ARTHRITIS & RHEUMATISM
Vol. 44, No. 4, April 2001, pp 912–920

Published by Wiley-Liss, Inc.

# A Disease-Specific Activity Index for Wegener's Granulomatosis

## Modification of the Birmingham Vasculitis Activity Score

John H. Stone,[1] Gary S. Hoffman,[2] Peter A. Merkel,[3] Yuan-I Min,[1] Misty L. Uhlfelder,[1] David B. Hellmann,[1] Ulrich Specks,[4] Nancy B. Allen,[5] John C. Davis,[6] Robert F. Spiera,[7] Leonard H. Calabrese,[2] Fredrick M. Wigley,[1] Nicola Maiden,[8] Robert M. Valente,[4] John L. Niles,[9] Kenneth H. Fye,[6] Joseph W. McCune,[10] E. William St.Clair,[5] and Raashid A. Luqmani,[8]
for the International Network for the Study of the Systemic Vasculitides (INSSYS)

***Objective.*** **To refine and validate the Birmingham Vasculitis Activity Score (BVAS) as a disease-specific activity index for Wegener's granulomatosis (WG).**

***Methods.*** **Sixteen members of the International Network for the Study of the Systemic Vasculitides (INSSYS) revised the BVAS, with 3 goals: to reduce the redundancy of some component items, to enhance its ability to capture important disease manifestations specific to WG, and to streamline the instrument for use in clinical research. We defined the items and weighted them empirically as either minor (e.g., nasal crusting = 1 point) or major (e.g., alveolar hemorrhage = 3 points). We then validated the new, disease-specific BVAS/WG in 2 simulation exercises and a clinical case series that involved 117 patients with WG.**

***Results.*** **We removed 38 items from the original BVAS, revised 9 items, and added 7 new items. Correlations between the scores on the BVAS/WG and the physician's global assessment (PGA) of disease activity were high, even when patients in remission were excluded. In the clinical case series, Spearman's rank correlation coefficient between the BVAS/WG and the PGA was r = 0.81 (95% confidence interval 0.73–0.87). The interobserver reliability using intraclass (within-case) correlation coefficients in the 2 simulation exercises was r = 0.93 for the BVAS/WG and r = 0.88 for the PGA in the first and r = 0.91 for the BVAS/WG and r = 0.88 for the PGA in the second. There was no significant observer effect in the scoring of the BVAS/WG or the PGA. The discriminant validity of the BVAS/WG was good: r = 0.73 (95% confidence interval 0.43–0.83).**

***Conclusion.*** **The BVAS/WG is a valid, disease-specific activity index for WG. Tested in simulation exercises and in actual patients, the BVAS/WG correlates well with the PGA, is sensitive to change, and has good inter- and intraobserver reliability. The INSSYS will use the BVAS/WG to assess the primary outcome in a phase II/III trial of etanercept in WG.**

The investigation of new therapies for the treatment of systemic vasculitis and the organization of

Supported by NIH grant N01-AR-9-2240 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases, by the Food and Drug Administration (Office of Orphan Products Development), and by Immunex Corporation, Seattle, Washington.

[1]John H. Stone, MD, MPH, Yuan-I Min, PhD, Misty L. Uhlfelder, BA, David B. Hellmann, MD, Fredrick M. Wigley, MD: Johns Hopkins University, Baltimore, Maryland; [2]Gary S. Hoffman, MD, Leonard H. Calabrese, DO: The Cleveland Clinic Foundation, Cleveland, Ohio; [3]Peter A. Merkel, MD, MPH: Boston University, Boston, Massachusetts; [4]Ulrich Specks, MD, Robert M. Valente, MD: Mayo Clinic, Rochester, Minnesota; [5]Nancy B. Allen, MD, E. William St.Clair, MD: Duke University, Durham, North Carolina; [6]John C. Davis, MD, MPH, MS, Kenneth H. Fye, MD: University of California, San Francisco; [7]Robert F. Spiera, MD: Beth Israel Hospital, New York, New York; [8]Nicola Maiden, MB, BS, MRCP (UK), Raashid A. Luqmani, B Med Sci BM BS, MRCP (UK), DM FRCP (E): University of Edinburgh, Edinburgh, Scotland; [9]John L. Niles, MD: Harvard University, Boston, Massachusetts; [10]Joseph W. McCune, MD: University of Michigan, Ann Arbor.

Address correspondence and reprint requests to John H. Stone, MD, MPH, The Johns Hopkins Vasculitis Center, 1830 East Monument Street, Suite 7500, Baltimore, MD 21205.

Submitted for publication March 30, 2000; accepted in revised form November 20, 2000.

multicenter clinical trials in these diseases demand rigorous methods for serial assessments of disease. Compared with other rheumatic diseases, e.g., systemic lupus erythematosus (1–6), there have been relatively few efforts to develop and validate disease activity indices for the vasculitides (7,8). This is principally because the term "vasculitis" includes up to 20 clinically distinct forms of vascular inflammation, many of which are rare diseases. Moreover, the complex, multisystem involvement of systemic vasculitis makes the concise assessment of disease activity challenging. No validated activity indices for individual forms of vasculitis exist.

The Birmingham Vasculitis Activity Score (BVAS) (7) was conceived for use in multiple forms of vasculitis and has been modified and used in clinical trials (9). The original BVAS has certain limitations, however, because of some redundancy of the items. The Vasculitis Activity Index (8), though subjected to a thorough validation study, is also intended for use in multiple forms of systemic vasculitis. Recently, attention has centered around the development of activity indices that are specific for individual forms of vasculitis. The Groningen Index (10) and the Disease Extent Index (11) were created for use in Wegener's granulomatosis (WG) but have not been subjected to validation studies.

