

Development, Sensibility, and Validity of a Systemic Autoimmune Rheumatic Disease Case Ascertainment Tool

Susan M. Armstrong, Joan E. Wither, Alan M. Borowoy, Carolina Landolt-Marticorena, Aileen M. Davis, and Sindhu R. Johnson

ABSTRACT. Objective. Case ascertainment through self-report is a convenient but often inaccurate method to collect information. The purposes of this study were to develop, assess the sensibility, and validate a tool to identify cases of systemic autoimmune rheumatic diseases (SARD) in the outpatient setting.

Methods. The SARD tool was administered to subjects sampled from specialty clinics. Determinants of sensibility — comprehensibility, feasibility, validity, and acceptability — were evaluated using a numeric rating scale from 1–7. Comprehensibility was evaluated using the Flesch Reading Ease and the Flesch-Kincaid Grade Level. Self-reported diagnoses were validated against medical records using Cohen’s κ statistic.

Results. There were 141 participants [systemic lupus erythematosus (SLE), systemic sclerosis (SSc), rheumatoid arthritis, Sjögren syndrome (SS), inflammatory myositis (polymyositis/dermatomyositis; PM/DM), and controls] who completed the questionnaire. The Flesch Reading Ease score was 77.1 and the Flesch-Kincaid Grade Level was 4.4. Respondents endorsed (mean \pm SD) comprehensibility (6.12 \pm 0.92), feasibility (5.94 \pm 0.81), validity (5.35 \pm 1.10), and acceptability (3.10 \pm 2.03). The SARD tool had a sensitivity of 0.91 (95% CI 0.88–0.94) and a specificity of 0.99 (95% CI 0.96–1.00). The agreement between the SARD tool and medical record was $\kappa = 0.82$ (95% CI 0.77–0.88). Subgroup analysis by SARD found κ coefficients for SLE to be $\kappa = 0.88$ (95% CI 0.79–0.97), SSc $\kappa = 1.0$ (95% CI 1.0–1.0), PM/DM $\kappa = 0.72$ (95% CI 0.49–0.95), and SS $\kappa = 0.85$ (95% CI 0.71–0.99). The screening questions had sensitivity ranging from 0.96 to 1.0 and specificity ranging from 0.88 to 1.0.

Conclusion. This SARD case ascertainment tool has demonstrable sensibility and validity. The use of both screening and confirmatory questions confers added accuracy. (First Release November 1 2016; J Rheumatol 2017;44:18–23; doi:10.3899/jrheum.160327)

Key Indexing Terms:

SYSTEMIC AUTOIMMUNE RHEUMATIC DISEASES
SYSTEMIC LUPUS ERYTHEMATOSUS
SJÖGREN SYNDROME

SCLERODERMA
SYSTEMIC SCLEROSIS
RHEUMATOID ARTHRITIS

Self-reported instruments may be helpful for case ascertainment in the population and are widely used for collecting information regarding a patient’s health status. However, their utility in research is limited by problems with accuracy, impeding their case-finding ability^{1,2,3}. This limitation can

greatly undermine the findings of studies that rely on self-report. Thus, availability of valid case-finding instruments is important.

Systemic autoimmune rheumatic diseases (SARD) such as systemic lupus erythematosus (SLE), systemic sclerosis

From the Toronto Scleroderma Program, Division of Rheumatology, Department of Medicine, Toronto Western and Mount Sinai Hospitals; Division of Genetics and Development, and Health Care and Outcomes Research, Toronto Western Research Institute, University Health Network; Arthritis Centre of Excellence, Division of Rheumatology, Departments of Medicine and Immunology, University Health Network; Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.

Funded by a Strategic Operating Grant from The Arthritis Society of Canada to Dr. J. Wither.

S.M. Armstrong, PhD, Toronto Scleroderma Program, Division of Rheumatology, Department of Medicine, Toronto Western and Mount Sinai Hospitals, University of Toronto; J.E. Wither, MD, PhD, Senior Scientist, Division of Genetics and Development, Toronto Western Research Institute, University Health Network, and Arthritis Centre of Excellence, Division of Rheumatology, Departments of Medicine and Immunology,

University Health Network, and University of Toronto; A.M. Borowoy, MD, Division of Rheumatology, Department of Medicine, Toronto Western Hospital, and University of Toronto; C. Landolt-Marticorena, MD, PhD, Toronto Western Research Institute, University Health Network; A.M. Davis, PhD, Senior Scientist, Health Care and Outcomes Research, Toronto Western Research Institute, University Health Network, and Institute of Health Policy, Management and Evaluation, University of Toronto; S.R. Johnson, MD, PhD, Toronto Scleroderma Program, Division of Rheumatology, Department of Medicine, Toronto Western and Mount Sinai Hospitals, and Institute of Health Policy, Management and Evaluation, University of Toronto.

