**

Three response options and one question were removed in this stage. There was a question on biologic drugs as a treatment for IBD “Have you been treated with any of the following: infliximab (Remicade), adalimumab (Humira) or other (please specify)?” with 3 responses for each drug “currently, previously, never”. Although no patient commented on this item, I decided to remove the response option “Never” for each drug, and change the question into the following “Has your inflammatory bowel disease ever been treated with a biologic drug such as Remicade™ or Humira™ or other biologic drug?” Again, some patients had difficulty with the question, “Does your back pain occur during the second half of the night?” We decided to remove it because it was difficult for patients to define exactly the second half of the night. These modifications led to a revised questionnaire (version 2) with 20 items, one diagram and one picture.

**2.2 Stage 2**

In stage 2, version 3 of the questionnaire was administered to 9 axSpA patients. All patients found the questionnaire transparent in term of appropriate response options for each item. All patients found the instructions adequate. On assessment of feasibility, all patients found the questionnaire acceptable. Again, all patients found the flow of questions smooth and the font size appropriate. The median time to complete the questionnaire was 4 minutes (range 3 – 5).

**Item Modifications**

Three patients made some comments regarding the clarity of some questions. I re-worded the question “Did your back pain or stiffness, that lasted for 3 months or more, improve after starting the drug(s)?” to the following “… improve after taking any biologic drug(s)?” to clarify the item and increase its sensitivity. There was still some difficulty in understanding the concept and response options for the question on the mode of onset of back pain. The latter was evident when some patients circled, for example, the response
“weeks” instead of providing a single number. One patient had the same misunderstanding upon answering the previous (version 2) questionnaire during the pilot 1 phase prompting us to modify it again and provide binary response options “When your back pain first began, did it develop quickly (over hours or days), or slowly (over weeks or months)” with instructions to choose one response. A similar issue was observed on an item which asked to provide a single number on the duration of morning stiffness in both hours and minutes but some patients circled one of the options. Therefore, I added an instruction beside the response options to type in numbers, as we still wanted to avoid restricting patients’ answers to a binary response option i.e. morning stiffness lasting 30 minutes or more vs. less than 30 minutes. This helps minimizes expected response bias. One patient suggested adding, “not related to injury” to the question, “Have you ever had heel pain”. I accepted this addition because it fits the true description of enthesitis.

One patient found a typographical error related to numbering of the question which was corrected.

Feedback From Free Text

A few patients suggested adding “hip pain” to the diagram of the back although they drew a line pointing towards the sacroiliac joints. One patient suggested adding the hip pain component but did not elaborate on its location on the diagram. The SpA team dropped this item during phase 4 (selection of the items) because of lack of a proper definition.

Item Reduction

One question and one sub-question were removed in this stage. I had to remove the item of alternating buttock pain partly due to difficulty in understanding the question. I also decided to remove a sub-question on “Prior diagnosis of fibromyalgia” (for details, please refer to the Discussion, section Content Validity).
**Image Reduction**

Three patients indicated that the picture of a toe dactylitis was confusing because it looked like peripheral arthritis. I accepted their suggestion of removing it, as the question was self-evident and I bolded the word “entire” in the question “Have you ever had pain and swelling of your entire finger or toe, not related to injury?” I avoided using the term “sausage-shaped swelling” because I felt it was an inappropriate metaphor.

The aforementioned modifications of the items led to creation of version 3 of the questionnaire which included 19 items.

### 2.3 Stage 3

Version 3 of the questionnaire was administered to 6 axSpA patients who were satisfied with the questionnaire items in terms of its sensibility. Minor modifications were performed related to the order of few questions.

### 3 Second Sensibility Assessment Completed by the Committee Members

I administered a sensibility questionnaire again to the committee members (SpA team and general rheumatologists) to evaluate version 3 of TASQ. There was a dramatic improvement in the sensibility of the final questionnaire among the members compared to the first sensibility assessment. There was a suggestion to modify the instruction on the item asking about the duration of morning stiffness to “number of hours” and “number of minutes”. Although patients did not have difficulty in understanding the instruction, the consensus was to make it clearer.

This time, I asked if the scoring was simple assuming that each question is weighted equally of a score of 1. All but 1 member thought the questionnaire scoring would be simple in this way. Table 10 shows the results of sensibility assessment among the members before and after pilot testing.
Table 10. Comparison of sensibility assessments completed by the committee members for the pre-piloted and post-piloted TASQ

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

* All the numbers in the table represent number of committee members who agreed on the corresponding principles (percentages in parenthesis).

† This was determined independently from the committee members.
4 Version 4 of the Questionnaire

Version 4 (dated 7th July 2011) was a one-page, double-sided questionnaire that was set out in a clear and uncluttered format accompanied by a diagram of the back. It took 5 minutes or less to complete the questionnaire. At the beginning of the questionnaire, demographic questions were asked, namely sex and date of birth. We did not include a question on identification of the patient’s name to reassure patients of confidentiality. Clear instructions on how to fill out the questionnaire were provided to ensure standardization.

The questionnaire consisted of 16 items. The response options were presented as follows: a binary response in 11 questions and 1 sub-question, 3-option responses in 2 questions and 1 sub-question, and provision of continuous variables in 1 question and 1 sub-question. We elected to retain additional 2 questions and 2 sub-questions in the final questionnaire, although they will not be scored. These questions did not increase the number of pages. Two questions serve as a confirmation that patients who complete the questionnaire have ever had a history of chronic back pain or stiffness lasting 3 months or more, and another question asks about the type of IBD. Two sub-questions help obtain descriptive data including types of biologic drugs and whether the intake is current or previous, and prior diagnosis of AS. The questionnaire is easy to read for a student at the 5th grade or above. The Flesch Reading Ease is 74.7% which is considered fairly easy.

5 Description of the Items in the Final Questionnaire

This section describes the item of the final version of the questionnaire. Figure 9 summarizes the domains which consists of IBD, inflammatory back symptoms, and extra-axial features.
Inflammatory bowel disease (IBD):
- Treatment with biologic agents (infliximab or adalimumab, others).
- Response of back symptoms after taking biologic agents.