WG, a major form of systemic vasculitis (12), is the prototype of disorders associated with antineutrophil cytoplasmic antibodies (13). Several features of WG, including the upper respiratory tract manifestations (often destructive), distinguish it from other forms of vasculitis and make the development of a WG-specific disease activity index important.

In preparation for the conduct of a clinical trial in WG, we revised the BVAS with 3 major goals: 1) reduction of the redundancy in some of the component items, 2) enhancement of the instrument's ability to capture disease manifestations relevant to WG, and 3) streamlining of the instrument for clinical use in multicenter studies. We report herein our efforts to revise the original BVAS, to construct a WG-specific disease activity index (BVAS/WG), and to validate the use of this instrument in patients with this disorder.

## PATIENTS AND METHODS

**Construction of the BVAS/WG.** *Overview.* The first draft of the BVAS/WG was developed at a meeting of 16 vasculitis investigators in Baltimore, Maryland on January 16–17, 1999. We composed an instruction manual for using the BVAS/WG, which includes a glossary of terms. Following the initial meeting, we tested and further refined the BVAS/WG in 2 simulation exercises (using "paper cases"; see below), and then employed the BVAS/WG in a clinical case series of 117 patients evaluated at 9 centers, a subset of whom had multiple visits. Data from these validation exercises were presented to the investigators at a followup meeting in Baltimore 1 year later (January 14, 2000), at which time data on test–retest reliability were obtained.

The BVAS/WG evaluation form (Figure 1), BVAS/WG glossary, and the simulation exercises are accessible on the Internet at http://vasculitis.med.jhu.edu.

*Consensus group.* The panel of investigators comprised 16 physicians whose clinical and basic research interests pertain to the vasculitides. All were members of the International Network for the Study of the Systemic Vasculitides (INSSYS). The group included physicians trained in several medical subspecialties: rheumatology, pulmonology, and nephrology.

*Item selection.* We retained the original BVAS arrangement of data items in 9 groups, which, with 1 exception ("general"), correspond to organ systems: 1) general, 2) cutaneous, 3) mucous membranes/eyes, 4) ear, nose, and throat, 5) cardiovascular, 6) gastrointestinal, 7) pulmonary, 8) renal, and 9) nervous system. We added an "other" section to permit the documentation of clinical manifestations of WG not covered by items listed in the 9 core groups. We excluded from the BVAS/WG items that occur often in other forms of vasculitis but are not generally associated with WG (e.g., bruits or loss of pulses). When appropriate, we merged or eliminated items deemed redundant. For example, bloody nasal discharge and nasal crusting were combined into a single item because we believed the assignment of separate points for these closely related manifestations of nasal inflammation was unnecessary. Similarly, 5 items related to changes in serum creatinine or proteinuria were replaced with one item (rise in creatinine $>30\%$ *or* fall in creatinine clearance $>25\%$), and we added red blood cell casts as a new item. Whenever possible, we replaced the symptoms and signs included as part of the original BVAS (e.g., hoarseness/stridor) with their known anatomic correlates in WG (i.e., subglottic involvement).

*Evaluation form.* The BVAS/WG evaluation form (Figure 1) consists of a 1-page form that is easy to read and use. A box containing instructions for completing the form is displayed at the top of the evaluation form. The form also includes: 1) 34 separate disease items, categorized into 9 groups; 2) an "other" section; 3) an asterisk by the 15 major items (see below); 4) tick boxes to indicate new/worse or persistent disease; 5) an area to total the scores; 6) a section for the designation of disease status; 7) the physician's global assessment (PGA) of disease activity scale; and 8) a box for administrative use that contains information about the patient identification code and clinical center. Items on the BVAS/WG evaluation form are counted only if they result from active WG, and not from damage from previously active WG or another medical condition.

*Item classification as major versus minor.* By consensus of the panel, "major" disease manifestations were those that constitute an immediate threat to the patient's life or to the function of a vital organ. Major items such as urinary red blood cell casts, pulmonary hemorrhage, and mononeuritis multiplex that constitute immediate threats to vital organs or the patient's life are indicated on the evaluation form by an asterisk and boldface type. The occurrence of such manifestations currently indicates the need for combination therapy with

## BVAS for Wegener's Granulomatosis Evaluation Form

Tick box (□ or ○) only if abnormality is ascribable to the presence of active Wegener's Granulomatosis (chronic damage should be scored separately in the Vasculitis Damage Index, VDI.)

□ Tick box only if the abnormality is **persistent disease activity** since the last assessment and not worse within the **previous 28 days.**

○ Tick box only if the abnormality is **newly present or worse** within the **previous 28 days.**

△ If no items are present in any section, tick "none".