Address correspondence to Dr. S.R. Johnson, Division of Rheumatology, Ground Floor, East Wing, Toronto Western Hospital, 399 Bathurst St., Toronto, Ontario M5T 2S8, Canada. E-mail: Sindhu.Johnson@uhn.ca

Accepted for publication September 21, 2016.

(SSc), inflammatory myositis [polymyositis (PM), dermatomyositis (DM)], Sjögren syndrome (SS), and rheumatoid arthritis (RA) are multisystem, chronic illnesses associated with high morbidity and mortality. Self-reported instruments are frequently used in studies of SARD, but inconsistencies between self-reported and physician-reported diagnoses have been described^{2,3,4,5}. In a study by Cooper, *et al*³ assessing whether the healthy relatives of patients with a known SARD were at a higher risk of developing an autoimmune disease than the general population, a questionnaire was administered to 893 family members of patients with SARD asking whether they had ever been diagnosed with any of 11 autoimmune diseases. Although 178 subjects reported an autoimmune disease, the authors were able to corroborate fewer than half these diagnoses.

The aim of our study was to develop a tool that would more accurately identify individuals with a SARD. Using the questionnaire of Cooper, *et al*³ as a template, several modifications were made with the objectives of improving the operating characteristics of the questionnaire and enhancing readability. We then evaluated the sensibility and validity of this new SARD tool.

MATERIALS AND METHODS

Instrument design. The questionnaire from the study of Cooper, *et al*³ was used as a template with 11 questions asking “Has a doctor ever told you that you had...” followed by the diseases SLE, RA, SSc, PM or DM, SS, antiphospholipid antibody syndrome (APS), hemolytic anemia, multiple sclerosis, thyroid disease, Type 1 diabetes mellitus (DM), and idiopathic thrombocytopenia purpura (ITP). This was followed by a dichotomous response option of “yes” or “no”. Because the terms “idiopathic thrombocytopenia purpura” and “antiphospholipid antibody syndrome” are less widely known to the general public, the question of whether a participant had the disease was preceded by a clinical question. For example, the ITP question was preceded by a question asking whether the participant had low platelets (yes or no). For APS, the preceding question asked participants if they had “recurrent blood clots or at least 3 pregnancy losses” with a response option of “yes” or “no”. These 11 questions were considered the screening questions.

The questionnaire was modified by the addition of “confirmatory questions” after the screening question. The use of confirmatory questions has previously been reported to improve the positive predictive value and agreement in patients who have been asked whether they had SLE or RA¹. The confirmatory questions inquired about medication taken for the condition, symptoms of the SARD, or treatment by a particular specialist. Each confirmatory question was followed by the response options “yes,” “no,” and “I don’t know” (the SARD Participant Questionnaire is available online at jrheum.org as a data supplement). In consultation with an expert in health instrument design and evaluation (AMD) and application of the Dillman methods⁶, the visual format and readability of the tool was enhanced by the addition of white space, blocking and shading of alternating questions, and increased font size. These methods have been shown to improve the readability of tools, particularly in the classification of SARD⁷. Because this questionnaire was designed for research purposes to identify patients who have a SARD, it was further reviewed by a community rheumatologist and an academic rheumatologist-scientist (AMB, JW) to assess its applicability for this intended use.

Subjects. Participants with SARD and disease controls (hemolytic anemia, multiple sclerosis, thyroid disease, and Type 1 DM) were recruited from hospital-based specialty clinics at the Toronto Western Hospital in Toronto,

Ontario, Canada, including the Scleroderma Clinic, the Sjögren’s Disease Clinic, the STAT Clinic (an urgent rheumatology referral clinic), and the Early Autoimmune Rheumatic Disease Clinic. Consecutive participants were approached in the waiting room. Patients were excluded if they were under the age of 18 or had difficulty reading and conversing in English. Healthy controls were recruited from hospital staff and associates. The participants were provided with a hard copy of the questionnaire and were allowed to complete it at their own pace. Study personnel were available to answer any questions that might arise during completion of the questionnaire. Ethics approval was obtained from the Research Ethics Board of the University Health Network (12-5189-BE). Participants provided written, informed consent.

