

Differentiation Between Wegener's Granulomatosis and Microscopic Polyangiitis by an Artificial Neural Network and by Traditional Methods

ROLAND LINDER, ISABELLE ORTH, E. CHRISTIAN HAGEN, FOKKO J. van der WOUDE,
and WILHELM H. SCHMITT

ABSTRACT. Objective. To investigate the operating characteristics of the American College of Rheumatology (ACR) traditional format criteria for Wegener's granulomatosis (WG), the Sørensen criteria for WG and microscopic polyangiitis (MPA), and the Chapel Hill nomenclature for WG and MPA. Further, to develop and validate improved criteria for distinguishing WG from MPA by an artificial neural network (ANN) and by traditional approaches [classification tree (CT), logistic regression (LR)].

Methods. All criteria were applied to 240 patients with WG and 78 patients with MPA recruited by a multicenter study. To generate new classification criteria (ANN, CT, LR), 23 clinical measurements were assessed. Validation was performed by applying the same approaches to an independent monocenter cohort of 46 patients with WG and 21 patients with MPA.

Results. A total of 70.8% of the patients with WG and 7.7% of the patients with MPA from the multicenter cohort fulfilled the ACR criteria for WG (accuracy 76.1%). The accuracy of the Chapel Hill criteria for WG and MPA was only 35.0% and 55.3% (Sørensen criteria: 67.2% and 92.4%). In contrast, the ANN and CT achieved an accuracy of 94.3%, based on 4 measurements (involvement of nose, sinus, ear, and pulmonary nodules), all associated with WG. LR led to an accuracy of 92.8%. Inclusion of antineutrophil cytoplasmic antibodies did not improve the allocation. Validation of methods resulted in accuracy of 91.0% (ANN and CT) and 88.1% (LR).

Conclusion. The ACR, Sørensen, and Chapel Hill criteria did not reliably separate WG from MPA. In contrast, an appropriately trained ANN and a CT differentiated between these disorders and performed better than LR. (First Release Feb 15 2011; J Rheumatol 2011;38:1039–47; doi:10.3899/jrheum.100814)

Key Indexing Terms:

DIFFERENTIAL DIAGNOSIS

WEGENER'S GRANULOMATOSIS

MICROSCOPIC POLYANGIITIS

Wegener's granulomatosis (WG) and microscopic polyangiitis (MPA) are closely related systemic vasculitides (SV). Both are associated with antineutrophil cytoplasmic antibodies (ANCA) and are grouped as ANCA-associated vasculitides. ANCA directed against proteinase 3 (PR3) are more common in WG and those directed against myeloperoxidase (MPO) are associated with MPA, but these associa-

tions do not reliably separate WG from MPA^{1,2,3}. The clinical presentation of both diseases may also overlap, as most organ systems can be affected by WG as well as MPA. Although involvement of the ears, nose, and throat (ENT) such as granulomatous rhinitis is a common finding in WG and is observed in only a minority of patients with MPA, it may also be more frequent in MPA than generally estimated: ENT involvement was described in up to 30% of MPA cases in some series^{4,5,6} and may even represent the first clinical symptom of vasculitis in up to 16% of cases⁶. Common clinical presentations in these patients with MPA include unspecific rhinitis, sinusitis, and deafness due to inner ear involvement. Bilateral nasal polyps with histological proof of vasculitis have also been described⁷. Granuloma formation, which is regarded to be indicative for WG, is often difficult to demonstrate^{8,9}. Therefore, differentiation between the 2 disorders is not always clear. Although the initial treatment of WG and MPA does not differ, the differentiation of both disorders is not only academically but also clinically important, because granuloma do not respond as well to immunosuppressive agents as sole vasculitic

From the Scientific Institute of TK for Benefit and Efficiency in Health Care (WINEG), Hamburg; Vth Medical Clinic, University Hospital Mannheim, University of Heidelberg, Mannheim; and Weinheim Kidney Center, Weinheim, Germany; and the Department of Internal Medicine, Meander Medical Center, Locatie "Lichtenberg," Amersfoort, The Netherlands.

R. Linder, MD, WINEG; I. Orth, MD; F.J. van der Woude, MD, PhD; W.H. Schmitt, MD, Vth Medical Clinic, University Hospital Mannheim, University of Heidelberg, and Weinheim Kidney Center; E.C. Hagen, PhD, Department of Internal Medicine, Meander Medical Center, Locatie "Lichtenberg."

Dr. van der Woude is deceased.

Address correspondence to Dr. W.H. Schmitt, Nierenzentrum Weinheim and Vth Medical Clinic, Mannheim University Hospital, University of Heidelberg, Röntgenstr. 1, D-69469 Weinheim, Germany.

E-mail: w.schmitt@nierenzentrum-weinheim.de

Accepted for publication December 22, 2010.

lesions do, and are associated with high relapse rates. Consequently, a refractory course of ANCA-associated vasculitis is observed more commonly in WG than MPA^{10,11,12}, and new treatment options such as B-cell depletion using rituximab may be promising to treat granuloma¹².

Classification and nomenclature of SV are complicated, and various systems have been developed. The American College of Rheumatology (ACR) classification criteria¹³ and the 1994 Chapel Hill Consensus Conference (CHCC) nomenclature¹⁴ are the systems most widely used, although the ACR criteria do not include MPA. The CHCC nomenclature provided names and working definitions on 10 different types of primary SV including MPA, but these definitions were not intended to be used as classification or diagnostic criteria. Nevertheless, they are frequently applied for this purpose under routine clinical conditions and in clinical research^{6,9}, especially in the case of MPA in the absence of alternative criteria¹⁵. Because genuine diagnostic criteria for SV are lacking, the CHCC nomenclature — supplemented with surrogate measurements for vasculitis and granuloma formation — has also been transformed into traditional format criteria sets⁹. However, application of these and modified criteria (Sørensen diagnostic criteria) to unselected cohorts of patients with vasculitis also resulted in the misclassification of both patients with WG and patients with MPA¹⁵. Therefore, criteria that reliably differentiate WG from MPA are still lacking.