**Major items are in bold and marked with ***

**All WG-related clinical features need to be documented on this form if they are related to active disease. Use "OTHER" category as needed.**

**1.** Clinic ID: __ __ __ __

**2.** Patient ID: __ __ __ __

**3.** Patient name code: __ __ __ __ __ __

**4.** Date form completed: __ __ - __ __ __ - __ __ (day / month / year)

**5.** Visit ID: __ __ __

| | Persistent | New/Worse | None |
|---|---|---|---|
| **6. GENERAL** | | | △1 |
| a. arthralgia/arthritis | □1 | ○2 | |
| b. fever (≥38.0 °C) | □1 | ○2 | |
| **7. CUTANEOUS** | | | △1 |
| a. purpura | □1 | ○2 | |
| b. skin ulcer | □1 | ○2 | |
| c. * **gangrene** | □1 | ○2 | |
| **8. MUCOUS MEMBRANES/EYES** | | | △1 |
| a. mouth ulcers | □1 | ○2 | |
| b. conjunctivitis/episcleritis | □1 | ○2 | |
| c. retro-orbital mass/proptosis | □1 | ○2 | |
| d. uveitis | □1 | ○2 | |
| e. * **scleritis** | □1 | ○2 | |
| f. * **retinal exudates/hemorrhage** | □1 | ○2 | |
| **9. EAR, NOSE & THROAT** | | | △1 |
| a. bloody nasal discharge/nasal crusting/ulcer | □1 | ○2 | |
| b. sinus involvement | □1 | ○2 | |
| c. swollen salivary gland | □1 | ○2 | |
| d. subglottic inflammation | □1 | ○2 | |
| e. conductive deafness | □1 | ○2 | |
| f. * **sensorineural deafness** | □1 | ○2 | |
| **10. CARDIOVASCULAR** | | | △1 |
| a. pericarditis | □1 | ○2 | |
| **11. GASTROINTESTINAL** | | | △1 |
| a. * **mesenteric ischemia** | □1 | ○2 | |
| **12. PULMONARY** | | | △1 |
| a. pleurisy | □1 | ○2 | |
| b. nodules or cavities | □1 | ○2 | |
| c. other infiltrate secondary to WG | □1 | ○2 | |
| d. endobronchial involvement | □1 | ○2 | |
| e. * **alveolar hemorrhage** | □1 | ○2 | |
| f. * **respiratory failure** | □1 | ○2 | |

| | Persistent | New/Worse | None |
|---|---|---|---|
| **13. RENAL** | | | △1 |
| a. hematuria (no RBC casts) (≥1+ or ≥10 RBC/hpf) | □1 | ○2 | |
| b. * **RBC casts** | □1 | ○2 | |
| c. * **rise in creatinine >30% *or* fall in creatinine clearance >25%** | | ○2 | |

**Note:** *If both hematuria and RBC casts are present, score only the RBC casts (the major item).*

| | | | |
|---|---|---|---|
| **14. NERVOUS SYSTEM** | | | △1 |
| a. * **meningitis** | □1 | ○2 | |
| b. * **cord lesion** | | ○2 | |
| c. * **stroke** | | ○2 | |
| d. * **cranial nerve palsy** | | ○2 | |
| e. * **sensory peripheral neuropathy** | | ○2 | |
| f. * **motor mononeuritis multiplex** | | ○2 | |
| **15. OTHER** (describe all items and * items deemed major) | | | △1 |
| _____ | □1 | ○2 | |
| _____ | □1 | ○2 | |
| _____ | □1 | ○2 | |
| _____ | □1 | ○2 | |

**16. TOTAL NUMBER OF ITEMS:**

| a. | b. | c. | d. |
|---|---|---|---|
| ____ | ____ | ____ | ____ |
| Major New/Worse | Minor New/Worse | Major Persistent | Minor Persistent |

**DETERMINING DISEASE STATUS:**

**Severe Disease/Flare:** ≥1 new/worse Major item.

**Limited Disease/Flare:** ≥1 new/worse Minor item.

**Persistent Disease:** Continued (but not new/worse) activity.

**Remission:** No active disease, including either new/worse or persistent items.

**17. CURRENT DISEASE STATUS** *(check only one):*

Severe Disease/Flare ( 1)

Limited Disease/Flare ( 2)

Persistent Disease ( 3)

Remission ( 4)

**18. PHYSICIAN'S GLOBAL ASSESSMENT (PGA)**

Mark line to indicate the amount of WG disease activity (not including longstanding damage) within the **previous 28 days:**

**Remission** (0) ——— (10) **Maximum activity**

**19. Value in item #18:** __ __ __ mm *(distance from 0 to tick mark in millimeters)*

**20. DATE FORM REVIEWED:** __ __ - __ __ __ - __ __ (day / month / year)

**21. STUDY PHYSICIAN ID:** __ __

**22. STUDY PHYSICIAN SIGNATURE:** _____

**23. CLINIC COORDINATOR ID:** __ __

**24. CLINIC COORDINATOR SIGNATURE:** _____

**Figure 1.** Evaluation form for the Birmingham Vasculitis Activity Score for Wegener's Granulomatosis (BVAS/WG). RBC = red blood cell; hpf = high-power field.

cyclophosphamide and glucocorticoids. In contrast, although "minor" items are important manifestations of active WG, they do not constitute immediate threats to vital organs or patients' lives and would normally be managed as limited WG, with methotrexate and glucocorticoids (14–16). Examples of minor items are swollen salivary gland, nasal crusting, and purpura. This dichotomous classification of disease severity is not meant to be applied rigidly in clinical practice but, rather, to be used as a standardization tool in research studies.

*Weighting.* We empirically chose to weight each major "new/worse" item as 3 points and each minor item as 1 point. We evaluated the appropriateness of these weights in the validation studies.