Inflammatory back symptoms:
- Age of onset of back pain or stiffness.
- Morning stiffness.
- Mode of onset of back pain.
- Location of back pain: neck, upper / middle / lower back, buttocks.
- Nocturnal pain with sleep disturbance.
- Response to:
  a. Physical activities.
  b. Rest.
  c. Non-steroidal anti-inflammatory drugs (NSAIDs).
- Family history (first degree relatives) of AS.

Extra-axial features:
- Prior diagnosis of iritis or uveitis.
- Prior diagnosis of psoriasis.
- Prior diagnosis of psoriatic arthritis.
- Heel pain (enthesitis) ever.
- Peripheral arthritis ever.
- Dactylitis ever.
5.1 Inflammatory Bowel Disease (IBD)

Treatment with Biologic Agents

Choices of biologic agents here are infliximab, adalimumab, or others. As stated in the Introduction, the first 2 TNF inhibitors are the main biologic drugs used in the management of IBD patients. I added “other” to account for future biologic agents. This question will not contribute to the total score.

Response of Back Pain or Stiffness to Biologic Agents

The questionnaire asks if patients ever had chronic back pain or stiffness lasting 3 months or more that improved after taking biologic agents. A positive score is when there is a response of the back pain or stiffness to biologic agents.

5.2 Inflammatory Back Symptoms

Chronic Back pain or Stiffness Lasting ≥ 3 Months

As discussed in the Introduction, since previous studies confirmed the strong association between back pain and stiffness with axSpA, we used a double-barreled screening question combining “chronic back pain” with “stiffness” lasting 3 months or more.

Age of Onset of Inflammatory Back Pain

This is an open-ended question. A positive score is when the age of onset of IBP is 45 years or less in accordance to the ASAS classification criteria for axSpA.

Mode of Onset of Back Pain

The item asks whether the initial development of the back pain was quickly “over hours or days” or slowly “over weeks or months”. The latter response option constitutes a positive score.
Morning Stiffness of the Back

This item starts with a closed-ended question and then branches if the answer is “Yes” to presence of morning stiffness. The branching question then asks about duration. A patient who has morning stiffness will be given a positive score. If morning stiffness lasts 30 minutes or more, a positive score will be given.

Nocturnal Pain with Sleep Disturbance

This item starts with a closed-ended question then branches if the patient answers “Yes” to presence of nocturnal pain. The branching question then asks if the pain improves upon getting out of bed, walking or stretching with 3 responses: “Yes”, “No. This would not improve my back pain”, or “No. I do not get up”. A positive score is given to the question and the sub-question i.e. presence of nocturnal pain with sleep disturbance, and improvement upon getting out of bed, walking or stretching. Improvement of pain upon getting up has been consistently found in previous studies to be a major finding in AS patients (93, 95, 97, 119).

Locations of the Back Pain

Traditionally the low back has been the main complaint among axSpA patients (120). Even though cervical spine was recently found to be significantly associated with primary AS compared to IBD-associated AS (36), I did not want to miss those patients. In our questionnaire, I provided a diagram of the back and asked the patients to check the areas of back pain or stiffness including the neck, upper back, middle back, lower back and buttock. A positive score is given to any of these locations with a total possible score of 5.

Response to Daily Physical Activity

I preferred to use this item over “response to exercise”, which is used in both Calin criteria for IBP and the modified New York criteria, because it is more practical to ask about response of back pain or stiffness to daily physical activities such as walking. This item has also been used in the case ascertainment questionnaire (107). To increase the sensitivity of this item, I asked, “Does your back pain or stiffness improve at all with
daily physical activities?” A positive score is given if the patient answers “Yes” to this question.

**Response to Rest**

IBP typically lacks improvement with rest as was found in previous studies (3, 93, 119). To increase the sensitivity of this item, I asked, “Does your back pain or stiffness improve at all with rest?” A positive score is given if the patient answers “No” to this question.

**Responsiveness to NSAIDs**

This question asks, “Does your back pain or stiffness improve after taking anti-inflammatory drugs such as Motrin™, Advil™, Voltaren™, Naprosyn™, Aleve™, Indocin™, Mobicox™, or Celebrex™?” The 3 responses are: “Yes”, (“No. These do not improve my symptoms”, or “No. I do not take anti-inflammatory drugs”. A positive score is given if the patient answers “Yes” to this question. Although this item may appear to select NSAIDs respondents especially when there are patients with axSpA who are NSAIDs non-respondents, this item addresses only one aspect of axSpA and that the non-respondents are likely to score other features.

**Family History of AS**

Family history (first or second degree) of AS is a specific item that has been part of the Amor and the ESSG criteria for SpA, and in the ASAS classification criteria for axSpA. In my questionnaire, this item was asked as “Do you have a parent or sibling who has been diagnosed with ankylosing spondylitis?” The 3 responses are: “Yes”, “No”, or “Unknown”. The latter response was mainly added to include patients who are adopted. A positive score is given if the patient answers “Yes” to this question.
5.3 Extra-axial manifestations

**Prior Diagnosis of Uveitis or Iritis**

In order to make this item more sensitive, I asked if a patient was diagnosed with iritis or uveitis without asking about details of its symptoms. A positive score is given if the patient answers “Yes” to this question.

**Prior Diagnosis of Psoriasis**

Psoriasis can co-exist in 10–25% of axSpA patients. This item is part of the ASAS classification criteria for axSpA. A positive score is given if the patient answers “Yes” to this question.

**Prior Diagnosis of Psoriatic Arthritis**

This item was initially added since some patients may have a prior diagnosis of psoriatic arthritis without psoriasis. It may co-exist in IBD and result in symptoms of axSpA. A positive score is given if the patient answers “Yes” to this question.

**Peripheral Arthritis**

This item is part of the ASAS classification criteria for axSpA. In order to make this item more specific for peripheral arthritis, I asked whether the patient had pain and swelling in any joint that was not related to injury. I then increased the sensitivity of this item by adding “ever” to the description of this symptom. A picture of peripheral arthritis of the hands was initially added to the questionnaire but was removed when the SpA team felt that it might be confused with the picture of dactylitis. A positive score is given if the patient answers “Yes” to this question.