Artificial neural networks (ANN) have been effectively applied to the classification of SV including giant cell arteritis¹⁶ and Churg-Strauss syndrome¹⁷. In the latter study, the ANN proved to be superior to the ACR classification criteria.

It was therefore our objective to investigate the operating characteristics of the ACR traditional format criteria for WG, the CHCC nomenclature for WG and MPA (transformed into classification criteria), and the Sørensen diagnostic criteria for WG and MPA, and then to generate and validate improved criteria for clinical differentiation between WG and MPA by using an ANN as a new approach.

MATERIALS AND METHODS

Patients from the multicenter cohort. Within a European collaboration for the standardization of ANCA assays (the EC/BCR Project for ANCA Assay Standardization), 13 referral centers in 10 European countries participated in the collection of patient data, biopsies, and sera¹. Within that project, each center had been asked to retrospectively include the last 20 consecutive patients with idiopathic SV seen in that center before the start of the study and prospectively the first 15 consecutive patients with SV after that date. Among other cases, 240 patients with WG and 78 patients with MPA had been recruited. Patients had been selected on clinical and histological criteria only, not on ANCA serology.

Classification of patients within the EC/BCR project. After entry, all patients had been classified based on data retrieved from records and at site visits. As described¹, a system for the classification of patients had been designed based on the diagnostic names and definitions adopted by the CHCC¹⁴. Patients were classified as having WG if they had histologically

proven vasculitis with granuloma and/or giant cells or if they had clinical evidence of at least 1 airway symptom or sign typical for granulomatous lesions of WG such as pulmonary nodules, subglottic stenosis, chronic rhinitis with massive crusting and epistaxis, or proliferative mastoiditis. Demonstration of orbital granuloma by computed tomography or magnetic resonance imaging also led to a classification of WG. Patients were classified as having MPA if they had systemic (extrarenal) manifestations compatible with or histology demonstrating small vessel vasculitis in the absence of airway symptoms typical for granulomatous lesions of WG. Signs and symptoms in the ENT region that could also be attributed to solely vasculitic (and not granulomatous) lesions such as inner ear deafness or unspecific rhinitis were allowed in patients with MPA.

Selection of measurements for classification. Within the EC/BCR project, a total of 79 clinical, immunological, and histological characteristics present between the date of diagnosis and the date of entry into the study had been scored¹. Clinical measurements were grouped to organ systems. For example, if either nasal crusting or bloody discharge from the nose had been present during the patient's course and had been regarded as due to WG/MPA, involvement of the nose was scored. Of those 79 measurements, 23 clinical ones, including chest radiograph, serum creatinine, and urinary findings, were assumed to be of possible importance for distinguishing WG from MPA and were considered for the development of new approaches to classification (Table 1). Missing data (0.3%) were replaced as recommended by Lee, *et al*¹⁸. It was further investigated whether the inclusion of ANCA test results would improve the classification. Therefore, the results of indirect immunofluorescence [cytoplasmic ANCA (cANCA) and perinuclear ANCA (pANCA)] and solid-phase assays (Copenhagen ELISA for antibodies against PR3 and MPO)¹ were added to the analysis. Because it was our aim to develop instruments that separate WG from MPA on the basis of clinical data, histological measurements were not considered.

Existing models of classification. We investigated how many of the patients met the ACR traditional format criteria for WG¹³ and specified how often the individual criteria were met (Table 2). Accuracy was defined as the percentage of correctly classified patients (sensitivity for the disease in question plus specificity for disease to be ruled out, divided by the number of all patients). Sensitivity was defined as the percentage of patients recognized by the set of criteria under investigation, and specificity as a negative test in the other patient group.

CHCC definitions of WG and MPA. In accordance with other investigators⁹, we tested the usefulness of the CHCC definitions (Table 3), supplemented with surrogate measurements for granuloma formation, glomerulonephritis, and small vessel vasculitis, for the classification of WG and MPA. These surrogate measurements⁹ had been adapted from the scoring systems created by the Birmingham Vasculitis Activity Score¹⁹. The association of WG and MPA with ANCA and certain clinical manifestations such as necrotizing glomerulonephritis, mentioned as "common" by the authors of the CHCC definitions, were listed as optional criteria. Their presence was not required for the assignment of a patient to the disease in question.

Sørensen diagnostic criteria for WG and MPA. Sørensen, *et al* proposed diagnostic criteria for WG and MPA (Table 4) based on the CHCC definitions⁹. They also allowed surrogate measurements for granulomatous inflammation and glomerulonephritis to replace histology⁹. Concerning WG, eosinophilia of the peripheral blood and tissues (of an undefined extent) was an exclusion criterion. However, a moderate elevation of the eosinophil count has been recognized repetitively in WG^{17,20} and was shown to result in the misclassification of patients with WG using the Sørensen criteria¹⁵. Therefore, these criteria have been modified by Lane, *et al*¹⁵, allowing eosinophilia of up to 1500 eosinophils/ μ l to be present in WG. We examined the ability of the original criteria proposed by Sørensen, *et al*⁹ and the modified criteria¹⁵ to distinguish WG from MPA (Table 4).