Sensibility assessment. Sensibility is an important measurement property of an instrument, evaluating attributes of its usefulness. Determinants of sensibility include its clinical function, justification, applicability, format (comprehensibility and readability), face and content validity, and feasibility⁸. Comprehensibility of the SARD tool was evaluated using the Flesch Reading Ease and the Flesch-Kincaid Grade Level tools (Microsoft Office Professional Plus 2013). An acceptable Flesch Reading Ease score of 60.0–70.0 indicates that the material would be easily understood by 13- to 15-year-old students and a score of 90.0–100.0 indicates that the material would be easily understood by an average 11 year old⁹. Based on the sensibility assessment guide of Rowe and Oxman⁸, which has been used to evaluate other self-reported health instruments^{10,11}, comprehensibility, feasibility, face validity, and acceptability were evaluated using a numeric rating scale from 1–7. The anchors varied based on the question stem. The sensibility assessment questionnaire included space for subjects to provide specific comments related to the SARD tool questions and space for general feedback.

Validation. Concurrent validity evaluates the agreement between 2 measures of a construct administered at about the same time¹². Participants’ self-reported diagnoses using the SARD tool were compared to their diagnosis recorded by their physician(s) in their medical record. In SARD, there is no gold standard diagnostic test. A diagnosis is made through the physician’s judgment, using an aggregation of symptoms, signs, and test results that conform to the construct of the disease as learned in physicians’ specialized training and years of experience¹³. An investigator blinded to the self-reported diagnoses reviewed the medical records. Data collected included patient age, sex, and SARD diagnosis.

Statistical analyses. The data were summarized using descriptive statistics. For dichotomous screening questions, an answer of “no” was recorded as 0 while “yes” was recorded as 1. The response option of “I don’t know” was recorded as a 0. Based on a sample size assessment, 80 participants were required to achieve 80% (95% CI 0.71–0.88) power with a score of 5 or higher on a given item on the sensibility questionnaire^{14,15}. To determine whether the confirmatory questions conferred added value, the sensitivity, specificity, and 95% CI were calculated for the screening questions with and without the confirmatory questions separately. Concurrent validity of the diagnosis reported using the SARD tool compared with the medical record was evaluated using Cohen’s κ statistic. The κ statistic may be interpreted as less than 0 indicating no agreement, 0.00–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as an almost perfect agreement¹⁶. *A priori*, we hypothesized that moderate agreement ($\kappa > 0.41$) would be acceptable because additional factors may affect self-reported and physician-reported diagnoses. Statistical analysis was done using RStudio.

RESULTS

Participants. The SARD tool was administered to 141 participants whose characteristics are summarized in Table 1. Our response rate was 100% with all subjects approached agreeing to participate. The majority of the participants (87.2%) were women. The number of participants varied by

Table 1. Participant characteristics. Values are n (%) unless otherwise specified.

Characteristics	Sample Population, n = 141
Female	123 (87.2)
Age, yrs, mean ± SD	44 ± 14
Diagnosis	
SLE	30 (21.3)
RA	10 (7.1)
SSc	24 (17.0)
DM or PM	8 (5.7)
Sjögren syndrome	13 (9.2)
APS	4 (2.8)
Hemolytic anemia	3 (2.1)
Multiple sclerosis	4 (2.8)
Thyroid disease	26 (18.4)
Type 1 diabetes mellitus	6 (4.2)
Healthy controls	47 (33.3)

SLE: systemic lupus erythematosus; RA: rheumatoid arthritis; SSc: systemic sclerosis; DM: dermatomyositis; PM: polymyositis; APS: antiphospholipid antibody syndrome.

SARD with the highest number in the SLE group (n = 30). There were no patients with ITP. The healthy controls were 87.2% women. They were of similar age to the sample population with a mean ± SD age of 41.9 ± 15.6 years compared with 44 ± 14 years in the SARD group. Fourteen subjects had overlapping SARD conditions (SLE and SSc, n = 5; SLE and SS, n = 3; RA and SS, n = 4; SSc and SS, n = 2). A subset of 87 participants (including 18 healthy controls) were separately recruited to complete the sensibility assessment.