**Dactylitis**

This item represents a highly specific feature of the SpA group of diseases, and is part of the ASAS classification criteria for axSpA. I asked this specific question by describing
pain and swelling of the entire finger or toe that was not related to injury. I then increased its sensitivity by adding “ever”. As in the case of “peripheral arthritis” item, a picture of dactylitis was initially added to the questionnaire but was removed when some patients felt the pictures of dactylitis looked like that of arthritis of a finger. A positive score is given if the patient answers “Yes” to this question.

Heel Pain (Enthesitis)

This item is also part of the ASAS classification criteria for axSpA. A positive score is given if the patient answers “Yes” to this question.

6 Reliability study

6.1 Response Rate of Reliability Questionnaires

Of the 77 mailed questionnaires, 34 were returned by the cut-off date. Mails for the questionnaires of 6 patients were returned due to incorrect addresses. One patient declined participation and the mail was returned. The response rate during the summer of 2011 was 44.2%. Since I was interested in a sample size of 24, the final number of respondents (i.e. 34) was excellent. Therefore, I did not pursue follow-up mails for non-respondents.

6.2 Missing Answers on the Questionnaire

Questionnaires from 5 of 34 responders had $\geq 15\%$ missing answers in their respective questionnaires. We were able to contact 4 of them by phone as soon as the mailed questionnaires were received. We read the unanswered questions for them over the phone
as written without providing any further explanation to avoid interviewer bias. Questionnaires from 33 out of 34 patients were usable for the descriptive analysis. We could not identify one male patient due to missing date of birth on the questionnaire and inability to crosscheck the code on his questionnaire with Spondylitis database (clerical error). For the test-retest reliability, the final usable pairs of questionnaires were 33 as one patient was excluded from the analysis because of incomplete data of the second questionnaire.

6.3 **Comparison between Responders and Non-Responders**

Table 11 summarizes the demographic and baseline characteristics of patients who did and did not respond to the questionnaire. Respondents were slightly older than non-respondents. The median of the highest level of education at the last clinic visit was university in responders and college in non-responders. Ulcerative colitis was slightly more common in non-responders than responders. The median ESR level was slightly higher in non-responders than responders. Responders and non-responders were not different in terms of average age, sex, highest level of education, types of IBD, activity of axSpA, functional activities, ESR or CRP levels.
Table 11. Demographic and clinical characteristics of responders and non-responders. Values are means (SD) unless otherwise indicated

<table>
<thead>
<tr>
<th></th>
<th>Responders</th>
<th>Non-responders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 33)</td>
<td>(n = 43)</td>
</tr>
<tr>
<td>Age, yrs</td>
<td>45.7 ± 13.5</td>
<td>40.5 ± 11.7</td>
</tr>
<tr>
<td>Male (%)</td>
<td>27 (79.4%)</td>
<td>31 (72.1%)</td>
</tr>
<tr>
<td>Education, median</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Crohn’s disease (%)</td>
<td>21 (63.6%)</td>
<td>26 (60.5%)</td>
</tr>
<tr>
<td>Ulcerative colitis (%)</td>
<td>11 (33.3%)</td>
<td>18 (41.9%)</td>
</tr>
<tr>
<td>Unclassified colitis (%)</td>
<td>1 (3%)</td>
<td>0</td>
</tr>
<tr>
<td>BASDAI</td>
<td>3.7 ± 2.6</td>
<td>4.1 ± 2.6</td>
</tr>
<tr>
<td>BASFI</td>
<td>3.7 ± 3.0</td>
<td>3.5 ± 3.1</td>
</tr>
<tr>
<td>ESR mm/hr, median</td>
<td>6.5</td>
<td>13.0</td>
</tr>
<tr>
<td>CRP mg/L, median</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

* The response of one patient was dropped because of incomplete information.

Note: All $p$ values are > 0.05.

**Abbreviations:** SD, Standard Deviation; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; ESR: Erythrocyte Sedimentation Rate; CRP: C-Reactive Protein
6.4 Descriptive Statistics of the Items

I coded the items to the direction of the construct (axSpA), that is, the positive or expected answers for a patient who has axSpA (for details, please refer to Description of the Items in the Final Questionnaire). All respondents confirmed that they had back or stiffness that lasted ≥ 3 months. Among the responses of the 34 patients, the most prominent items are as follows: male gender, age of onset of back pain ≤ 45 years, slow development of back pain initially (over weeks or months), current morning stiffness of the back, current morning stiffness of the back lasting 30 minutes or more, pain or stiffness in the neck, lower back and the buttock, improvement of back pain or stiffness at all with daily physical activities, and history of peripheral arthritis. All patients were aware that they had a diagnosis of AS. The least common associated features were psoriatic arthritis and psoriasis successively. Table 12 summarized these findings.
Table 12. Summary of the responses for features of axSpA among patients with axSpA and IBD

