A Review of the Sapporo and Revised Sapporo Criteria for the Classification of Antiphospholipid Syndrome. Where Do the Revised Sapporo Criteria Add Value?

RAJA S. BOBBA, SINDHU R. JOHNSON, and AILEEN M. DAVIS

ABSTRACT. Objective. The preliminary classification criteria for antiphospholipid syndrome (APS), or the Sapporo criteria, are widely used for the inclusion of patients with APS into clinical studies. Revised Sapporo criteria have been proposed as an improved criteria set. Whether these criteria sets fulfill the current standards of measurement science are unknown. The purpose of this study was (1) to evaluate the developmental methodology and measurement properties of the Sapporo and the revised Sapporo criteria for use in clinical trials; and (2) to evaluate if the revised Sapporo criteria provide added value over the Sapporo criteria.

Methods. A computer search for articles describing use of the Sapporo and the revised Sapporo criteria was performed. Item generation, item reduction, sensibility, validity, and reliability of the criteria were evaluated.

Results. The Sapporo criteria set has incremental face and content validity over its predecessors. However, through separation of anti-ß2-glycoprotein I antibodies as a sub-item, the specification of a wider time interval between serologic testing, the specification of a time interval between serology and clinical manifestations, and specification of definitions for clinical manifestations and laboratory titer thresholds, the revised Sapporo criteria set has incremental face and content validity over the Sapporo criteria. The complexity of the criteria, diagnostic tests, and immunologic tests limits their feasibility. The reliability of each criterion is unknown. The discriminative capacity of the Sapporo criteria is good, with sensitivity, specificity, and positive and negative predictive values of 0.71, 0.98, 0.95, and 0.88, respectively, compared to patients with systemic lupus erythematosus. The discriminative capacity of the revised Sapporo criteria is unknown.

Conclusion. The revised Sapporo criteria set has incremental face and content validity compared its predecessors. Reliability testing of each criterion is needed before these criteria can be confidently used in multicenter APS trials. Discriminatory testing of the revised Sapporo criteria is required.

(First Release June 1 2007; J Rheumatol 2007;34:1522–7)

Key Indexing Terms: ANTIPHOSPHOLIPID SYNDROME SAPPORO CRITERIA REVISED SAPPORO CRITERIA

The antiphospholipid antibody syndrome (APS) is an autoimmune condition with a variety of clinical, hematologic, and serologic manifestations. Due to the heterogeneity of presentations, investigators have proposed criteria to classify patients with APS for inclusion of homogeneous groups of patients in clinical studies. Previous iterations of criteria development have included the criteria for “antiphospholipid syndrome,” the criteria for primary APS and APS within systemic lupus erythematosus (SLE), and the criteria for the “antiphospholipid/cofactor syndromes”. However, these criteria have been criticized for being rudimentary; for not accounting for gestational duration at time of miscarriage; for inability to account for the lupus anticoagulant (LAC); or for inappropriately including livedo reticularis and pulmonary hypertension, while inappropriately excluding cardiac valve abnormalities. Thus, there remained a need for uniformity and international consensus within the field.
This led to the development of the preliminary classification criteria for definite APS, or the Sapporo criteria, at a post-conference workshop in Sapporo, Japan, following the Eighth International Symposium on Antiphospholipid Antibodies. Using the Sapporo criteria, definite APS is present if one of 2 clinical criteria and one of 2 laboratory criteria are met. Clinical criteria include the presence of vascular thrombosis (arterial, venous, or small-vessel thrombosis confirmed by diagnostic imaging or histopathology, but excluding superficial venous thrombosis) or pregnancy morbidity (fetal loss at various stages of gestational duration). Laboratory criteria include the presence of anticardiolipin antibodies (aCL: IgG or IgM isotype, present in medium or high titer, on 2 or more occasions at least 6 weeks apart) or the presence of the LAC (present in plasma, on 2 or more occasions separated by 6 weeks). Since their development in 1999, the Sapporo criteria have been widely used in clinical research. However, over the same period, new clinical and laboratory insights have been gained. In 2006, a revised version of the Sapporo criteria was developed at a consensus workshop in Sydney, Australia, before the Eleventh International Congress on Antiphospholipid Antibodies. Although the requirement of one clinical and one pregnancy morbidity remains unchanged, modifications to the pregnancy morbidity and laboratory criteria have been made.

Due to the importance of criteria in research, a committee on Classification and Response Criteria, a subcommittee of the American College of Rheumatology (ACR) Quality Measures Committee, has been charged with the responsibility of encouraging development and validation of new and improved classification criteria for various rheumatic diseases. Recommendations for development and validation of criteria sets have been developed based on the current standards of measurement. Although widely accepted, the measurement properties of the Sapporo criteria have not been comparatively related to the recommended standards of measurement, thus it is difficult for investigators to ascertain if these criteria are appropriate for classification of patients for clinical trials. Further, there is uncertainty whether the revised Sapporo criteria confer an advantage over the Sapporo criteria.