Newly developed models of classification. A prototypical software tool called approximation and classification of medical data (ACMD)²¹ was used to train the network. ACMD uses various strategies to improve the

Table 1. Clinical measurements used for the classification models of WG versus MPA.

| Measurement | Multicenter Cohort | | | Monocenter Cohort | | |
|----------------------------------|--------------------|--------------------|----------|-------------------|--------------------|----------|
| | WG, n = 240 (%) | MPA, n = 78 (%) | p | WG, n = 46 (%) | MPA, n = 21 (%) | p |
| Clinical measurements | | | | | | |
| Age, mean ± SD | 53.1 ± 17.0 | 62.2 ± 13.5 | < 0.0001 | 56.7 ± 16.1 | 63.4 ± 12.7 | NS |
| Men | 111 (46.3) | 46 (59.0) | NS | 27 (58.7) | 10 (47.6) | NS |
| Nose | 170 (70.8) | 4 (5.1) | < 0.0001 | 33 (71.7) | 2 (9.5) | < 0.0005 |
| Sinus | 117 (48.8) | 0 (0) | < 0.0001 | 29 (63.0) | 0 (0) | < 0.0005 |
| Ear | 85 (35.4) | 4 (5.1) | < 0.0005 | 11 (23.9) | 1 (4.8) | NS |
| Trachea | 12 (5.0) | 1 (1.3) | NS | 7 (15.2) | 1 (4.8) | NS |
| Lungs/bronchi | 161 (67.1) | 38 (48.7) | 0.005 | 28 (60.9) | 2 (9.5) | < 0.0005 |
| Kidney | 174 (72.5) | 74 (94.9) | < 0.0005 | 37 (80.4) | 21 (100) | NS |
| Creatinine (mg/dl), mean ± SD | 3.4 ± 3.5 | 4.4 ± 4.0 | < 0.005 | 3.4 ± 3.7 | 5.1 ± 4.4 | < 0.01 |
| Hematuria | 151 (62.9) | 70 (89.7) | < 0.0005 | 37 (80.4) | 21 (100) | NS |
| Proteinuria | 142 (59.2) | 64 (82.1) | < 0.0005 | 35 (76.1) | 21 (100) | < 0.05 |
| Skin | 72 (30.0) | 26 (33.3) | NS | 9 (19.6) | 1 (4.8) | NS |
| Peripheral nerves | 53 (22.1) | 22 (28.2) | NS | 19 (41.3) | 6 (16.7) | NS |
| Central nervous system | 24 (10.0) | 2 (2.6) | NS | 3 (6.5) | 0 (0) | NS |
| Eyes | 82 (34.2) | 7 (9.0) | < 0.0005 | 14 (30.4) | 1 (4.8) | NS |
| Orbital pseudotumor | 7 (2.9) | 0 (0) | NS | 1 (2.2) | 0 (0.0) | NS |
| Abdomen | 40 (16.7) | 7 (9.0) | NS | 2 (4.4) | 1 (4.8) | NS |
| Joints/muscles | 140 (58.3) | 32 (41.0) | < 0.05 | 28 (60.9) | 8 (38.1) | NS |
| Large vessels | 18 (7.5) | 2 (2.6) | NS | 3 (6.5) | 2 (9.5) | NS |
| Heart | 29 (12.1) | 6 (7.7) | NS | 4 (4.4) | 1 (4.8) | NS |
| Constitutional symptoms | 200 (83.3) | 64 (82.1) | NS | 34 (73.9) | 8 (38.1) | 0.01 |
| Radiological measurements | | | | | | |
| Pulmonary nodule | 92 (38.3) | 1 (1.3) | < 0.0005 | 12 (26.1) | 0 (0) | < 0.05 |
| Pulmonary infiltrate | 79 (32.9) | 28 (35.9) | NS | 21 (45.7) | 2 (9.5) | < 0.01 |
| ANCA serology* | | | | | | |
| PR3 ANCA | 124 (51.7) | 22 (28.2) | < 0.0005 | 17 (40.0) | 8 (38.1) | NS |
| MPO ANCA | 55 (22.9) | 38 (48.7) | < 0.0005 | 1 (2.2) | 13 (61.9) | < 0.0005 |
| cANCA | 126 (52.5) | 16 (20.5) | < 0.0005 | 20 (43.5) | 8 (38.1) | NS |
| pANCA | 45 (18.8) | 34 (43.6) | < 0.0005 | 1 (2.2) | 13 (61.9) | < 0.0005 |

* ANCA serology was available for 203 patients with WG and 66 patients with MPA from the multicenter cohort and for all cases from the monocenter cohort. WG: Wegener's granulomatosis; MPA: microscopic polyangiitis; ANCA: antineutrophil cytoplasmic antibodies; PR3 ANCA: proteinase 3 ANCA; MPO ANCA: myeloperoxidase ANCA; cANCA: cytoplasmic ANCA; pANCA: perinuclear ANCA; NS: not significant.

training of self-learning ANN, e.g., early stopping to avoid overfitting the network²² and ensembles to improve robustness and generalization performance²³. Adaptive propagation²⁴ was used as the learning algorithm, a further development of the back-propagation algorithm²⁵. To control for the so-called peaking phenomenon²⁶, feature selection was performed by the neural net clamping technique²⁷.

Figure 1 shows the ANN structure used in our study. During training, datasets with a known outcome were entered at the input neurons (input variables, either continuous or categorical) and at the output neuron (binary output variables, namely WG or MPA). After starting the network, the input data were processed in the hidden and output layers, resulting in a number between 1 and 0 at the output neuron, representing assignment to WG or MPA. The activity of the output neuron depended on the inputs and the weights at the connections. The key feature of ANN is that the weights at the connections are "learned" during training of the network. "Experience" in the trained network is stored in these interconnection weights²⁸. The system begins with random weights at the connections between the neurons. The software correlates the network output with the actual outcome and calculates an error value. The ANN attempts to mini-

mize the error by adjusting the weights at the connections according to a learning algorithm²⁸. This process is repeated a predefined number of times during the training phase. At the end of the learning process, the optimum weight factors are fixed. In the user phase, data from cases not previously interpreted by the network are entered, and an output is calculated based on now-fixed weight factors²⁹.

Binary logistic regression (LR) was calculated by the software SPSS V.11.5.1 (SPSS Inc., Chicago, IL, USA). Feature selection was done by backward search using the Wald test. Exclusively default settings were chosen.

As a tool for generating the classification tree (CT) we used the chi-squared automatic interaction detector, a module of the SPSS analysis software (Answer Tree V.3.1, SPSS Inc.). While building up the CT, the number of levels below the root was restricted to 5, the minimum number of records per main node was set at 10, and the minimum number of records per end node was set at 5. Otherwise, default settings were chosen.