<table>
<thead>
<tr>
<th>Items</th>
<th>Number of patients with positive responses, n = 34 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>27 (79.4)</td>
</tr>
<tr>
<td>Treatment of IBD with biologic drug that improved the back pain or stiffness</td>
<td>16 (47.1)</td>
</tr>
<tr>
<td>Age of onset of back pain ≤ 45 years</td>
<td>33 (97.1)</td>
</tr>
<tr>
<td>Slow development of back pain initially</td>
<td>26 (76.5)</td>
</tr>
<tr>
<td>Current morning stiffness of the back</td>
<td>24 (70.6)</td>
</tr>
<tr>
<td>Current morning stiffness of the back lasting ≥ 30 minutes</td>
<td>22 (64.7)</td>
</tr>
<tr>
<td>Current back pain awakening at night</td>
<td>17 (50)</td>
</tr>
<tr>
<td>Current back pain awakening at night and improving upon getting up, walking or stretching</td>
<td>11 (32.4)</td>
</tr>
<tr>
<td>Locations of back pain or stiffness</td>
<td></td>
</tr>
<tr>
<td>Neck</td>
<td>25 (73.5)</td>
</tr>
<tr>
<td>Upper back</td>
<td>14 (41.2)</td>
</tr>
<tr>
<td>Middle back</td>
<td>10 (29.4)</td>
</tr>
<tr>
<td>Lower back</td>
<td>26 (76.5)</td>
</tr>
<tr>
<td>Buttock</td>
<td>25 (73.5)</td>
</tr>
<tr>
<td>Improvement of back pain or stiffness at all with daily physical activities</td>
<td>26 (76.5)</td>
</tr>
<tr>
<td>No improvement of back pain or stiffness at all with rest</td>
<td>19 (55.9)</td>
</tr>
<tr>
<td>Improvement of back pain after taking NSAIDs</td>
<td>19 (55.9)</td>
</tr>
<tr>
<td>Family history (first degree) of AS</td>
<td>7 (20.6)</td>
</tr>
<tr>
<td>Prior diagnosis of AS</td>
<td>34 (100)</td>
</tr>
<tr>
<td>Prior diagnosis of iritis or uveitis</td>
<td>12 (35.3)</td>
</tr>
<tr>
<td>Prior diagnosis of psoriasis</td>
<td>3 (8.8)</td>
</tr>
<tr>
<td>Prior diagnosis of psoriatic arthritis</td>
<td>1 (2.9)</td>
</tr>
<tr>
<td>History of peripheral arthritis</td>
<td>26 (76.5)</td>
</tr>
<tr>
<td>History of dactylitis</td>
<td>12 (35.3)</td>
</tr>
<tr>
<td>History of heel enthesitis</td>
<td>21 (61.8)</td>
</tr>
</tbody>
</table>
At this stage, the committee members agreed on removing a response option “Prior diagnosis of psoriatic arthritis” from item 13 (for more details, see Discussion - Test-Retest Reliability). The final version 5 of TASQ (dated May 2012) is shown in Figure 10.
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:

<table>
<thead>
<tr>
<th>Remicade</th>
<th>Currently</th>
<th>Previously</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Humira</th>
<th>Currently</th>
<th>Previous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>Currently</th>
<th>Previous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

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

© Copyright. Version 5 (May 2012)    END

Figure 9. The Toronto Axial Spondyloarthritis Questionnaire (TASQ)
6.5 **Test-Retest (Intra-rater) Reliability**

Thirty-two pairs of questionnaires, indicating patients who completed 2 questionnaires at both times, were used for analysis out of the 33 respondents. At this stage, I calculated the kappa (κ) coefficients for the questionnaire items including the ones which are not going to be ultimately scored such as the IBD type and the different types of biologic agents (Table 13). The κ coefficients of all items ranged between 0.81 to 1.00 which indicate almost perfect agreements for test-retest reliability. The absolute percentage of agreement across all items ranged from 91% to 100%.

The range of κ coefficients for each domain is as follows: 0.84 – 1.00 for the (IBD domain), 0.85 – 1.00 for (Back pain or stiffness) domain, and 0.81 – 1.00 for (extra-axial manifestations) domain. The ranges of absolute agreement between different domains are as follows: 91 – 100% for the (IBD domain), 94–100% for (inflammatory back symptoms) domain, and 91 – 100% for (extra-axial manifestations) domain.

The lower limit of 95% confidence interval of κ coefficients for some items ranged from 0.60 to 0.78 (moderate to substantial agreement) although κ was > 0.80 (almost perfect correlation). These items include IBD type, pain and stiffness in the neck, middle back and buttock, no improvement of back pain or stiffness with rest, history of peripheral arthritis and history of dactylitis.
Table 13. Results of the test-retest reliability analysis

<table>
<thead>
<tr>
<th>Questions</th>
<th>Kappa coefficient (95% CI)</th>
<th>Percentage of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male gender</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>IBD type</td>
<td>0.84 (0.66 – 1.00)</td>
<td>91</td>
</tr>
<tr>
<td>Biologic agents for IBD</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Any type of biologic agent</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Improvement of back pain or stiffness after taking biologic agents</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Age of onset of back pain ≤ 45 years</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Time to develop</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Presence of AM stiffness</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Morning stiffness lasting ≥ 30 minutes</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Back pain awakens at night</td>
<td>0.94 (0.82 – 1.00)</td>
<td>97</td>
</tr>
<tr>
<td>Current back pain awakening at night improves upon getting up, walking or stretching</td>
<td>0.95 (0.86 – 1.00)</td>
<td>97</td>
</tr>
<tr>
<td>Locations of back pain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neck</td>
<td>0.91 (0.75 – 1.00)</td>
<td>97</td>
</tr>
<tr>
<td>Upper back</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Middle back</td>
<td>0.93 (0.78 – 1.00)</td>
<td>97</td>
</tr>
<tr>
<td>Lower back</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Buttock</td>
<td>0.85 (0.65 – 1.00)</td>
<td>94</td>
</tr>
<tr>
<td>Improvement with physical activity</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>No improvement with rest</td>
<td>0.88 (0.71- 1.00)</td>
<td>94</td>
</tr>
<tr>
<td>Improvement with NSAIDs</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Family history of AS</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>AS</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Iritis / Uveitis</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Peripheral arthritis</td>
<td>0.92 (0.77 – 1.00)</td>
<td>97</td>
</tr>
<tr>
<td>Dactylitis</td>
<td>0.81 (0.60 – 1.00)</td>
<td>91</td>
</tr>
<tr>
<td>Heel pain</td>
<td>1.00</td>
<td>100</td>
</tr>
</tbody>
</table>

*Abbreviations: CI, Confidence Interval; IBD, Inflammatory Bowel Disease; NSAIDs, Non-Steroidal Anti-Inflammatory Drugs; AS, Ankylosing Spondylitis*
Chapter 5

Discussion
1 Questionnaire Development

The TASQ self-administered questionnaire was designed to serve as a case finding instrument that could facilitate referrals of IBD patients with suspected axSpA to rheumatologists for further evaluation. Early diagnosis allows for earlier intervention, an important concept especially since TNF inhibitors have been found to be efficacious in early disease. Because focusing only on characteristics of IBP cannot be sufficient to make a diagnosis of axSpA, additional features are required, such as other relevant clinical information (family history, articular symptoms, extra-axial manifestations), laboratory tests (HLA-B27, ESR, CRP), and imaging (X-ray, MRI).