Sensibility assessment. The Flesch Reading Ease score for the SARD tool was 77.1 and the Flesch-Kincaid Grade Level was 4.4. As shown in Table 2, respondents endorsed comprehensibility (mean rating 6.12 ± 0.92), feasibility (mean rating 5.94 ± 0.81), face validity (mean rating 5.35 ± 1.10), and acceptability (mean rating 3.10 ± 2.03 where 1 = very unlikely).

Table 2. Sensibility assessment. Values are mean ± SD.

Sensibility Assessment Questions	Rating
How easy were the questions to understand? (7 = Excellent)	6.12 ± 0.92
Is the way in which the questions are presented confusing to you? (7 = Not at all)	6.55 ± 0.88
How would you rate the amount of time taken to complete this part of the questionnaire? (7 = Perfect)	5.94 ± 0.81
In your opinion, how likely is the questionnaire to detect the presence of autoimmune diseases? (7 = Very likely)	5.35 ± 1.10
Do you think that this part of the questionnaire would be acceptable to the general population and family members of patients with systemic autoimmune rheumatic diseases? (7 = Excellent)	5.85 ± 1.05
Do you think that others will be uncomfortable with the questionnaire? In other words, will they think that the information gathered is too probing or too personal for them? (1 = Very unlikely)	3.10 ± 2.03
Now that you have finished this part of the questionnaire, would you agree to complete this again? (7 = Very likely)	6.15 ± 1.01

Validity. Participants' self-reported diagnoses were compared with the diagnoses found in their medical record (Table 3). For the total sample population, the agreement between the SARD tool and medical record was $\kappa = 0.82$ (95% CI 0.77–0.88). Subgroup analysis by SARD found κ coefficients for SLE to be $\kappa = 0.88$ (95% CI 0.79–0.97), SSc $\kappa = 1.0$ (95% CI 1.0–1.0), PM/DM $\kappa = 0.72$ (95% CI 0.49–0.95), SS $\kappa = 0.85$ (95% CI 0.71–0.99), and RA $\kappa = 0.52$ (95% CI 0.32–0.72) improving to $\kappa = 0.61$ (95% CI 0.34–0.87) with the confirmatory question. All diseases in the questionnaire have been included in the table for completeness. Those with sample sizes of fewer than 10 patients have been indicated with an asterisk to aid in data interpretation.

The screening questions had high sensitivity ranging between 0.96 and 1.0, with the exception of the screening question for hemolytic anemia (sensitivity 0.33, n = 3; Table 3). The specificity of the screening questions was high, ranging between 0.88 and 1.0. Modest increases to near perfect specificity were achieved through the addition of the confirmatory question in the SLE, RA, PM/DM, SS, and hemolytic anemia groups. In the subset of patients with overlapping conditions, the SARD tool correctly identified those with SLE and SSc overlap (5/5 for SLE and 5/5 for SSc). Among those subjects with a SARD overlapping with SS, there were 4 subjects who did not have the term "Sjögren's disease" documented in the clinical chart. However, 3 of these subjects had documentation of sicca symptoms of dry eyes or dry mouth in their chart.

DISCUSSION

Case ascertainment tools are important instruments in clinical research. We have developed a SARD case ascertainment tool with demonstrable readability, comprehensibility, feasibility, and face and concurrent validity. This SARD tool is low cost and does not require specialized personnel, nomograms, or computation. It can be easily implemented in the outpatient setting to identify participants for research

Table 3. Validation.