Validation of allocation methods. All 3 methods of allocation (ANN, LR, and CT) were validated by 2 approaches: using the leave-one-out method, and using an independent monocenter cohort of patients with WG and MPA.

Table 2. ACR 1990 traditional format criteria for the classification of WG applied to the multicenter cohort of 240 patients with WG and 78 patients with MPA.

| Criteria | No. Patients Fulfilling Criteria | | Accuracy, % |
|--|----------------------------------|---------------|----------------|
| | WG, n (%) | MPA, n (%) | |
| Nasal or oral inflammation (purulent or bloody nasal discharge, oral ulcers) | 178 (74.2) | 7 (9.0) | 78.3 |
| Abnormal chest radiograph (fixed infiltrates, nodules, cavities) | 92 (38.3) | 1 (1.3) | 53.7 |
| Urinary sediment (red cell casts, > 5 red blood cells) | 137 (57.1) | 66 (84.6) | 49.9 |
| Granulomatous inflammation on biopsy (any) | 65 (27.1) | 0 (0) | 45.0 |
| Patients to be classified as WG (≥ 2/4 criteria positive) | 170 (70.8) | 6 (7.7) | 76.1 |

ACR: American College of Rheumatology; WG: Wegener's granulomatosis; MPA: microscopic polyangiitis.

Table 3. Chapel Hill nomenclature for WG and MPA as applied to the multicenter cohort of 240 patients with WG and 78 patients with MPA.

| Criteria | No. Patients Fulfilling Criteria | | Accuracy, % |
|---|----------------------------------|--------------------------|--------------------|
| | WG, n (%) | MPA, n (%) | |
| Mandatory criteria for WG | | | |
| Granulomatous inflammation involving the respiratory tract Including patients fulfilling surrogate measurements of granulomatous inflammation (respiratory tract) | 56 (23.3) | 0 (0) | 42.1 |
| Necrotizing vasculitis affecting small to medium-size vessels Including patients fulfilling surrogate measurements of necrotizing vasculitis (small to medium-size vessels) | 231 (96.3) | 8 (10.3) | 94.7 |
| Patients classified as WG without surrogate measurements | 95 (39.6) | 31 (39.7) | 44.7 |
| Patients classified as WG including surrogate measurements | 214 (89.2) | 70 (89.7) | 69.8 |
| Patients classified as WG without surrogate measurements | 32 (13.3) | 0 (0) | 35.0 |
| Patients classified as WG including surrogate measurements | 206 (85.8) | 8 (10.3) | 86.8 |
| Mandatory criteria for MPA | | | |
| Necrotizing vasculitis with few or no immune deposits affecting small to medium-size vessels Including patients fulfilling surrogate measurements of necrotizing vasculitis (small to medium-size vessels) | 95 (39.6) | 31 (39.7) | 55.3 |
| Patients classified as MPA without surrogate measurements | 214 (89.2) | 70 (89.7) | 30.2 |
| Patients classified as MPA including surrogate measurements | 95 (39.6) | 31 (39.7) | 55.3 |
| Patients classified as MPA including surrogate measurements | 214 (89.2) | 70 (89.7) | 30.2 |
| Optional criteria for both WG and MPA | | | |
| Necrotizing glomerulonephritis Including patients fulfilling surrogate measurements of glomerulonephritis | 118 (49.2) | 60 (76.9) | 42.8* |
| Association with ANCA | 163 (67.9) | 74 (94.9) | 44.2 [†] |
| | 183 (90.2) of 203 cases | 56 (84.9) of 66 cases | 71.7 ^{††} |
| Additional optional criteria for MPA only | | | |
| Necrotizing arteritis in small to medium-size arteries | No data | No data | — |
| Pulmonary capillaritis Including patients fulfilling surrogate measurements of pulmonary capillaritis | 23 (9.6) | 0 (0) | 68.2 |
| | 88 (36.7) | 21 (26.9) | 54.4 |

* Accuracy for WG; for MPA 49.7%. [†] Accuracy for WG; for MPA 39.9%. ^{††} Accuracy for WG; for MPA 28.3%. WG: Wegener's granulomatosis; MPA: microscopic polyangiitis; ANCA: antineutrophil cytoplasmic antibodies.

The leave-one-out method means that the algorithms for classification are established by using $n - 1$ of all cases, and to use the missing case for validation. By performing n rounds of classification, every case is used once for the purpose of validation.

The independent monocenter cohort consisted of 46 consecutive patients with WG and 21 consecutive patients with MPA from the Mannheim University hospital (academic referral center; nephrology/rheumatology unit). Independently of our study, patients had been assessed by an interdisciplinary team including nephrologists, rheumatologists, ENT, and eye specialists, and had undergone an extensive imaging proce-

dures. The clinical diagnosis (WG or MPA) had been made according to the same guidelines as described for the multicenter cohort.

Statistical methods. The 2-tailed Wilcoxon test was used for numerical variables such as age and creatinine. Otherwise, the chi-squared test was used.

RESULTS

Histology. By definition, granuloma formation or giant cells could be present only in biopsy specimens of patients with

Table 4. Sørensen diagnostic criteria for WG and MPA⁹ applied to the multicenter cohort for the classification of WG (n = 155)* vs MPA (n = 46)*.