Thus, our objectives were (1) to evaluate the developmental methodology (item generation, item reduction) and measurement properties (sensibility, reliability, and validity) of the Sapporo and revised Sapporo criteria to classify patients with APS for clinical trials; and (2) to evaluate whether the revised Sapporo criteria provide added value over the Sapporo criteria.

MATERIALS AND METHODS

Search strategy. Studies were identified using Medline (1966 to week 4 May 2006), Embase (1980 to 2006 week 22), PsychInfo (1985 to week 5 May 2006), and Cumulative Index to Nursing, Allied Health Literature (1982 to week 4 May 2006) databases. Subject headings (antiphospholipid antibody syndrome OR Hughes syndrome) AND (Sapporo criteria OR Sydney criteria OR classification OR criteria) were used. The search was limited to human and English language studies. The bibliographies of included studies and published reviews were also searched. Titles and abstracts were screened to identify articles that discussed the development, validity, or reliability of the criteria. Studies were excluded if they (1) did not evaluate the APS, (2) studied animal models, (3) were non-English language, (4) were case reports or case series, (5) evaluated exclusively pedi- atric patients (<16 years of age), or (6) evaluated criteria exclusively for classification of catastrophic APS.

Developmental methodology and measurement properties. Item generation, item reduction, administration, coding, scoring, sensibility, reliability, and validity were attributes used to evaluate the Sapporo and revised Sapporo criteria. “Item generation” refers to the process by which potential items in the criteria set are collected. “Item reduction” refers to the process by which redundant and less relevant items are eliminated. “Administration coding and scoring” reflects the process of how the criteria set is implemented and how components of the criteria set are incorporated to generate a quantitative representation (i.e., a score) of a specific construct.

“Sensibility” is used to describe whether a measurement instrument is useful and makes sense. Important attributes of sensibility include the purpose for which the measurement tool will be used, the population and the setting in which the instrument will be applied, content validity, face validity, and feasibility. “Content validity” reflects the ability of the measurement tool to incorporate relevant items and whether they accurately reflect the conceptual framework of the tool. “Face validity” evaluates biologic coherence of the domains and whether at face value the measurement tool actually makes sense in connecting the items. “Feasibility” refers to the ease of use of the instrument in terms of time of completion and ease of scoring.

“Reliability” describes the consistency of the criteria across repeated use by a single rater at 2 different points of time (intrarater) or 2 different raters at the same point (intrarater) in time. “Criterion validity” is a reflection of how the criteria compare to a gold standard. “Concurrent criterion validity” compares the criteria to a gold standard that is administered concurrently. The “discriminatory capacity” describes the ability of the criteria to differentiate the presence or absence of disease. Sensitivity and specificity are indicators of discriminative capacity.

RESULTS

Literature search. The search strategy resulted in the identification of 405 citations. Evaluation of titles and abstracts identified 36 articles that discussed the development, validity, or reliability of the criteria. The majority of articles (25 of 36) discussed the historical development of APS classification criteria. The remaining 11 articles included reliability and validity data for varying aspects of the criteria, which are summarized below. None of the articles discussed the same criteria, so a “head to head” comparison was not possible.

Purpose, population, and setting. Both the Sapporo and revised Sapporo criteria sets were developed to classify groups of patients with APS for inclusion in clinical studies and were intended to be applied in the outpatient setting. The criteria were not designed for diagnostic purposes.

Item generation and reduction. Item generation and reduction methods for the Sapporo criteria included expert judgment, review of the literature, and discussion of case scenarios, followed by open debate, voting, and revision of the criteria until consensus was achieved. Methods for item generation and reduction for the revised Sapporo criteria
included review of the literature, grading of evidence, and discussion among experts until consensus was achieved.

**Content validity.** The Sapporo criteria incorporate the relevant domains of thrombosis, pregnancy loss, and antiphospholipid antibodies. The Sapporo criteria are an improvement over their predecessors by including the aCL and the LAC. The criteria specify the immunoglobulin isotype required, the need for an assay to measure anti-ß2-glycoprotein I antibodies, the tests required to demonstrate the presence of a LAC, and the need for retesting of aCL at least 6 weeks apart (reducing the false-positive rate due to transient aCL resulting from concurrent infection but which are not associated with thromboembolic events19). The pregnancy morbidity criteria incorporate gestational duration, which was not present in previous iterations of criteria.