Disease	Questions	Sensitivity (95% CI)	Specificity (95% CI)	κ (95% CI)
Sample population	—	0.91 (0.88–0.94)	0.99 (0.96–1.00)	0.82 (0.77–0.88)
SLE	Screening	1.00 (0.89–1.00)	0.95 (0.89–0.98)	0.88 (0.79–0.97)
	Confirmatory	0.73 (0.56–0.86)	0.98 (0.94–0.99)	0.77 (0.64–0.91)
RA	Screening	1.00 (0.72–1.00)	0.88 (0.82–0.93)	0.52 (0.32–0.72)
	Confirmatory	0.60 (0.31–0.83)	0.98 (0.93–0.99)	0.61 (0.34–0.87)
SSc	Screening	1.00 (0.86–1.00)	1.00 (0.97–1.00)	1.00 (1.00–1.00)
	Confirmatory 1	0.71 (0.51–0.85)	1.00 (0.97–1.00)	0.80 (0.66–0.94)
	Confirmatory 2	0.92 (0.74–0.98)	1.00 (0.97–1.00)	0.95 (0.88–1.00)
PM or DM*	Screening	1.00 (0.65–1.00)	0.96 (0.92–0.98)	0.72 (0.49–0.95)
	Confirmatory 1	1.00 (0.65–1.00)	0.96 (0.92–0.98)	0.72 (0.49–0.95)
	Confirmatory 2	0.71 (0.36–0.92)	0.97 (0.93–0.99)	0.60 (0.31–0.89)
Sjögren syndrome	Screening	1.00 (0.77–1.00)	0.97 (0.92–0.99)	0.85 (0.71–0.99)
	Confirmatory 1	0.92 (0.67–0.99)	0.98 (0.94–1.00)	0.88 (0.74–1.00)
	Confirmatory 2	0.61 (0.36–0.82)	0.99 (0.96–1.00)	0.71 (0.48–0.93)
APS*	Screening	1.00 (0.44–1.00)	1.00 (0.97–1.00)	1.00 (1.00–1.00)
	Confirmatory	1.00 (0.44–1.00)	1.00 (0.97–1.00)	1.00 (1.00–1.00)
Hemolytic anemia*	Screening	0.33 (0.06–0.79)	0.98 (0.95–0.99)	0.32 (–0.18 to 0.81)
	Confirmatory	0.33 (0.06–0.79)	1.00 (0.97–1.00)	0.50 (–0.11 to 1.00)
Multiple sclerosis*	Screening	1.00	0.99 (0.96–0.99)	0.89 (0.66–1.00)
	Confirmatory	1.00 (0.51–1.00)	0.99 (0.96–1.00)	0.89 (0.66–1.00)
Thyroid	Screening	0.96 (0.80–0.99)	0.98 (0.94–1.00)	0.93 (0.85–1.00)
	Confirmatory	0.88 (0.72–0.96)	0.98 (0.94–1.00)	0.88 (0.77–0.98)
Type 1 diabetes mellitus*	Screening	1.00 (0.61–1.00)	0.99 (0.96–1.00)	0.92 (0.76–1.00)
	Confirmatory	0.00 (0.00–0.39)	0.99 (0.96–1.00)	–0.01 (–0.03 to 0.01)

* Indicates diseases with $n < 10$. Screening: screening question; confirmatory 1: first confirmatory question; confirmatory 2: second confirmatory question. SLE: systemic lupus erythematosus; RA: rheumatoid arthritis; SSc: systemic sclerosis; DM: dermatomyositis; PM: polymyositis; APS: antiphospholipid antibody syndrome.

studies or by mail for family studies examining the development of SARD in the first-degree relatives of patients with this diagnosis. The use of both screening and confirmatory questions has resulted in good sensitivity and specificity.

The published accuracy of self-reported SARD diagnoses is highly variable depending on the population being studied and the method used to validate the self-report^{1,3,17,18}. Our SARD tool had demonstrable validity with “almost perfect agreement”¹⁶ with the medical record for our sample population. For each of the SARD (SLE, RA, SSc, PM/DM, SS, and APS), the κ values indicated moderate to almost perfect agreement¹⁶. Among the disease controls, the κ values were moderate to substantial for multiple sclerosis and thyroid disease. For SSc, the screening question alone was sufficient. The addition of confirmatory questions following a screening question is a method that has been used previously to enhance questionnaire validity¹. Indeed, in our study, the addition of the confirmatory question improved specificity. For RA and hemolytic anemia, the confirmatory questions inquired about the use of methotrexate and prednisone, respectively. These medications are frequently used in these conditions and are easily recognizable, thereby improving the specificity of the SARD tool for these conditions.

In comparison to the original version of this tool, the new version has improved visual formatting (addition of white space, blocking and shading of alternating questions, and

increased font size). These methods have been shown to improve the readability of tools, particularly in the classification of SARD⁷. This SARD tool retains the original screening questions. The addition of 1–2 confirmatory questions adds improved specificity. In some cases, the confirmatory questions result in reduced sensitivity. Because we report the operating characteristics of both the screening and confirmatory questions by SARD in the same sample, future investigators may make an informed choice regarding the use of only the screening questions or both screening and 1 or more confirmatory questions, depending on their needs and preferences.