*Activity, damage, and comorbidity.* A cardinal principle of the BVAS/WG is that any clinical item scored on the evaluation form must reflect *active* WG, as opposed to WG-related damage or an intercurrent medical problem. This occasionally necessitates delays in completing the evaluation form until the etiology of a particular clinical finding can be defined. For example, a rise in serum creatinine may have several potential etiologies (drug toxicity, volume depletion, infection, recurrent glomerulonephritis secondary to WG, and others). Until the etiology of the creatinine elevation is apparent, the renal organ system is not scored. This requirement reflects the challenges that confront clinicians in the care of patients with WG, and it enhances the accuracy of the BVAS/WG as a measure of disease activity.

*New/worse versus persistent disease.* The instruction box at the top of the evaluation form instructs investigators to indicate whether clinical items that are present are new/worse since the previous BVAS/WG evaluation or whether the items represent persistent WG activity. All items present are scored as either new/worse or persistent (but not both). Distinction between new/worse and persistent disease features is critical to the accurate determination of WG flares, as opposed to poorly controlled, ongoing WG activity.

*Scoring.* In calculating the final BVAS/WG score, the number of major items (either new/worse or persistent) is multiplied by 3 and added to the total number of minor items. The maximum BVAS/WG score, therefore, is 68, assuming that not more than 1 major and 1 minor "other" items are present. The BVAS/WG score ranged from 0 to 13, with a median score of 2, among the 117 clinic patients evaluated at baseline in this study (Figure 2).

*Determining disease status.* To establish unambiguous definitions of disease status for use in both clinical trials and clinical practice, the BVAS/WG includes categorical ratings that incorporate major and minor items into the definitions of disease status. These 4 disease statuses are as follows: severe disease/flare (occurrence of any new/worse item that is major), limited disease/flare (occurrence of any new/worse item that is minor), persistent disease (presence of $\geq$1 item representing active disease that has continued since the patient's previous evaluation), and remission (no active disease; that is, no new/worse and no persistent items present).

*Physician's global assessment.* A new feature of the BVAS/WG is the inclusion of a PGA of WG activity. Evaluators are instructed to indicate the degree of WG disease activity within the 28 days prior to the evaluation by marking a vertical line on the 10-cm visual analog scale.

**Validation studies of the BVAS/WG.** *Simulation exercises.* We designed 2 simulation exercises using "paper cases" (10 in each exercise, for a total of 20 cases) to provide investigators with experience before they used the instrument for real patients. Most of the cases used in the simulation exercises were based on actual patients from the investigators' clinical practices. We selected these cases to represent the broad range of clinical manifestations in WG and to test the extent to which the BVAS/WG reflects clinical practice. The range of disease included cases with fulminant, life-threatening disease; other types of severe disease flares; limited WG; chronic, "grumbling" WG; and complete remission. Some cases included unusual manifestations of WG (e.g., genitourinary disease) to emphasize the importance of the "other" items group in capturing all manifestations of active disease.

Prior to the completion of each simulation exercise, each investigator scored 3 "practice" cases. We provided a guide to the scoring of these 3 cases, detailing the rationale for the scoring in each. Only after completing the training cases did investigators read and score the cases in the simulation exercises.

*Intraobserver (test–retest) reliability.* Six months after the investigators had completed the simulation exercises, they rescored 3 of the cases (randomly selected from among the original 20).

*Clinical case series.* From March to November 1999, investigators used the BVAS/WG instrument for WG patients evaluated at their centers in the context of routine clinical care (117 patients). Serial evaluations of 36 patients with return visits during this period permitted the measurement of sensitivity to change (discriminant validity). During this same period, 17 patients with WG were assessed independently at the same visit by 2 of the investigators at the same center (NM and RAL). Observations from these patients permitted the evaluation of interobserver reliability of the BVAS/WG and the PGA in clinical settings.

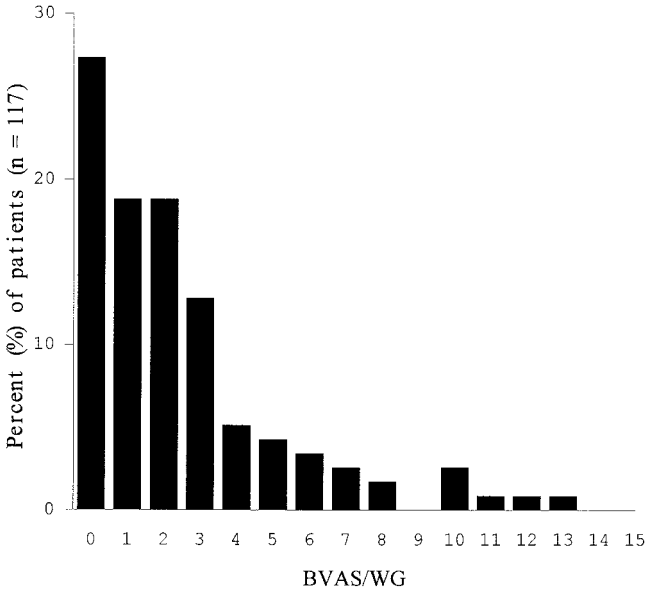


**Figure 2.** Scores of 117 clinic patients on the Birmingham Vasculitis Activity Score for Wegener's Granulomatosis (BVAS/WG) at baseline visits.

**Table 1.** Interobserver reliability: intraclass correlation coefficients for the BVAS/WG and the PGA for patients evaluated by one or more examiners*

| Series | No. of patients | No. of observers | Intraclass correlation (*P*)† | |
|---|---|---|---|---|
| | | | BVAS/WG | PGA |
| Simulation I | 10 | 16 | 0.93 (0.99) | 0.88 (0.99) |
| Simulation II | 10 | 13 | 0.91 (1.00) | 0.88 (0.99) |
| Clinical cases | 17 | 2 | 0.97 (0.68) | 0.96 (0.77) |

* BVAS/WG = Birmingham Vasculitis Activity Score for Wegener's granulomatosis; PGA = physician's global assessment (of disease activity).
† *P* values for assessing the significance of the observer effect.