The revised Sapporo criteria have improved content validity. Definitions for eclampsia, severe preeclampsia, placental insufficiency, and titers for abnormal aCL isotypes are specified. A requirement that the time between a positive antiphospholipid antibody test and the clinical manifestation is less than 5 years has been added. The criteria also separate the presence of an anti-ß2-glycoprotein I antibody defined by specific isotypes and titers as a unique sub-item. Together, these changes reduce the potential for misclassification error.

However, other potentially important domains that have been excluded are hematologic and immunologic factors such as thrombocytopenia, hemolytic anemia, antinuclear antibody, and extractable nuclear antigen3,6. Potentially important antibodies, such as anti-prothrombin20 and antiantigenic V6 antibodies, which are known to be present in the APS, are not present in the Sapporo criteria or the revised Sapporo criteria.

**Face validity.** The Sapporo criteria are complex in their implementation21,22. Criteria contain multiple components. For example, pregnancy morbidity criterion contains 3 options: the number of fetal deaths or premature births, the gestational cutoff, and anatomic specifications. Another threat to face validity is that the relationship between thrombosis, pregnancy loss, and antiphospholipid antibodies is not explicit at first glance. It does not clearly “look like” it measures what it intends to measure. Knowledge of the underlying pathophysiology is required to understand the relationship between domains. Thus face validity is limited to specialists familiar with the syndrome and its pathophysiology.

The revised Sapporo criteria have improved face validity over the Sapporo criteria with respect to time requirements. The time requirement between 2 positive LAC tests was revised from 6 weeks to 12 weeks, resulting in a theoretical reduction in false-positive tests. How this change in time requirement affects the false-negative rate is uncertain. Similarly, the revised Sapporo criteria stipulate that APS classification should be avoided if a period of more than 5 years separates the positive antiphospholipid antibody test and clinical manifestations. Again, there is a theoretical reduction in misclassification by eliminating an implied causal relationship between the presence of antiphospholipid antibodies and clinical manifestations. Whether the introduction of these time specifications truly reduces misclassification error requires testing.

**Feasibility.** Feasibility of both the Sapporo and revised Sapporo criteria is limited to specialists with access to diagnostic imaging, histopathology, genetic testing, and a laboratory that performs hematological testing, and the ability to interpret the test results. Feasibility of both criteria sets is limited if applied in the community setting.

**Administration, coding, and scoring.** The requirement that only one clinical and one laboratory feature are needed to classify a patient with APS suggests that these criteria are simple to use. However, since each criterion has multiple components, errors in coding may occur. When scoring, each criterion is equally weighted and is denoted as present or absent. Further research is needed to evaluate if this weighting produces scores that most accurately classify APS.

The revised Sapporo criteria are similar to the Sapporo criteria in regard to administration, coding, and scoring. As a result, they too suffer from complex, multi-item criteria. Uncertainty remains whether equal weighting of items is appropriate.

**Reliability.** No formal assessments of reliability of the Sapporo or revised Sapporo criteria have been performed. The greatest threat to the reliability of these criteria lies in the laboratory testing of antibodies. The guidelines by the Scientific Subcommittee for Lupus Anticoagulants/Phospholipids Dependent Antibodies indicate that lupus anticoagulant screening must use more than one test system23. The dilute Russell viper venom test is the most commonly used venom-based assay and is more sensitive than the dilute prothrombin time. Textarin time and Taipan Venom Time are other assays; however, Textarin is not sensitive to abnormalities involving factor V and the Taipan Venom Time is not sensitive to abnormalities involving factors V or X19. Whether one test or specific combination of LAC screening tests has incremental reliability is uncertain.

There is a significant lack of concordance in aCL testing between laboratories19, and marginal to moderate agreement between assays (Cohen’s kappa 0.2 to 0.66)24. The lack of concordance may be related to the choice of microtiter plates for ELISA assays, buffering reagents, technical issues within the laboratory, and variable plate readers19.

A cross-sectional multicenter study also demonstrated considerable interlaboratory variability of anti-ß2-glycoprotein I antibody measurement25. Only 11/190 (6%) of centers achieved an excellent concordance (kappa 0.8–1.0), and 37/190 (13%) of centers achieved good concordance (kappa
Incremental validity refers to the degree that one measure precedes is its incremental face and content validity. Features associated with APS including hemolytic anemia, for SLE include the presence of antiphospholipid antibodies, and to guide therapy. Indeed, the differentiation of classification criteria of the Sapporo criteria have not been reported. However, the revised Sapporo criteria have improved specificity when applied to obstetric medicine. In a retrospective study of 107 women with antiphospholipid antibodies, 16% of patients were classified as definite APS based on the pregnancy morbidity criteria of the Sapporo criteria. When the revised Sapporo criteria were applied, 21% (6 additional patients) were classified as having definite APS.