A potential limitation to implementing medication use as confirmatory questions for a particular disease is that the response may be susceptible to recall bias or bias the sample to those individuals with more severe disease. Second, it should be noted that reliance on confirmatory questions might reduce sensitivity because the medication(s) may not be indicated exclusively for 1 disease. The question of the use of artificial tears for treatment of SS is limited by the fact that these are commonly used for people with dry eyes who do not have a diagnosis of SS^{19,20}. The use of symptoms as confirmatory questions is reasonable in diseases such as SS and SSc with classic and symptom-centered diagnoses^{21,22}. The high sensitivity of the screening questions makes it reasonable to use them to rule out a particular SARD before

asking confirmatory questions to increase specificity. Several of the diseases (SSc, PM/DM, and SS) had 2 confirmatory questions. Because patients may vary in their presentation as well as in their method of treatment, these particular questions should be interpreted separately. For example, a patient with SS may experience only 1 sicca symptom or may have dry eyes and mouth together²³. It should be noted that the recruiting of participants from specialty clinics may affect the performance of the questionnaire. Diagnostic test accuracy is dependent on the prevalence of disease, which may be the primary driver of setting-dependence. Application of this questionnaire to a more general population may alter the performance characteristics. Finally, our cases and controls were comparable with regards to age and sex. We did not collect information on education level, socioeconomic status, or health literacy. These could affect the findings.

Overall, the aspects of sensibility were scored highly by the participants, with scores over 5 in each of the areas assessed. It should be noted that the question pertaining to being uncomfortable with the SARD tool had the most variable response. This may be because the anchors for the response scale were reversed, as compared to the other questions in the questionnaire. Some participants may have completed the tool quickly and responded highly in all categories, not realizing that lower scores for this domain were more positive. Indeed, 1 subject who marked a 7 on this scale (indicating that people would likely find the questionnaire very uncomfortable) wrote “none” in the space provided to indicate which particular questions people would find uncomfortable. The branching questions may have been another source of confusion for some of the participants. Some participants answered branching questions (particularly when the answer was affirmative) after they had answered the screening question in the negative. For example, a participant may have indicated that they did not have SSc, but said that they did have Raynaud phenomenon. Modifying the design of the tool by use of arrows or clearer instructions may eliminate this confusion²⁴.

Moderate correlations were observed for RA, DM/PM, and SS. This was not surprising because subjects may have been told that they have these diagnoses as a means of explaining their symptoms. However, clinicians may consider inflammatory arthritis, inflammatory myositis, or sicca symptoms manifestations of the underlying SARD such as SSc or SLE. Indeed, it has been recognized that 15% of SSc subjects will have a symmetrical, inflammatory arthritis²⁵. As a point of comparison, the Connective Tissue Screening Questionnaire is a 30-item questionnaire with additional demographic, socioeconomic, and comorbidity questions²⁶. This questionnaire was validated among SARD cases and controls attending a tertiary, academic rheumatology center in the United States. The specificity of our SARD tool in SSc and PM/DM is improved because they reported specificities of 92% (95% CI 88–96) for SSc and 83% (95% CI 78–88)

for PM/DM. The point estimates for specificity of our tool are better for SLE and RA, but have overlapping CI.

We have developed a case ascertainment tool for the identification of individuals with a SARD for use in the outpatient setting. This tool has demonstrable sensibility (readability, comprehensibility, acceptability, face validity, feasibility). This tool has demonstrable validity compared with medical records. The use of both screening and confirmatory questions confers added accuracy, with a sensitivity of 91% and a specificity of 99%.

ONLINE SUPPLEMENT

Supplementary data for this article are available online at jrheum.org.