| Criteria | No. Patients Fulfilling Criteria WG, n* (%) | MPA, n* (%) | Accuracy, % |
|--|---|----------------|----------------|
| Criteria for WG | | | |
| 1. Biopsy/surrogate measurement for granulomatous inflammation in the respiratory system | 149 (96.1) | 5 (10.9) | 94.5 |
| 2. Biopsy-verified necrotizing vasculitis in small to medium-size vessels or biopsy/surrogate measurement for glomerulonephritis or positive PR3 ANCA | 146 (94.2) | 46 (100) | 72.6 |
| 3. Lack of eosinophilia (blood/biopsy samples) | 106 (68.4) | 35 (76.1) | 58.2 |
| Patients classified as WG (criteria 1–3 positive) | 92 (59.4) | 3 (6.5) | 67.2 |
| Patients classified as WG as modified by Lane, <i>et al</i> ¹⁵ (criteria 1 and 2 positive) | 133 (85.8) | 5 (10.9) | 86.6 |
| Criteria for MPA | | | |
| 1. Biopsy-verified necrotizing vasculitis in small vessels and/or glomerulonephritis with few or no immune deposits | 171 (71.3) | 73 (93.6) | 44.7 |
| 2. Involvement of > 1 organ system, as indicated by biopsy-verified vasculitis in small to medium-size vessels or a surrogate measurement for glomerulonephritis | 145 (60.4) | 70 (89.7) | 51.9 |
| 3. Lack of biopsy and surrogate measurement or granulomatous inflammation in the respiratory system | 9 (3.8) | 70 (89.7) | 94.7 |
| Patients to be classified as MPA (3/3 criteria positive) | 4 (1.7) | 58 (74.4) | 92.4 |

* Results of testing for PR3 ANCA plus determination of eosinophil counts of the peripheral blood were available for 155/240 patients with WG and 46/78 patients with MPA, resulting in a lower number of patients available to test the criteria for WG. WG: Wegener's granulomatosis; MPA: microscopic polyangiitis; PR3 ANCA: proteinase 3 antineutrophil cytoplasmic antibodies.

WG and could be demonstrated in 27% of the 240 patients with WG from the multicenter cohort (monocenter cohort 33%). Pauciimmune crescentic glomerulonephritis in conjunction with airway symptoms compatible with WG was shown in 49% (54%). Nonrenal vasculitis was demonstrated by histology in 31% (22%). In 10% (24%) of WG cases, no histology was available. By renal biopsy, crescentic glomerulonephritis compatible with MPA was found in 77% of the 78 cases of the multicenter cohort (monocenter cohort 21%), although these findings of cause were insufficient to

separate these cases from WG. Extrarenal biopsy demonstrated MPA in 29% (0%) of cases. No histology was available in 5% (0%) of the patients with MPA.

The distribution of the affected organs is given in Table 1. Involvement of the nose, sinuses, and ears, as well as pulmonary nodules, was associated with WG in both cohorts ($p < 0.0005$ in multicenter cohort). In contrast, renal involvement was significantly associated with MPA in the multicenter cohort ($p < 0.005$) with a similar trend in the monocenter cohort ($p = 0.08$). Among patients with MPA,

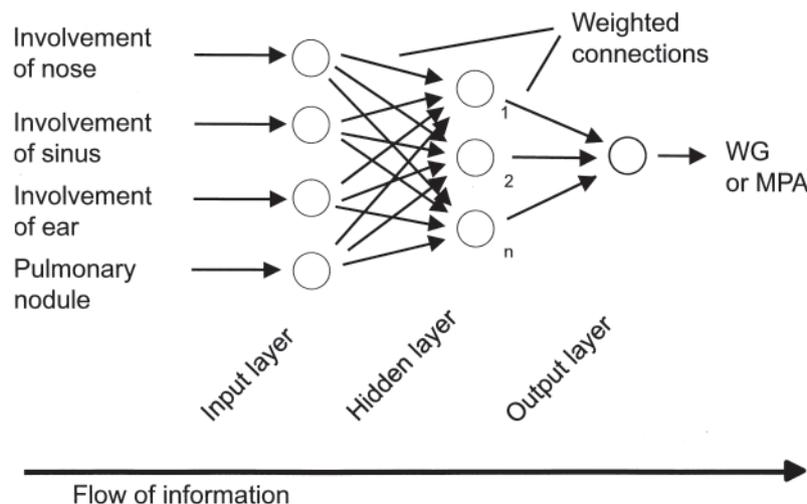


Figure 1. Structure of the artificial neural network (ANN) applied to the classification of patients with Wegener's granulomatosis (WG) versus patients with microscopic polyangiitis (MPA). Numbers 1 and 2 refer to the number of hidden neurons within the hidden layer; n = any number.

serum creatinine was significantly higher in both cohorts as compared with WG ($p < 0.01$). In the multicenter, but not in the monocenter cohort, cANCA and PR3 ANCA were significantly associated with WG ($p < 0.0005$). MPO ANCA and pANCA were associated with MPA in both cohorts ($p < 0.0005$).

Existing classifications. A total of 70.8% of patients with WG fulfilled the ACR traditional format criteria for WG (Table 2, accuracy 76.1%). The criteria “granulomatous inflammation on biopsy” and “abnormal chest radiograph” were least often met. The specificity for WG was 92.3%. False-positive cases (patients with MPA) mostly had oral ulcers in conjunction with a nephritic urinary sediment.

Traditional format criteria of the CHCC definition for WG were met by only 13.3% of the patients with WG and none of the MPA cases (Table 3, accuracy 35%). It is noteworthy that granulomatous inflammation of the respiratory tract was histologically proven in only 23.3% of cases. However, substitution of clinical findings with measurements used as surrogate histological data for both granulomatous inflammation and vasculitis increased the sensitivity for WG to 85.8%, with a specificity of 89.7% (accuracy 86.8%). Traditional format criteria for MPA, from the CHCC definition for MPA, had a sensitivity of only 39.7% for MPA, with a specificity of 60.4% (Table 3, accuracy 55.3%). Substitution of histological findings by surrogate measurements increased sensitivity to 89.7%, at the cost of a reduced specificity (10.8%; accuracy 30.2%).

Applying the Sørensen criteria for WG to the cohort of WG cases led to a sensitivity of 59.4%. Because eosinophilia of the tissues or blood (eosinophil count $> 500/\mu\text{l}$) was present in 31.6% of WG cases, the criterion “lack of eosinophilia” was least often met (Table 4, accuracy 67.2%). Allowing the eosinophil count to be as high as $1500/\mu\text{l}$ increased the sensitivity to 86.9% (accuracy 86.6%). Positivity for PR3 ANCA, which is part of these criteria, was found in 61% of patients with WG and 33% of patients with MPA and did not help to separate these 2 disorders. Sørensen criteria for MPA had a specificity for MPA of 97.3% and a sensitivity of 74.4% (Table 4, accuracy 92.4%).