Previous questionnaires focused on refining the characteristics of IBP and did not specifically target IBD patients. TASQ is to be administered at a single point in time to patients with IBD who have chronic back pain or stiffness that has ever lasted 3 months or more. This cut-off time which was chosen to minimize the referrals of mechanical back pain was also used in previous studies (93, 106, 121). TASQ is cheap, easy to administer and easy to understand by an average student in the 5th grade. Completion of the questionnaire takes 5 minutes or less which is in accordance to the recommended completion time between 5 and 15 minutes (122). This also facilitates its feasibility especially when this questionnaire is handed to IBD patients in a busy gastroenterology clinic or even when used online.

Designing a new questionnaire requires multiple steps to ensure a proper methodology in order to yield a reliable instrument. During the development stage, I included sensitive and specific items to help capture different aspects of the construct being sought i.e. axSpA in IBD. I organized it into 3 domains (IBD, inflammatory back symptoms, and extra-axial features). I strove to simplify the words in order to achieve our aims. Hypothetically these questions represent a wide spectrum of discriminatory features that increase the likelihood of axSpA and, therefore, can help gastroenterologists refer those patients in a timely fashion.
The clinimetric method, as introduced by Feinstein, advocates for the use of clinical judgment when developing a measuring instrument. The tools that I used for this approach included sensibility and reliability. In 1993, Rowe and Oxman paved the road shown by Feinstein and developed a questionnaire for sensibility assessment. Twelve items were developed based on the principles of sensibility. The response options were constructed on a 7-point Likert scale. Some of their questions appeared redundant, and some contained double-barreled questions. This questionnaire was adapted during development of recent measuring instruments such as the Pediatric Cardiopulmonary Physiotherapy Discharge Tool (123), and HIV Disability Questionnaire (124). I also adapted the sensibility questionnaire by modifying some of their questions and adding new ones for feasibility. New items related to the concept of feasibility (ease of use) consist of time to completion (in minutes), flow of questions, readability, typographical errors, font size, and use of illustrations. I also changed the format of scaling responses into dichotomous scale which goes along the theme of the scaling responses of TASQ.

The sensibility assessment proved to be a valuable and comprehensive approach during selection of the items and pilot testing that enabled me to look at different facets of the questionnaire and remedy its weaknesses. Using the sensibility assessment, the questionnaire underwent 4 revisions incorporating feedback from the SpA team, general rheumatologists, and patients before reaching the fourth and final version.

2 Clinimetric Measures in the Literature

Psychometric methods use classical test theory and item response theory. The former involved the statistical analysis of the data to yield reliability (test retest reliability, internal consistency, split half reliability) and validity. Item response theory uses other statistical models with different assumptions and requires larger sample size. It is obvious that these methods rely heavily on statistical analyses to select items of an index or questionnaire. With regards to assessing the internal consistency of an instrument, it is required to have high correlations between the items (inter-item correlation) and the
whole instrument (items-total correlation). While having a high internal consistency coefficient is preferred, it may give a false impression that items are homogenous. For example, the coefficient can increase by only increasing the number of questions (even if they are redundant). In this case, the sensitivity of the items may actually be low which consequently lowers the discriminatory property of the items (125). Having a high coefficient of internal consistency requires an appropriate, and usually large, sample size. Psychometric methods are not concerned with face validity and sensibility (usefulness of the instrument). In addition, they do not necessarily result in constructing an accurate measure (125).

Feinstein emphasized that judgment (by physicians and/or patients) is an essential part during development of an instrument measuring a complex clinical phenomenon and that psychometric techniques may be used to aid this process. Assessing internal consistency is of less relevance here because the aim of developing a measuring tool is to capture different aspects of the same (uni-dimensional) construct. Therefore, the internal consistency is likely to be low although this is not always the case (126). There are some indices which fell out of favor because they were not sensible and were difficult to apply such as the Norris prognostic index for myocardial infarction (110). The clinimetric approach has been used in the literature to construct indices which can comprise classification or diagnostic criteria, guidelines, prognostic criteria, and questionnaires. Examples of clinimetric measures are the New York Heart Association Functional Classification (127), the Jones criteria for rheumatic fever (128), the Apgar criteria to score the health of a newborn immediately after birth (129), and the Pittsburgh Sleep Quality Index (130).

Two independent studies compared 2 different versions of questionnaires (Quality of Life After Myocardial Infarction and Asthma Quality of Life Questionnaire) using methods of clinical judgment and statistical analysis for item reduction. In each study, the clinical judgment method performed slightly better than the statistical method when both versions were tested for concurrent validity and responsiveness.
There are important issues that were raised for discussion during the sensibility assessment of TASQ. These will be discussed next.

3 Sensibility Assessment of TASQ

3.1 Comprehensibility (Transparency)

Assigning appropriate weighting to each item of the questionnaire is a challenge. There are 4 methods that can shed some light on item weighting. The first method of scoring is based on clinical judgment and intuition. For example, a total score can simply be the sum of each item that is weighted equally with a score of 1. This gives the questionnaire the advantages of being simple, easy to score and transparent. Pediatric Crohn’s Disease Activity Index (PCDAI) was scored using this method (131). In addition, the contribution of each item weight to the cumulative score will be clear as opposed to a questionnaire that assigns different weights to each item. The disadvantage of this approach is the implication that every item is of equal importance. The second method assigns a weighted score for each domain e.g. double the total score of Domain 1 and half Domain 2. This indicates the greater significance of a domain and the lesser significance of another. The third method uses a statistical test (factor analysis) where assigning weights is based on factor loadings. However, this often requires large sample size i.e. 5 patients for each item. In addition, the interpretation of factor analysis requires clinical judgment. The fourth method relies on another statistical analysis i.e. multivariate logistic regression analysis that quantifies each item depending on its importance in explaining the outcome (dependent variable). This is the basis of scoring the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) and Crohn’s Disease Activity Index (CDAI) (132, 133). While it may be scientifically appropriate to weight items based on statistical analysis, the different statistical tests may give different final scores which also depend on the characteristics of the population and the setting. Some scales were scored using statistical models that resulted in assigning sophisticated scoring that requires a
special calculator. A clear example is the scoring system for the case ascertainment questionnaire.