REFERENCES

1. Walitt BT, Constantinescu F, Katz JD, Weinstein A, Wang H, Hernandez RK, et al. Validation of self-report of rheumatoid arthritis and systemic lupus erythematosus: the Women's Health Initiative. *J Rheumatol* 2008;35:811-8.
2. Formica MK, McAlindon TE, Lash TL, Demissie S, Rosenberg L. Validity of self-reported rheumatoid arthritis in a large cohort: results from the Black Women's Health Study. *Arthritis Care Res* 2010;62:235-41.
3. Cooper GS, Wither J, McKenzie T, Claudio JO, Bernatsky S, Fortin PR; CaNIOS GenES Investigators. The prevalence and accuracy of self-reported history of 11 autoimmune diseases. *J Rheumatol* 2008;35:2001-4.
4. Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med* 2006;119:503.e1-9.
5. McAlindon TE, Formica M, Palmer JR, Lafyatis R, Rosenberg L. Assessment of strategies for identifying diagnosed cases of systemic lupus erythematosus through self-report. *Lupus* 2003;12:754-9.
6. Dillman DA, Smyth JD, Christian LM. *Internet, mail, and mixed-mode surveys: the tailored design method*. Hoboken: John Wiley & Sons Inc.; 2009:65-106.
7. Johnson SR, Naden RP, Fransen J, van den Hoogen F, Pope JE, Baron M, et al. Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. *J Clin Epidemiol* 2014;67:706-14.
8. Rowe BH, Oxman AD. An assessment of the sensibility of a quality-of-life instrument. *Am J Emerg Med* 1993;11:374-80.
9. Flesch R. A new readability yardstick. *J Appl Psychol* 1948; 32:221-33.
10. Ellerton C, Davis A, Brooks D. Preliminary development and validation of a paediatric cardiopulmonary physiotherapy discharge tool. *Physiother Can* 2011;63:34-44.
11. Alnaqbi KA, Touma Z, Passalent L, Johnson SR, Tomlinson GA, Carty A, et al. Development, sensibility, and reliability of the Toronto Axial Spondyloarthritis Questionnaire in inflammatory bowel disease. *J Rheumatol* 2013;40:1726-35.
12. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. Oxford: Oxford University Press; 2008.
13. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. *J Clin Epidemiol* 2010;63:355-69.
14. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998; 17:857-72.
15. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Amer Statist Assoc* 1927;22:209-12.

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
17. Rasooly I, Papageorgiou AC, Badley EM. Comparison of clinical and self reported diagnosis for rheumatology outpatients. *Ann Rheum Dis* 1995;54:850-2.
18. Karlson EW, Lee IM, Cook NR, Manson JE, Buring JE, Hennekens CH. Comparison of self-reported diagnosis of connective tissue disease with medical records in female health professionals: the Women's Health Cohort Study. *Am J Epidemiol* 1999;150:652-60.
19. Brewitt H, Sistani F. Dry eye disease: the scale of the problem. *Surv Ophthalmol* 2001;45 Suppl 2:S199-202.
20. Hay EM, Thomas E, Pal B, Hajeer A, Chambers H, Silman AJ. Weak association between subjective symptoms or and objective testing for dry eyes and dry mouth: results from a population based study. *Ann Rheum Dis* 1998;57:20-4.
21. Sankar V, Noll JL, Brennan MT. Diagnosis of Sjögren's syndrome: American-European and the American College of Rheumatology classification criteria. *Oral Maxillofac Surg Clin North Am* 2014;26:13-22.
22. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Ann Rheum Dis* 2013; 72:1747-55.
23. Shiboski SC, Shiboski CH, Criswell L, Baer A, Challacombe S, Lanfranchi H, et al; Sjögren's International Collaborative Clinical Alliance (SICCA) Research Groups. American College of Rheumatology classification criteria for Sjögren's syndrome: a data-driven, expert consensus approach in the Sjögren's International Collaborative Clinical Alliance cohort. *Arthritis Care Res* 2012;64:475-87.
24. Dillman DA, Smyth JD, Christian LM. *Internet, mail, and mixed-mode surveys: the tailored design method*. 3rd ed. Hoboken: John Wiley & Sons Inc.; 2009:151-233.
25. Muangchan C; Canadian Scleroderma Research G, Baron M, Pope J. The 15% rule in scleroderma: the frequency of severe organ complications in systemic sclerosis. A systematic review. *J Rheumatol* 2013;40:1545-56.
26. Karlson EW, Sanchez-Guerrero J, Wright EA, Lew RA, Daltroy LH, Katz JN, et al. A connective tissue disease screening questionnaire for population studies. *Ann Epidemiol* 1995;5:297-302.