Newly developed models of allocation. The ANN was initially trained with 23 clinical measurements, excluding ANCA test results (Table 1). The backward search revealed 4 of these measurements to be relevant for distinguishing between WG and MPA (Figure 1): pulmonary nodules and involvement of nose, sinuses, and ears (all associated with WG). Using these 4 measurements as input neurons and 1 hidden layer, the ANN correctly assigned 230/240 patients with WG (95.8%) and 70/78 patients with MPA (89.7%; accuracy 94.3%, Table 5). Inclusion of the other 19 clinical measurements listed in Table 1 did not further improve the assignment of patients. Validation using the monocenter cohort resulted in the correct assignment of 91% of cases. Involvement of the nose (6 cases), ears (5 cases), and a pul-

monary nodule (1 case) in patients with MPA was associated with the incorrect assignment to WG in both cohorts. Including ANCA test results during the training phase of the ANN did not improve the assignment of patients from the multicenter cohort. However, 1 additional patient with MPA (MPO ANCA-positive) with involvement of the ear (deafness of the inner ear due to suspected vasculitis in the absence of otitis media; reversed by glucocorticosteroids) was correctly classified as MPA when the ANCA status had been included during the training phase, with a resulting accuracy of 92.5%.

In logistic regression, feature selection considered 10 measurements significant: male sex, nose, sinus, ear, lungs/bronchi, hematuria, eyes, joints/muscles, large vessels, and pulmonary nodules. Use of these measurements resulted in the correct classification of 92.8% of cases from the multicenter and 88.1% from the monocenter cohort, using the leave-one-out methodology for validation (Table 5). Inclusion of ANCA test results slightly reduced the number of correctly assigned patients to 86.6%.

Allocation of patients by CT employed the same measurements that were also selected by the ANN, i.e., involvement of nose, sinuses, and ears, as well as pulmonary nodules. Presence of each of these 4 measurements leads to the classification as WG, otherwise MPA. Thus, the CT could also be described in the traditional format as a table with 4 criteria. The accuracy obtained was equal to that of the ANN (94.3% for the multicenter and 91.0% for the monocenter cohort; Table 5). Adding ANCA test results to the measurements under consideration did not change the results.

DISCUSSION

Our study shows that an ANN can distinguish between WG and MPA with an accuracy of over 90% based solely on clinical data. An identical accuracy was achieved by a CT approach. These newly developed models were superior to established instruments (ACR traditional format criteria for WG, CHCC nomenclature for WG and MPA used as classification criteria, and Sørensen diagnostic criteria).

The differentiation between WG and MPA based on clinical data is not easy. First, ENT involvement is commonly regarded as a hallmark of WG. However, it was observed in 13.9% of the MPA cases described here. Moreover, studies addressing the frequency of ENT lesions in different patients with SV described ENT involvement in up to 30% of patients with MPA^{4,5,6}. Second, formation of granulomata and giant cells is restricted to WG. But proof of these lesions by histology is often difficult to obtain and is commonly available in $\leq 50\%$ of cases^{9,17} (28% of the patients described here). In a prospective study performed under routine conditions, only 6 of 25 patients with newly diagnosed WG had biopsy-proven disease⁸. This illustrates that classification criteria that depend solely on histology are difficult to work with in clinical practice. That is why the classifica-

Table 5. Distinction of WG from MPA by artificial neural network (ANN), logistic regression (LR), and classification tree (CT)/traditional format criteria (TF). Sensitivity means the ratio of patients correctly classified with WG; specificity is the ratio of patients correctly classified with MPA.

| Method | Measurements Used, n | Sensitivity, % | Specificity, % | Accuracy, % |
|--------------------|----------------------|----------------|----------------|-------------|
| ANN | | | | |
| Multicenter cohort | 4 | 95.8 | 89.7 | 94.3 |
| Monocenter cohort | 4 | 93.5 | 85.7 | 91.0 |
| LR | | | | |
| Multicenter cohort | 23 | 94.2 | 88.5 | 92.8 |
| Monocenter cohort | 23 | 93.5 | 76.2 | 88.1 |
| CT/TF | | | | |
| Multicenter cohort | 4 | 95.8 | 89.7 | 94.3 |
| Monocenter cohort | 4 | 93.5 | 85.7 | 91.0 |

WG: Wegener's granulomatosis; MPA: microscopic polyangiitis.

tion procedures presented here do without histology. Accordingly, criteria for WG/MPA delineated from the CHCC nomenclature were met by only 13.3% of the patients with WG and 39.6% of patients with MPA of the multicenter cohort, frequencies that are in line with previous studies⁹. Also, the ACR traditional format criteria for WG depend partially on histology, and when applied to the multicenter cohort of patients led to a sensitivity of only 70.8%. This is in line with Rao, *et al*, who found the positive predictive value of the ACR classification criteria to be as low as 29%³⁰.

Surrogate measurements that replace the histological proof of granulomata, giant cells, and vasculitis by clinical measurements have been used by various investigators. We demonstrated that the use of surrogate measurements increased the accuracy of the CHCC criteria for WG from 35.0% to 86.8% and contributed to the accuracy of the Sørensen criteria. However, the performance of the CHCC criteria for MPA did not improve with surrogate measurements (accuracy 55.3% without and 30.2% with surrogate measurements). These criteria did not separate WG from MPA, which suggests that they are not of value in classification. On the other hand, it should be stressed that the use of surrogate measurements should not replace the search for appropriate histological confirmation, as biopsy material is essential to rule out other nonvasculitic conditions such as neoplasia and infection.