**Statistical analysis.** For both the BVAS/WG and the PGA, we assessed the inter- and intraobserver reliability of repeated measures using intraclass correlation coefficients (ICC) (17), calculated as the ratio of the between-patient variance to the total variance in the scores by means of random effects analysis of variance (ANOVA) models, treating both "observer" and "patient" as random effects. We compared the precision of the BVAS/WG and the PGA by applying the signed rank test (18) to the differences between their respective coefficients of variation (CVs) within each of the 20 cases in the simulation exercises (8).

We also determined the correlation of the BVAS/WG to the PGA (construct validity) and of changes in disease activity (discriminant validity). Comparison with the PGA has a common precedent not only in the creation of activity indices in vasculitis (8) but also in other forms of rheumatic disease such as systemic lupus erythematosus and rheumatoid arthritis (1,3,19). Both construct and discriminant validity were estimated using Spearman's rank correlation coefficient (18). After determining that there was no significant observer effect in scoring, the mean scores of the BVAS/WG and the PGA for each case in the simulation exercises were used to assess construct validity. For the clinical cases, only scores from the patients' first visits were used to assess construct validity; the changes in scores between the first and the last visits for patients who had at least 2 visits were used to assess discriminant validity. All statistical analyses were performed using SAS version 6.12 (20).

## RESULTS

**Interobserver reliability.** To assess the interobserver reliability of the BVAS/WG and the PGA, we calculated the ICCs from the 2 simulation exercises, as well as in 17 real patients assessed by 2 independent observers at a single institution (NM and RAL). The ICCs were 0.93, 0.91, and 0.97 for the BVAS/WG and 0.88, 0.88, and 0.96 for the PGA, respectively (Table 1). There was no significant observer effect in the scoring of the BVAS/WG or the PGA.

**Intraobserver (test–retest) reliability.** Three randomly selected cases from the simulation exercises were rescored by 12 investigators, with a period of more than 6 months between the 2 evaluations. The ICCs for the BVAS/WG and the PGA were 0.62 and 0.28, respectively. There was no significant observer effect in the scoring of either the BVAS/WG or the PGA ($P = 1.00$ and $P = 0.96$, respectively). The low ICC for the PGA is due mainly to the fact that all 3 cases selected for reevaluation had high disease activity, and the PGA has an upper bound of 10.0. The mean PGA scores for the 3 cases were 7.98, 9.19, and 7.98, respectively; the mean BVAS/WG scores were 8.00, 13.69, and 10.08, respectively.

**Precision.** We used the data from the simulation exercises to compare the precision (interobserver variation) of the BVAS/WG and the PGA. The differences in coefficients of variation (DCVs) were calculated by computing the CVs for both BVAS/WG and PGA in each of the 20 simulation exercise cases. The DCV, which compares the BVAS/WG and the PGA with regard to the level of agreement between evaluators of the same case, was calculated as the CV of the PGA minus the CV of the BVAS/WG. A positive number in the final column of Table 2 indicates a relatively lower

**Table 2.** Precision: comparison of the differences in the CV between the BVAS/WG and the PGA*

| Case | BVAS/WG, mean ± SD | PGA, mean ± SD | Difference in CV† |
|---|---|---|---|
| 1 | 0.00 ± 0.00 | 0.07 ± 0.14 | 2.04 |
| 2 | 0.31 ± 0.79 | 0.28 ± 0.79 | 0.32 |
| 3 | 2.69 ± 0.60 | 2.22 ± 1.34 | 0.38 |
| 4 | 4.06 ± 0.25 | 3.52 ± 1.38 | 0.33 |
| 5 | 3.75 ± 0.45 | 3.78 ± 1.31 | 0.23 |
| 6 | 5.12 ± 1.20 | 5.22 ± 1.83 | 0.12 |
| 7 | 4.06 ± 1.18 | 6.12 ± 1.24 | −0.09 |
| 8 | 8.00 ± 0.00 | 7.98 ± 1.03 | 0.13 |
| 9 | 12.5 ± 1.15 | 8.68 ± 0.75 | −0.01 |
| 10 | 13.7 ± 3.00 | 9.19 ± 0.54 | −0.16 |
| 11 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.00 |
| 12 | 0.08 ± 0.28 | 0.22 ± 0.72 | −0.39 |
| 13 | 1.00 ± 0.41 | 1.01 ± 0.67 | 0.26 |
| 14 | 1.31 ± 0.48 | 2.41 ± 1.47 | 0.24 |
| 15 | 2.54 ± 0.97 | 3.28 ± 1.87 | 0.19 |
| 16 | 4.62 ± 0.51 | 4.84 ± 1.34 | 0.17 |
| 17 | 5.31 ± 1.38 | 6.08 ± 1.29 | −0.05 |
| 18 | 7.15 ± 1.52 | 7.72 ± 1.17 | −0.06 |
| 19 | 10.1 ± 2.53 | 7.98 ± 1.17 | −0.10 |
| 20 | 11.4 ± 1.50 | 8.28 ± 0.80 | 0.04 |

* Data are from simulation exercises I and II. BVAS/WG = Birmingham Vasculitis Activity Score for Wegener's granulomatosis; PGA = physician's global assessment (of disease activity); CV = coefficient of variation.
† Mean difference in CV for all cases was 0.18 ($P = 0.060$, by signed rank test).