A recent Canadian questionnaire for inflammatory back pain (for details, refer to Section 4 of the Discussion) compared 2 methods of weighting items: summative (adding a score of 1 for each item) and weighted items based on multivariate regression analysis. The sensitivity, specificity, positive LR and the area under the curve were similar when any method was used (134).

Another recent Canadian study compared weighting of the Pediatric Ulcerative Colitis Activity Index (PUCAI) using mathematical and judgmental models. The PUCAI has 8 items including 6 historical and 2 laboratory parameters. Judgmental approach was found to be as good as the mathematical approach only if the 2 laboratory items were removed (135).

In my questionnaire, although there were 2 questions which are not going to be scored (type of IBD and previous diagnosis of AS), I retained them in the final version to provide important descriptive data especially that their removal would slightly reduce the length of the questionnaire.

In conclusion, I will explore the scoring of the questionnaire in the next validation phase and will compare the different scoring methods.

3.2 Content Validity

Many items had to be dropped because I wanted to keep the questionnaire simple and concise (on 2 pages). A recent Cochrane systematic review found that the response rates of patients in clinical trials increased when shorter self-reported questionnaires were mailed (136). For example, the odds ratios (ORs) per page increase were 0.90 (95% CI 0.83–0.98) for one page compared with ≥ 4 pages, and 0.98 (95% CI 0.96–0.99) for one page compared with ≥ 2 pages.
During the pilot testing, the item “fibromyalgia” was removed for 3 reasons related to low discriminatory property: 1) patients with true axSpA may have enthesitis that could have been misdiagnosed with fibromyalgia, 2) enthesitis may overlap with tender points of fibromyalgia, and 3) fibromyalgia can co-exist independently in both axSpA and IBD as part of its association with some chronic diseases (137, 138). In addition, patients did not endorse this item in the pilot study.

3.3 Feasibility

3.3.1 Acceptability

An item on “race” was included in the Phase 2 (item generation) since axSpA is more common in Caucasian people of European extraction. Therefore, I considered ethnicity in this instance a risk factor for axSpA. During Stage 3 of the development of TASQ, one member of our team pointed towards the inappropriate inclusion of an item about race. In a recent Cochrane systematic review, the likelihood of not responding to mailed questionnaires was shown to increase when they asked about sensitive questions (139). I then removed this item in order to increase the response rate.

3.3.2 Readability

Clarity of Questions

Constructing a clear questionnaire proved to be a challenge. Patients may not necessarily read and understand what health care workers perceive as a straightforward question since each comes from different backgrounds. In this study, I had to rephrase certain questions a few times and even drop certain items because they were unclear to patients,
even thought they were relevant to the complex construct of axSpA. I will elaborate more on this in the following sections.

**Insidious onset of back pain:** The item has long been considered vague, as previous studies did not agree on its exact definition. It is intended to measure the mode of onset of back pain. However, does it mean slow onset or gradual onset? Can it be measured that in weeks or months or even years? As a result of this unclear definition, studies showed an overlap between mechanical low back pain and IBP when the poorly defined terms “chronic” or “insidious onset” were used (93, 94). In the case ascertainment questionnaire of Weisman and colleagues, the questionnaire developers defined “insidious onset” by asking about the length of time (in months) since patients started to have back pain or stiffness. At the end, I decided to keep the question as “When your back pain first began, did it develop quickly (over hours or days) or slowly (over weeks or months). All patients during the pilot and reliability studies were able to answer it easily.

**Sleep disturbance at the second half of the night:** Rudwaleit and colleagues found this item an important characteristic of IBP when they administered their questionnaire (in German) by an interviewer (93). I am unaware of any published cross-cultural adaptation of their questionnaire. However, I used this item during the pilot testing phase but decided to drop it because some patients found it difficult to define a cut-off for the second half of the night.

**Hip arthritis:** Hip joint is localized in a deep-seated site, which may prevent patients from expressing their pain clearly. The diagnosis of hip arthritis is often based on clinical (history and physical examination) and radiographic evidence. There is a strong association between hip arthritis and AS (37). A study showed that hip arthritis is significantly associated with primary AS compared to IBD-associated AS (36). However, I believe that this item is important and merits asking in a questionnaire. The problem was how to ask about it properly using a single and relatively short question to fit in a case-finding questionnaire. In clinical practice, rheumatologists ask patients different questions about hip pain to increase the pre-test probability of hip arthritis. For example, they may ask whether patients have pain or stiffness in the upper frontal side of the thigh.
that radiates to the groin. Some ask whether patients feel or even hear a click during hip joint movements. Some patients point towards the painful area by cupping their thumb and index fingers around the hip joint, which can also be seen in other conditions such as trochanteric bursitis or fibroacetabular impingement syndrome. In the pilot study, some patients drew a line (without being asked) to point towards hip pain but they directed the line at either the buttock or sacroiliac joint. At the end, I could not include this item in our questionnaire due to lack of a standardized definition and its susceptibility to misinterpretation. A recent study attempted to describe hip pain for patients by using of a diagram with pre-shaded area for localizing the hip pain and compare it with a standardized question asking, “In the past month have you had any pain in the hip lasting one day or longer?” This study concluded that the use of combined methods was associated with signs of hip disease (abnormal physical examination and radiographic changes). However, if these combined methods were used to exclusively define hip pain, the sensitivity of capturing those patients would be low which does not go in the same direction of our questionnaire (140). Future studies are needed to best describe hip pain to patients.

**Alternating buttock pain:** This item was found to be one of the specific characteristics of IBP (although not sensitive) and was therefore incorporated in Amor criteria and in Rudwaleit’s study (93, 102). However, we had to drop it for 3 reasons: 1) some patients had difficulty understanding the question, 2) buttock pain can still be a complaint in some patients with AS and mechanical back pain, and 3) this particular item was not included in the description of IBP of the ASAS classification criteria for axSpA (105).

**Illustrations**

TASQ is unique compared to the previous questionnaires in that it shows for the first time a diagram of the back to allow patients to identify the locations of their back pain or stiffness. I initially wanted to include a location for the ribcage since this has been noted in some studies. But I decided against this idea because it will make the diagram cluttered with too many arrows.
The decision to remove the colored pictures of peripheral arthritis and dactylitis was reached by consensus from the committee members and from the feedback of patients. These pictures could be mutually misinterpreted. The picture of iritis was also removed during selection of the items because it may be misinterpreted as conjunctivitis. However, a future study is required to refine symptoms of iritis particularly if associated with axSpA in IBD patients. A disadvantage of using colored pictures is the expense which may limit the questionnaire feasibility. Questionnaires with colored pictures can be used on the website but can only be used by patients who have access to the internet.