Our second objective was to develop new instruments that better distinguish between WG and MPA on clinical grounds: the rarity of ENT involvement in MPA, which was certainly influenced by the definition of MPA and might underestimate the true frequency of ENT involvement in this disease, made the ENT-associated measurements the most important discriminators, together with the presence or absence of pulmonary nodules on the chest radiograph. Both the ANN and CT relied on these measurements and led to a better separation between WG and MPA than all previously established methods of assignment (accuracy 94.3% in

training and 91.0% in validation cohort). One may argue that accuracy is not the appropriate measurement when data are not balanced (WG: n = 240; MPA: n = 78). However, this imbalance reflects the real *a priori* distribution of WG:MPA of about 3:1³¹, with an even more pronounced predominance of WG in northern latitudes. The results demonstrate that WG and MPA can be distinguished solely by clinical measurements in the vast majority of cases. Only 18 of 318 patients from the multicenter cohort (monocenter validation cohort 6 of 67 patients) were misclassified. Besides ear and nose manifestations, there were further very selective measurements such as an orbital pseudotumor (seen only in WG) and tracheal involvement that strongly favored the diagnosis of WG. However, the measurements were associated with a low sensitivity and were not used by both the CT and the ANN.

Within the multicenter cohort, no differences could be detected in the sensitivity and specificity dependent on whether data on the patients were collected retrospectively or prospectively. This is probably because ENT involvement is usually present very early during the course of WG.

Although PR3 ANCA are more common in WG than in MPA^{1,2,3}, the inclusion of ANCA test results did not improve allocation within the monocenter cohort, and only 1 additional case of MPA from the validation cohort was correctly allocated by the ANN. This made the ANN the most accurate model of classification for the validation cohort. The good performance of the ANN is in line with previous reports that demonstrated a superiority of appropriately trained ANN compared to conventional methods in the classification of SV^{16,17}. Further, multiple other applications underline that ANN are promising tools to address clinical problems such as prognosis estimation and risk assessment. Examples include the prediction of survival and complications after percutaneous endoscopic gastrostomy³², the prediction of outcome in subdural hematoma³³, and the classification of schizophrenic patients³⁴, where the ANN proved to be superior to established models.

The advantage of ANN is that they reveal nonlinear relationships and have the ability to analyze the interaction between many variables at different levels. For example, there is evidence that the superiority of ANN is influenced by their ability to adjust the importance of certain measurements depending on the presence or absence of other variables³⁵. Usually, the ANN software allows identification of the specific input variables that have most value in terms of predictive accuracy. However, these should not be viewed as independent predictive variables as perceived by a clinician. Because ANN process data in a nonlinear way, the network logic of classification cannot be broken down into simple elements of clinical reasoning³⁵.

There are limitations to the ANN-based model and the CT/LR presented here. First, the number of possible outcomes was limited, as the approach was restricted to only 2 different SV. More meaningful and probably more complex models would have arisen if further vasculitic disorders had been incorporated into the analysis. But within the EC/BCR study, the number of datasets from these diseases was by far too small to train an ANN or to develop other reliable models of allocation. Nevertheless, because explicit classification criteria for the differentiation of WG and MPA have so far not been developed, the results presented here may help to separate these 2 disorders. Because WG and MPA differ from each other also with respect to their further course, for example in terms of relapse rate³⁶, a correct allocation will also have prognostic implications. Second, the ANN has not been tested against nonvasculitic disorders. This approach would help to develop diagnostic criteria that are urgently awaited to separate SV from other disorders that mimic vasculitis. In contrast to the separation of WG and MPA, the ANCA test might be very important in this setting, provided the pretest probability for the presence of vasculitis is rather high. To rule out or confirm vasculitis, histology will also remain important. Models that solve these diagnostic problems may be of higher value to the clinician than models for classification. However, datasets from vasculitis and control patients that can be used to develop diagnostic criteria are not currently available, to our knowledge.

Established criteria such as the ACR traditional format criteria for WG, the Sørensen criteria for WG and MPA, and criteria of the CHCC nomenclature did not reliably separate WG from MPA. In contrast, both a newly formulated and easy to use CT and an appropriately trained ANN — based on clinical data and not on histology — correctly assigned the majority of patients, and were associated with an accuracy of 91% when validated by application to an independent cohort. Both methods were superior to LR. The addition of ANCA test results to the criteria under consideration slightly improved the performance of the ANN, but not of the CT and LR. ANN offer exciting prospects for many applications in clinical medicine and warrant prospective testing in the classification of vasculitides.