**Table 3.** Correlation between the BVAS/WG and the PGA*

| Series | All observations | | Active disease only | |
|---|---|---|---|---|
| | No. | Spearman's r (95% CI) | No. | Spearman's r (95% CI) |
| Simulation I | 10 | 0.96 (0.83–0.99) | – | – |
| Simulation II | 10 | 1.00 (−) | – | – |
| Clinical cases | 117 | 0.92 (0.89–0.94) | 86 | 0.81 (0.73–0.87) |

* Patients in remission (defined as 0 on the physician's global assessment [PGA] [of disease activity] and 0 on the Birmingham Vasculitis Activity Score for Wegener's granulomatosis [BVAS/WG]) were excluded. 95% CI = 95% confidence interval.

CV for the BVAS/WG within the individual case (signifying greater precision); a negative number indicates a relatively lower CV for the PGA. The DCVs for the 20 cases are shown in Table 2. The mean DCV for all cases suggested greater precision for the BVAS/WG, but this comparison did not reach statistical significance ($P$ = 0.06).

**Correlation with PGA.** The correlations between the BVAS/WG and the PGA for all 3 series (the 2 simulation exercises and the clinical case series) are shown in Table 3. Figures 3A and B illustrate the BVAS/WG and PGA correlations in simulation exercise I and the clinical case series.

The 2 simulation exercises and the clinical case series included 10, 10, and 117 case evaluations, respectively. In the clinical case series, even following the exclusion of patients whose WG was in remission (the 31 cases whose scores of 0 for both the PGA and the BVAS/WG would be expected to inflate the correlations artificially), correlation between the BVAS/WG and the PGA was high (Spearman's rank correlation coefficient r = 0.81, 95% confidence interval [95% CI] 0.73–0.87).

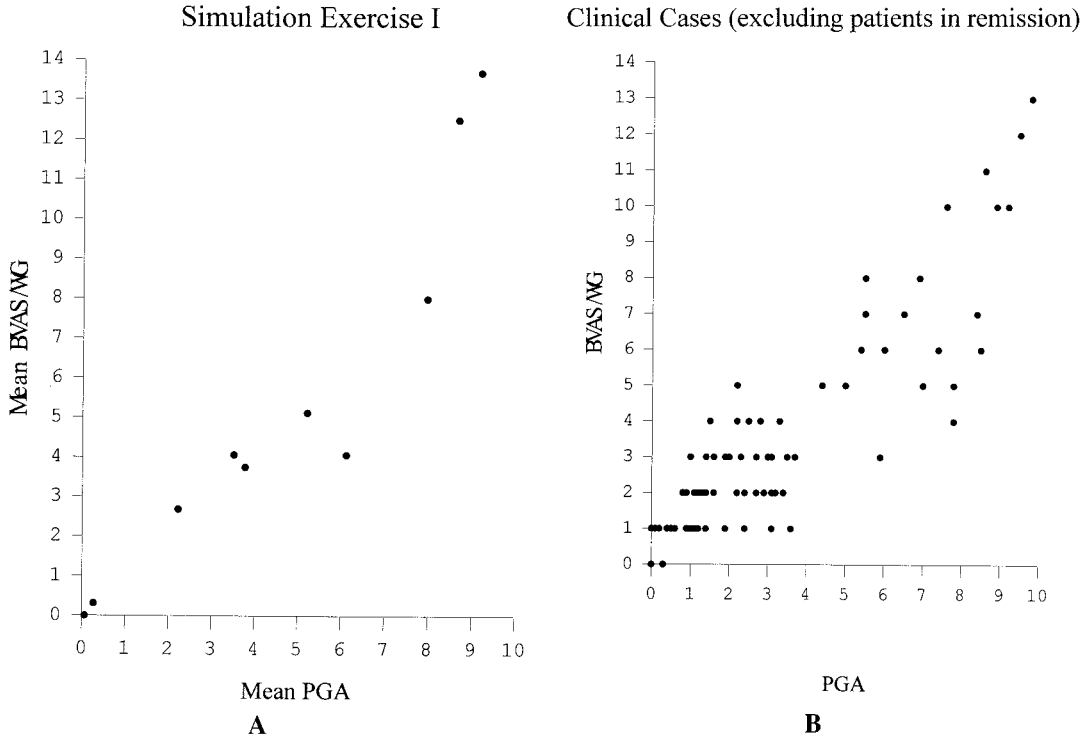


**Figure 3. A** and **B,** Correlation between the physician's global assessment (PGA) of disease activity and the Birmingham Vasculitis Activity Score for Wegener's Granulomatosis (BVAS/WG) in simulation exercise I and in the clinical case series.

**Table 4.** Discriminant validity (sensitivity to change)*

| | No. of patients | BVAS/WG, mean ± SD change† | PGA, mean ± SD change† | Spearman's r (95% CI) |
|---|---|---|---|---|
| All patients | 36 | −1.78 ± 2.43 | −1.45 ± 2.28 | 0.73 (0.53–0.85) |
| Patients with active disease only‡ | 32 | −2.00 ± 2.49 | −1.63 ± 2.88 | 0.67 (0.43–0.83) |

* BVAS/WG = Birmingham Vasculitis Activity Score for Wegener's granulomatosis; PGA = physician's global assessment (of disease activity); 95% CI = 95% confidence interval.
† Mean change between first and last visits.
‡ Excluding patients in remission (defined as 0 on the PGA and on the BVAS/WG for both visits).