Recently, Gladman and colleagues developed a screening questionnaire for psoriatic arthritis among patients with psoriasis and in general population; the Toronto Psoriatic Arthritis Screening (ToPAS) (141). ToPAS has 3 domains, namely skin, joint and nail. Among all the screening questionnaires for psoriatic arthritis, ToPAS was the first to include colored pictures of the skin rash, and nail changes. A Dutch study found that ToPAS performed slightly better compared to the Psoriatic Arthritis Screening and Evaluation (PASE) questionnaire which could be partly related to the use of illustrations (142). An update of ToPAS with more colored pictures is underway (143).

4 Reliability

4.1 Rating Responses of the Questionnaire Items

Upon looking at the responses of TASQ in the reliability study, all but one patient started to have back pain or stiffness below the age of 45 years. The most prominent items among patients with established IBD and axSpA were as follows: male gender, age of onset of back pain $\leq$ 45 years, slow development of back pain initially (over weeks or months), current morning stiffness of the back, current morning stiffness of the back lasting 30 minutes or more, pain or stiffness in the neck, lower back and the buttock,
improvement of back pain or stiffness at all with daily physical activities, and history of peripheral arthritis.

4.2 Test-Retest Reliability

I chose test-retest (intra-rater) reliability to assess the stability of the items over time in order to quantify measurement error rather than true change in disease activity. I 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 2011 Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (115).

There are various criteria for the interpretation of $\kappa$ coefficients. These include the criteria of Landis and Koch, Cicchetti and Sparrow, and Fleiss (118, 144, 145). It is important to note that all these criteria are arbitrary. In our study, I chose the criteria of Landis and Koch. My a priori hypothesis was that the reliability coefficient (kappa or $\kappa$) would be estimated as 0.90. The $\kappa$ coefficient for each item was more than 0.80 indicating almost perfect agreement. Similarly, the absolute agreement for each item was more than 91%. Although it may seem that some items had reliability coefficients (range 0.81 – 0.88) of less than the hypothesized reliability of 0.90, the minimally accepted reliability for this study was 0.80. Therefore, those items are still within acceptable range of reliability i.e. > 0.80 indicating almost perfect agreement.

I observed that the 95% CIs of $\kappa$ coefficients for some items were wide and that may be due to our small sample size. If we used a larger sample size, we may improve the precision around the $\kappa$ statistics.

The sub-question “Prior diagnosis of psoriatic arthritis” was removed after the reliability study for 2 reasons. First, this item was rarity endorsed by patients (only 1 patient). Second, during the validation phase of the case ascertainment questionnaire, this item did not help differentiating AS from chronic back pain (107). We considered psoriasis
adequate for this questionnaire, as previous studies found it co-exist in 10-25% of AS patients. Furthermore, psoriasis was recently shown to be more associated with non-radiographic axSpA (OR 3.6) compared to classic AS (146). This additional item reduction led to version 5 (dated May 1st, 2012).

5 Recent Questionnaire for Back Pain (April 2012)

A newly self-reported questionnaire has been published this year (April 2012). It is a screening questionnaire for IBP developed by the Edmonton group of SPARCC. It consists of 6 items with 3 branching questions. Items address morning stiffness of the back and/or hip (branching question asked about the most noticeable time), nocturnal pain (branching question asked about time of nocturnal awakening), diurnal variation of pain in the back and/or hip, peripheral arthritis, response to exercise, and response to rest. The questionnaire was administered to patients with established AS and mechanical back pain (MBP).

In univariate analysis, 4 items distinguished IBP from MBP: morning stiffness, diurnal variation, response to exercise and response to rest. In multivariate logistic regression analysis, the last 3 items were independently associated with IBP, especially diurnal variation (p = 0.0001). The “diurnal variation” item asked, “At what time of day your back and/or hip symptoms the worse? Morning, Afternoon, Evening, Night, or Not Applicable.” Using all the 6 items, the questionnaire outperformed Calin criteria and the IBP criteria of ASAS in identifying patients with IBP. In my opinion, the way in which the item “diurnal variation” was asked seems easier to understand than asking about nocturnal pain at the second half of the night as suggested from Rudwaleit’s study (93). In addition, this item showed the strongest association with IBP (OR 11.2, CI 3.7 – 34.2), with a specificity of 92% and a sensitivity of 49%. As in the case of the previously published questionnaires (Calin and case ascertainment questionnaire), the 6-item
questionnaire is not targeted to IBD patients, and focused on features of IBP with one question on peripheral arthritis.

As discussed earlier, investigators, who are interested in early diagnosis of axSpA, have long been searching for the optimal characteristics of IBP which can improve the current increase in post-test probability from 5 to 14%. Further, the interpretation of IBP needs some clinical experience. Therefore, IBP cannot be used solely as a referral criterion. The recommendation is to combine IBP with other features of the complex disease (axSpA) in order to achieve a high post-test probability of axSpA (more than 90%) (17, 104, 146). Indeed, IBP is no longer considered by ASAS to be a mandatory element in the definition of axSpA. My questionnaire is in agreement with this, since other elements in addition to characteristics of the back pain, are given equal importance.
Chapter 6

Conclusions
TASQ is a newly developed self-reported instrument to be administered to patients with IBD who have ever had chronic back pain or stiffness that has lasted 3 months or more. It consists of 3 domains (inflammatory bowel disease, inflammatory back symptoms, and extra-axial features) and 16 items. The items were chosen based on their high sensitivity (i.e. found in most patients with axSpA) and reasonable specificity which is in accordance with a purpose of case-finding questionnaire. I have demonstrated that TASQ is sensible, reliable, and inexpensive. It is easy-to-apply and requires 5 minutes of less to complete. It is suitable for a student in the 5th grade.