**DISCUSSION**

The Sapporo criteria for the classification of individuals with APS are widely accepted; however, the revised Sapporo criteria set is believed to be an improvement on its predecessor, as it incorporates clinical and laboratory insights gained in the past 15 years. The results of our review/critique suggest that there are several issues with the Sapporo and the revised Sapporo criteria that would benefit from additional research to ensure that misclassification is minimized. This is critical in evaluating interventions in clinical trials as inclusion of misclassified participants will increase the type II error.

The first area of confusion is the intent of the criteria. Although the explicit purpose of criteria is classification of individuals for research purposes, many investigators and clinicians have used these criteria for diagnostic purposes and to guide therapy. Indeed, the differentiation of classification criteria and diagnostic criteria has been unclear. Adding to the confusion is the characterization of patients for whom the criteria were intended. Primary APS implies that the patient does not have features of SLE. However, the current ACR classification criteria for SLE include many features associated with APS including hemolytic anemia, false-positive serologic test for syphilis, and thrombocytopenia. Indeed, the 1997 updated classification criteria for SLE include the presence of antiphospholipid antibodies. Piette, et al have proposed exclusion criteria for primary APS in an attempt to make the delineation more transparent.

The major advantage of the Sapporo criteria set over its predecessors is its incremental face and content validity. Incremental validity refers to the degree that one measure makes a contribution over a simpler one, and includes the dimensions of incremental content validity, incremental predictive validity, and ecologic validity or generalizability across settings. The Sapporo criteria were the first set of classification criteria to be developed by international consensus. However, since its development, additional knowledge has been gained in relation to clinical manifestations associated with APS (e.g., cognitive dysfunction, small high-density brain lesions on magnetic resonance imaging, nephropathy, and the pathophysiologic mechanism of disease). Critics of the Sapporo criteria believe that they suffer from inappropriate inclusions and exclusions. The inclusion of IgM aCL remains controversial. Although IgG aCL was strongly associated with recurrent fetal loss, its association with IgM and IgA were not statistically significant. Conversely, some investigators believe that important clinical and hematological features have been omitted.

The revised Sapporo criteria have incremental face and content validity over the Sapporo criteria through (1) the separation of anti-β2-glycoprotein I antibodies as a unique sub-item, (2) the specification of a wider time interval between serologic tests, (3) the specification of a time interval between serology and clinical manifestations, and (4) specification of definitions for clinical manifestations and laboratory titer thresholds. Although the specification of time intervals confers improved face validity, testing of these time intervals is needed. The inclusion of IgM aCL in the revised Sapporo criteria remains controversial.

The revised Sapporo criteria identify a subset of patients with APS, but exclude other clinical subsets of patients. Criteria currently exist for patients with catastrophic APS. Patients with antiphospholipid antibodies and non-criteria clinical manifestations are excluded using the revised Sapporo criteria. However, it has been proposed that patient subsets be classified as “antiphospholipid antibody associated disease.” Investigators should continue to define and evaluate non-criteria features of APS to determine their association with APS and clinical and prognostic significance. One innovative proposal to deal with the large spectrum of features is the development of a comprehensive scoring system that encompasses multiple clinical and immunological items.

Our ability to evaluate the measurement properties of the Sapporo and revised Sapporo criteria was limited by their design. Each set of classification criteria does not comprise a continuous scale, but rather a categorical one with components and subcomponents. Thus the reliability and validity of each criterion needs to be evaluated before one can determine if these criteria comply with the current standards of measurement. Reliability testing of each criterion is a critical requirement if these criteria are to be used to classify patients for multicenter studies.

The ACR Subcommittee on Classification and Response Criteria has made recommendations for developing and validating classification criteria. Based on these recommendations, the criterion validity (sensitivity, specificity, predictive value) of the revised Sapporo criteria will need to
be tested before they can be used confidently in clinical studies. Criterion validity testing involves evaluation of the level of agreement between the new criteria and a cohort of patients that expert clinicians agree have the disease. Ideally this cohort represents the full spectrum of disease. A common pitfall is the use of circular reasoning where the same experts and patients are used to derive and validate the criteria. Thus, experts and patients who were not involved in the development of these criteria should become involved in the validation studies.

We conclude the revised Sapporo criteria have incremental face and content validity over previous criteria, and these criteria have incremental value over the Sapporo criteria. Critical steps are needed in ensuring misclassification does not occur, including reliability testing of the individual criteria and head-to-head comparison of the Sapporo and revised Sapporo criteria in accuracy of classification.

ACKNOWLEDGMENT
The authors thank Dr. Paul Fortin, Dr. Carl Laskin, and Dr. Michael Lockshin for their insightful comments on the historical development and use of the Sapporo and revised Sapporo criteria.

REFERENCES