**Discriminant validity.** We assessed the discriminant validity of the BVAS/WG using the changes in scores between the first and the last visits among 36 patients from the clinical case series. The intervals between visits ranged from 10 days to 224 days, with a mean of 117 days. As noted, the scales of the BVAS/WG and the PGA differ slightly (PGA has a finite upper limit of 10), but the mean changes in the BVAS/WG and the PGA occurred in the same direction and with similar magnitude (Table 4). The correlation between the BVAS/WG and the PGA was modest: Spearman's r = 0.73 (95% CI 0.43–0.83) among all patients and 0.67 (95% CI 0.43–0.83) among patients with active disease on at least 1 visit. A graphic representation of the correlation between the BVAS/WG and the PGA is shown in Figure 4.

## DISCUSSION

Multicenter trials in WG and related disorders are now under way in both the United States and Europe. The organizational strides accomplished by vasculitis investigators worldwide require parallel improvements in the ability to measure and record vasculitis activity in a uniform manner. The distinctive features of certain forms of systemic vasculitis justify the development of individual, disease-specific indices of disease activity, particularly when such efforts have direct applications in clinical trials or other patient-related research. This report describes the development and validation study of the first disease-specific activity index for any individual form of vasculitis.

In this study, we demonstrated that the BVAS/WG: 1) correlates well with the PGA of disease activity; 2) is sensitive to change; 3) has good interobserver reliability; 4) performs well in test–retest evaluations; and 5) is simple and easy to use following training. The excellent performance of the BVAS/WG throughout this series of validation studies indicates that its use in clinical investigations of this disorder is appropriate. We have used it to assess the primary outcome in a phase I trial of etanercept in WG (21) and are employing it in a phase II/III trial in this disease.

The assessment of disease activity in WG is complex and will remain so until more precise understanding of the pathophysiology of the disease yields specific, readily measured biomarkers of disease activity. For this reason, we composed an instruction manual with guidelines for use of the BVAS/WG (http://vasculitis.med.jhu.edu). We recommend that clinician-investigators read this material carefully before using the BVAS/WG with patients. Prior to the use of the BVAS/WG as a means of assessing outcomes in clinical trials, training sessions and simulation exercises using the BVAS/WG are appropriate.

Among the limitations we faced in designing the

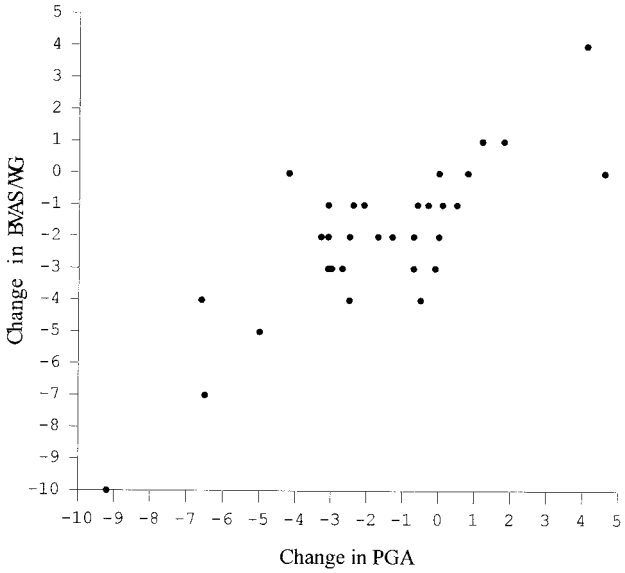


**Figure 4.** Discriminant validity: correlation between changes in the physician's global assessment (PGA) of disease activity and changes in the Birmingham Vasculitis Activity Score for Wegener's Granulomatosis (BVAS/WG) between visits.

BVAS/WG was a lack of data on grading the severity of all disease abnormalities with regard to their attendant risk of mortality or permanent disability. In patients with small and medium-sized vessel vasculitides, it is known that greater mortality occurs in the presence of pulmonary hemorrhage, glomerulonephritis, and mesenteric ischemia (22–25). In other circumstances, where risk data were not available, the panel of INSSYS investigators required consensus to determine whether a particular disease manifestation should be deemed major. Examples of conditions empirically judged to be major include central nervous system involvement, scleritis, retinal vasculitis, sensory and/or motor neuropathy, and gangrene. We acknowledge that grading these features as severe on the basis of consensus clinical judgment does not carry the weight of grades derived from data on real patients, but no such data are available.

Another limitation of our study was the use of paper cases for part of the validation process. However, the use of simulation cases is a well-established practice in outcome measure development and is the only practical method of measuring interobserver reliability among a large group of investigators studying a rare disease. Furthermore, we supplemented the validation process with a large number of real patients evaluated in the context of clinical practice.

Our study has a number of important strengths. First, we developed the new instrument from an established tool (the original BVAS), an instrument with which most of the investigators were familiar. Second, we involved clinical investigators who had both extensive practical experience in the care of patients with WG and expertise in clinical research. Third, we conducted the validation process with a variety of methods to refine the instrument's utility. Finally, we constructed not only a valid and robust tool for clinical research but also an instrument that is practical for use in clinic settings.

As the BVAS/WG is used more widely, further revisions to the instrument may be necessary. For example, weights for the minor and major BVAS items were chosen empirically at the outset of the study. Although the validation studies confirmed the appropriateness of these weights, more precise (albeit possibly less user-friendly) numeric equivalents of minor and major manifestations may result from analyses of future clinical trials.