TASQ is a very promising questionnaire that might facilitate early patients’ referral to rheumatologists and avoid delay in diagnosis of axSpA.
Chapter 7
Limitations and Future Directions
While the project achieved it primary goals, there are some limitations to point out. The following sections will discuss the limitations of this research and future directions.

1 Development

The pilot study was conducted at the Spondylitis clinic 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. However, we only had 2 questions specific to IBD patients (type of IBD, response of back pain or stiffness to biologic drugs). In addition, patients in the pilot study provided their feedback on the sensibility of the IBD items. Ideally, a random sample of axSpA patients would be representative as it should include a mixture of patients with different levels of disease severity and levels of education. The sampled patients represented a wide spectrum of axSpA disease activity. The questionnaire is limited to patients who have finished grade 5. It could potentially be too sophisticated for patients who have lower educational levels. If this were truly shown in a future study, 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.

Some features of IBP need to be clarified in future studies. As described in the Results, defining “alternating buttock pain” using simple words was challenging and led to dropping this item from the final version. On the same token, finding an accurate definition for “hip pain” needs to be studied in the future since hip arthritis is a feature of axSpA. The classic item of Calin criteria “improvement of back pain or stiffness with exercise” remains to be further clarified. It should be noted that this item is also observed in patients with mechanical back pain. Does improvement of IBP imply any degree of exercise? Is the improvement expected to occur after a certain period of time?
Future modifications for this questionnaire may emphasize shoulder involvement in axSpA patients. Chronic rotator cuff and enthesitis at the insertion of the supraspinatus muscle to the bone are common manifestations of axSpA. Inclusion of this item can be achieved with the use of a diagram.

2 Future Validation Studies

The TASQ was demonstrated to be reliable in patients who had IBD and axSpA and attended a hospital-based rheumatology clinic. An important next step is validation is this questionnaire in gastroenterology clinics. We are currently in the process of validating the TASQ through collaboration with the gastroenterologists at the University Health Network including Mt. Sinai, Toronto General and Toronto Western Hospitals. Inclusion criteria for the study include age ≥ 18 years, fluency in English, established IBD diagnosis, chronic back pain or stiffness ever present ≥ 3 months in duration, and competency to consent. Patients will be recruited and referred to the Spondylitis Clinic for a thorough assessment according to a standardized protocol. They will complete the TASQ in the waiting area. An advanced practice physiotherapist, who will be blinded to the patients’ answers of the questionnaire, will then assess them. The diagnosis of back symptoms will be determined. The ideal situation for validation of the questionnaire requires 2 groups of IBD patients: patients with mechanical back pain and patients with axSpA. Finally, the type of validity will be assessed such as concurrent construct validity.

Another validation of TASQ is underway in which we have collaborated with gastroenterologists at a hospital-based IBD clinic in the US. This validation is based on a case-control study. IBD patients have undergone CT scans of the abdomen and pelvis at some point of time as part of investigations for their IBD. We determined that the clinic has on file a bank over 900 CT scans of which approximately 130 patients were found to have definite sacroiliitis on CT scan. These patients likely have IBD-associated AS. The controls will consist of 200 IBD patients with no sacroiliitis on CT scan. After having just
recently obtained approval of the institutional research board at that hospital, my new questionnaire will be administered to those patients either via email or mail. The scores of the questionnaire will be determined using the receiver operator curve with the aim of having high sensitivity.

A third future study is to validate the TASQ in the primary care setting in patients who have IBD. This will involve approaching the Family Practice Units of the University of Toronto affiliated teaching hospitals.

The reliability study of this questionnaire was conducted in patients with established IBD and axSpA. This questionnaire applies to a subset of seronegative SpA patients who have IBD and undetected axSpA. With the exception of IBD domain, there is no reason to assume that this questionnaire cannot be applied to any patient at risk for axSpA such as patients with reactive arthritis or uveitis. However, this would require future validation.
15. 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

Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-
B27 status in early inflammatory back pain predict radiographically evident

17. Rudwaleit M, Sieper J. Referral strategies for early diagnosis of axial

18. Baklond G, Gran JT, Nossent JC. Increased mortality in ankylosing spondylitis is

Clinical Features of Late-onset Ankylosing Spondylitis: Comparison with

response to a local infliximab injection in a patient with chronic sacroiliitis.

gender differences in severity of ankylosing spondylitis? Results from the

22. McGonagle D, Gibbon W, Emery P. Classification of inflammatory arthritis by


24. Zeboulon N, Dougados M, Gossec L. Prevalence and characteristics of uveitis in
the spondyloarthropathies: a systematic literature review. Ann Rheum Dis.

25. Van Praet L, Van den Bosch F, Mielants H, Elewaut D. Mucosal inflammation in
spondylarthritides: past, present, and future. Curr Rheumatol Rep

26. El Maghraoui A. Extra-articular manifestations of ankylosing spondylitis:
2011;22(6):554-60.

Mosby; 2010.

approach to defining disease status in ankylosing spondylitis: the Bath
Ankylosing Spondylitis Disease Activity Index. J Rheumatol.

approach to defining functional ability in ankylosing spondylitis: the
development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol.
1994;21(12):2281-5.


42. Paccou J, Solau-Gervais E, Houvenagel E, Salleron J, Luraschi H, Philippe P, et al. Efficacy in current practice of switching between anti-tumour necrosis factor-


Appendix A: Consent to participate in a research study

Date: 13 July 2011

Dear Ms. / Mrs. / Mr.,

Many patients with inflammatory bowel disease such as Crohn’s disease, ulcerative colitis, or indeterminate (unclassified) colitis are at high risk for developing inflammatory back disease which is the main symptom of a group of chronic back diseases called spondyloarthritis. We have developed a new questionnaire that could help in earlier detection of inflammatory back disease in patients with inflammatory bowel disease in order to facilitate early treatment, with the aim to improve their quality of life. Our goal now is to make sure this questionnaire is reliable.

We are asking you to complete a questionnaire (marked as 1), and then mail it back to us. After about 1 week, fill out the same questionnaire (marked as 2) and send it back to us by 25th July 2011.

Two stamped, self-addressed envelopes have been provided.

We appreciate you time and cooperation with our request.

Sincerely,

Dr. Robert D. Inman, MD
Director of the Spondylitis Program
Director of the Arthritis Centre of Excellence