REFERENCES

- Hagen EC, Daha MR, Hermans J, Andrassy K, Csernok E, Gaskin G, et al. Diagnostic value of standardized assays for anti-neutrophil cytoplasmic antibodies in idiopathic systemic vasculitis. EC/BCR Project for ANCA Assay Standardization. *Kidney Int* 1998;53:743-53.
- Wiik AS. Clinical use of serological tests for antineutrophil cytoplasmic antibodies. What do the studies say? *Rheum Dis Clin North Am* 2001;27:799-813.
- Falk RJ, Hogan S, Carey TS, Jennette JC. Clinical course of anti-neutrophil cytoplasmic autoantibody-associated glomerulonephritis and systemic vasculitis. The Glomerular Disease Collaborative Network. *Ann Intern Med* 1990;113:656-63.
- Paulsen JI, Rudert H. Manifestations of primary vasculitis in the ENT region [German]. *Z Rheumatol* 2001;60:219-25.
- Metaxaris G, Prokopakis EP, Karatzanis AD, Sakelaris G, Heras P, Velegarakis GA, et al. Otolaryngologic manifestations of small vessel vasculitis. *Auris Nasus Larynx* 2002;29:353-6.
- Lane SE, Watts RA, Shepstone L, Scott DG. Primary systemic vasculitis: clinical features and mortality. *QJM* 2005;98:97-111.
- Kokan N, Hosomi Y, Inamoto S, Onishi K, Tanimoto H, Nibu K. Microscopic polyangiitis histologically confirmed by biopsy from nasal cavity and paranasal sinuses: a case report. *Rheumatol Int* 2006;26:936-8.
- Rao JK, Allen NB, Feussner JR, Weinberger M. A prospective study of antineutrophil cytoplasmic antibody (c-ANCA) and clinical criteria in diagnosing Wegener's granulomatosis. *Lancet* 1995;346:926-31.
- Sorensen SF, Slot O, Tvede N, Petersen J. A prospective study of vasculitis patients collected in a five year period: evaluation of the Chapel Hill nomenclature. *Ann Rheum Dis* 2000;59:478-82.
- Schmitt WH, Birk R, Heinzel PA, Göbel U, Choi M, Warnatz K, et al. Prolonged treatment of refractory Wegener's granulomatosis with 15-deoxyspergualin: An open study in seven patients. *Nephrol Dial Transplant* 2005;20:1083-92.
- Schmitt WH, Hagen EC, Neumann I, Nowack R, Flores-Suarez LF, van der Woude FJ; European Vasculitis Study Group. Treatment of refractory Wegener's granulomatosis with antithymocyte globulin (ATG): An open study in 15 patients. *Kidney Int* 2004;65:1440-8.
- Jones RB, Ferraro AJ, Chaudry AN, Brogan P, Salama AD, Smith KGC, et al. A multicenter survey of rituximab for refractory anti-neutrophil cytoplasm antibody-associated vasculitis. *Arthritis Rheum* 2009;60:2156-68.
- Leavitt RY, Fauci AS, Bloch DA, Michel BA, Hunder GG, Arend WP, et al. The American College of Rheumatology 1990 criteria for the classification of Wegener's granulomatosis. *Arthritis Rheum* 1990;33:1101-7.
- Jennette JC, Falk RJ, Andrassy K, Bacon PA, Churg J, Gross WL, et al. Nomenclature of systemic vasculitides. Proposal of an international consensus conference. *Arthritis Rheum* 1994;37:187-92.
- Lane SE, Watts RA, Barker TH, Scott DG. Evaluation of the Sorensen diagnostic criteria in the classification of systemic vasculitis. *Rheumatology* 2002;41:1138-41.
- Astion ML, Wener MH, Thomas RG, Hunder GG, Bloch DA. Application of neural networks to the classification of giant cell arteritis. *Arthritis Rheum* 1994;37:760-70.
- Schmitt WH, Linder R, Reinhold-Keller E, Gross WL. Improved differentiation between Churg-Strauss syndrome and Wegener's granulomatosis by an artificial neural network. *Arthritis Rheum* 2001;44:1887-96.
- Lee IN, Liao SC, Embrechts M. Data mining techniques applied to medical information. *Med Inform Internet Med* 2000;25:81-102.
- Luzmani RA, Bacon PA, Moots RJ, Janssen BA, Pall A, Emery P, et al. Birmingham Vasculitis Activity Score (BVAS) in systemic

- necrotizing vasculitis. *QJM* 1994;87:671-8.
20. Potter MB, Fincher RK, Finger DR. Eosinophilia in Wegener's granulomatosis. *Chest* 1999;116:1480-3.
 21. Linder R, Pöpl SJ. ACMD: A practical tool for automatic neural net based learning. *Lect Notes Comp Science* 2001;2199:168-73.
 22. Prechtelt L. Automatic early stopping using cross validation: quantifying the criteria. *Neural Netw* 1998;11:761-7.
 23. Bishop CM. *Neural networks for pattern recognition*. Oxford: Clarendon Press; 1995.
 24. Linder R, Wirtz S, Pöpl SJ. *Speeding up backpropagation learning by the APROP algorithm*. Millet, Alberta: ICSC Academic Press; 2000.
 25. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986;323:533-6.
 26. Kanal L. On dimensionality and sample size in statistical pattern recognition. *Pattern Recognit* 1971;3:225-34.
 27. Wang W, Jones P, Partridge D. *Ranking pattern recognition features for neural networks*. Berlin: Springer; 1998.
 28. Geddes CC, Fox JG, Allison ME, Boulton-Jones JM, Simpson K. An artificial neural network can select patients at high risk of developing progressive IgA nephropathy more accurately than experienced nephrologists. *Nephrol Dial Transplant* 1998;13:67-71.
 29. Van Biesen W, Sieben G, Lameire N, Vanholder R. Application of Kohonen neural networks for the non-morphological distinction between glomerular and tubular renal disease. *Nephrol Dial Transplant* 1998;13:59-66.
 30. Rao JK, Allen NB, Pincus T. Limitations of the 1990 American College of Rheumatology classification criteria in the diagnosis of vasculitis. *Ann Intern Med* 1998;129:345-52.
 31. Gibson A, Stamp LK, Chapman PT, O'Donnell JL. The epidemiology of Wegener's granulomatosis and microscopic polyangiitis in a Southern Hemisphere region. *Rheumatology* 2006;45:624-8.
 32. Takayama T, Takayama K, Inoue N, Funakoshi S, Serizawa H, Watanabe N, et al. Prediction of survival and complications after percutaneous endoscopic gastrostomy in an individual by using clinical factors with an artificial neural network system. *Eur J Gastroenterol Hepatol* 2009;21:1279-85.
 33. Abouzari M, Rashidi A, Zandi-Toghiani M, Behzadi M, Asadollahi M. Chronic subdural hematoma outcome prediction using logistic regression and an artificial neural network. *Neurosurg Rev* 2009;32:479-84.
 34. Bose SK, Turkheimer FE, Howes OD, Mehta MA, Cunliffe R, Stokes PR, et al. Classification of schizophrenic patients and healthy controls using [18F] fluorodopa PET imaging. *Schizophr Res* 2008;106:148-55.
 35. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;346:1135-8.
 36. Jayne D, Rasmussen N, Andrassy K, Bacon P, Tervaert JW, Dadoniené J, et al; European Vasculitis Study Group. A randomized trial of maintenance therapy for vasculitis associated with antineutrophil cytoplasmic autoantibodies. *N Engl J Med* 2003;349:36-44.