Finally, as studies of the vasculitides advance, disease-specific instruments for the evaluation of other clinically distinct forms of vasculitis may prove useful. These instruments may include not only tools for the measurement of disease activity but also instruments for the assessment of disease damage and the impact of a disease on the quality of patients' lives.

## ACKNOWLEDGMENT

The investigators thank James Tonascia, PhD, for thoughtful review of the manuscript.

## REFERENCES

1. Petri M, Hellmann D, Hochberg M. Validity and reliability of lupus activity measures in the routine clinic setting. J Rheumatol 1992;19:53–9.
2. Liang MH, Socher SA, Roberts WN, Esdaile JM. Measurement of systemic lupus erythematosus activity in clinical research. Arthritis Rheum 1988;31:817–25.
3. Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum 1989;32: 1107–18.
4. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH, and the Committee on Prognosis Studies in SLE. Derivation of the SLEDAI: a disease activity index for lupus patients. Arthritis Rheum 1992;35:630–40.
5. Symmons DP, Coppock JS, Bacon PA, Bresnihan B, Isenberg DA, Maddison P, et al, and Members of the British Isles Lupus Assessment Group (BILAG). Development and assessment of a computerized index of clinical disease activity in systemic lupus erythematosus. QJM 1988;69:927–37.
6. Urowitz MB, Gladman DD, Tozman EC, Goldsmith CH. The lupus activity criteria count (LACC). J Rheumatol 1984;11:783–7.
7. Luqmani R, Bacon P, Moots R, Janssen B, Pall A, Emery P, et al. Birmingham Vasculitis Activity Score (BVAS) in systemic necrotizing vasculitis. QJM 1994;87:671–8.
8. Whiting-O'Keefe QE, Stone JH, Hellmann DB. Validity of a Vasculitis Activity Index for systemic necrotizing vasculitis. Arthritis Rheum 1999;42:2365–71.
9. Jayne D, Rasmussen N. Treatment of antineutrophil cytoplasm autoantibody-associated systemic vasculitis: initiatives of the European Community Systemic Vasculitis Clinical Trials Study Group. Mayo Clin Proc 1997;72:737–47.
10. Kallenberg CG, Cohen Tervaert JW, Stegeman CA. Criteria for disease activity in Wegener's granulomatosis: a requirement for longitudinal clinical studies. APMIS Suppl 1990;19:37–9.
11. Reinhold-Keller E, Kekow J, Schnabel A, Schmidt WH, Heller M, Beigel A, et al. Influence of disease manifestation and antineutrophil cytoplasmic antibody titer on the response to pulse cyclophosphamide therapy in patients with Wegener's granulomatosis. Arthritis Rheum 1994;37:919–24.
12. Hoffman G, Kerr G, Leavitt R, Hallahan C, Lebovics R, Travis W, et al. Wegener's granulomatosis: an analysis of 158 patients. Ann Intern Med 1992;116:488–98.
13. Hoffman GS, Specks U. Antineutrophil cytoplasmic antibodies. Arthritis Rheum 1998;41:1521–37.
14. Hoffman GS, Leavitt RY, Kerr GS, Fauci AS. The treatment of Wegener's granulomatosis with glucocorticoids and methotrexate. Arthritis Rheum 1992;35:1322–9.
15. Sneller MC, Hoffman GS, Talar-Williams C, Kerr GS, Hallahan CW, Fauci AS. An analysis of forty-two Wegener's granulomatosis patients treated with methotrexate and prednisone. Arthritis Rheum 1995;38:608–13.
16. Stone J, Tun W, Hellmann D. Treatment of non-life threatening

Wegener's granulomatosis with methotrexate and daily prednisone as the initial therapy of choice. J Rheumatol 1999;26:1134–9.
17. Dunn G. Design and analysis of reliability studies: the statistical evaluation of measurement errors. New York: Halstead Press; 1989.
18. Armitage P, Berry G. Statistical methods in medical research. Boston: Blackwell Scientific Publications; 1987.
19. Boers M, Tugwell P. The validity of pooled outcome measures (indices) in rheumatoid arthritis clinical trials. J Rheumatol 1993; 20:568–74.
20. SAS Institute. SAS/STAT user's guide. Cary (NC): SAS Institute; 1989.
21. Stone JH, Uhlfelder ML, Hellmann DB, Crook S, Bedocs N-M, Hoffman GS. Etanercept combined with conventional treatment in Wegener's granulomatosis. Arthritis Rheum. In press.
22. Hogan SL, Nachman PH, Wilkman AS, Jennette JC, Falk RJ. Prognostic markers in patients with antineutrophil cytoplasmic autoantibody-associated microscopic polyangiitis and glomerulonephritis. J Am Soc Nephrol 1996;7:23–32.
23. Serra A, Cameron JS, Turner DR, Hartley B, Ogg CS, Neild GH, et al. Vasculitis affecting the kidney: presentation, histopathology and long-term outcome. QJM 1984;210:181–207.
24. Savage COS, Winearls CG, Evans DJ, Rees AJ, Lockwood CM. Microscopic polyarteritis: presentation, pathology and prognosis. QJM 1985;56:467–83.
25. Guillevin L, Lhote F, Gayraud M, Cohen P, Jarrousse B, Lortholary O, et al. Prognostic factors in polyarteritis nodosa and Churg-Strauss syndrome: a prospective study in 342 patients. Medicine (Baltimore) 1996;75:17–27.