Original Investigation

Clinical Vestibular Testing Assessed With Machine-Learning Algorithms

Adrian J. Priesol, MD; Mengfei Cao, BA; Carla E. Brodley, PhD; Richard F. Lewis, MD

IMPORTANCE Dizziness and imbalance are common clinical problems, and accurate diagnosis depends on determining whether damage is localized to the peripheral vestibular system. Vestibular testing guides this determination, but the accuracy of the different tests is not known.

OBJECTIVE To determine how well each element of the vestibular test battery segregates patients with normal peripheral vestibular function from those with unilateral reductions in vestibular function.

DESIGN, SETTING, AND PARTICIPANTS Retrospective analysis of vestibular test batteries in 8080 patients. Clinical medical records were reviewed for a subset of individuals with the reviewers blinded to the vestibular test data.

INTERVENTIONS A group of machine-learning classifiers were trained using vestibular test data from persons who were "manually" labeled as having normal vestibular function or unilateral vestibular damage based on a review of their medical records. The optimal trained classifier was then used to categorize patients whose diagnoses were unknown, allowing us to determine the information content of each element of the vestibular test battery.

MAIN OUTCOMES AND MEASURES The information provided by each element of the vestibular test battery to segregate individuals with normal vestibular function from those with unilateral vestibular damage.

RESULTS The time constant calculated from the rotational test ranked first in information content, and measures that were related physiologically to the rotational time constant were 10 of the top 12 highest-ranked variables. The caloric canal paresis ranked eighth, and the other elements of the test battery provided minimal additional information. The sensitivity of the rotational time constant was 77.2%, and the sensitivity of the caloric canal paresis was 59.6%; the specificity of the rotational time constant was 89.0%, and the specificity of the caloric canal paresis was 64.9%. The diagnostic accuracy of the vestibular test battery increased from 72.4% to 93.4% when the data were analyzed with the optimal machine-learning classifier.

CONCLUSIONS AND RELEVANCE Rotational testing should be considered the primary test to diagnose unilateral peripheral vestibular damage in patients with dizziness or imbalance. Most physicians, however, continue to rely on caloric tests to guide their diagnoses. Our results support a significant shift in the approach used to determine diagnoses in patients with vestibular symptoms.

Author Affiliations: Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts (Priesol, Lewis); Department of Computer Science, Tufts University, Medford, Massachusetts (Cao, Brodley); now with the College of Computer and Information Science, Northeastern University, Boston, Massachusetts (Brodley); Department of Neurology, Harvard Medical School, Boston, Massachusetts (Lewis).

Corresponding Author: Richard F. Lewis, MD, Departments of Otolaryngology and Laryngology and of Neurology, Harvard Medical School, 243 Charles St, Boston, MA 02114 (richard_lewis@meei.harvard.edu).

izziness and imbalance are very common problems and are responsible for significant patient morbidity.[1] A crucial aspect of the diagnostic process is determining whether the peripheral vestibular system is damaged in one ear or is normal; the former situation suggests an otologic disorder (eg, Ménière disease) and the latter suggests an extralabyrinthine locus of dysfunction (eg, vestibular migraine). Quantitative vestibular testing is often used to supplement the clinical history and physical examination to help make this determination, but the accuracy of the different tests is not known and test results are frequently inconsistent or contradictory. The ideal way to address this problem would be to calculate how well each variable in the vestibular test battery segregates individuals with normal peripheral vestibular function from those with unilateral vestibular damage. This calculation requires a large population of patients who have accurate diagnoses based on clinical criteria but independent of vestibular testing. Unfortunately, accurate diagnoses cannot be determined in most patients with this approach, and prior studies[2-4] are inadequate because patient populations were small and diagnoses were based on subjective, imprecise criteria.

We addressed this problem by using machine-learning algorithms to help categorize patients. Because the 2 principal tests in clinical use (ie, caloric and rotational tests) only assess the function of the lateral canals and their afferent inner-

vation, our primary goal was to determine the optimal approach to categorize patients as having either normal or abnormal lateral canal function. Using a large set of vestibular test batteries (8080 people) performed in our clinical vestibular laboratory at the Massachusetts Eye and Ear Infirmary, we first examined the results of the 2 vestibular tests that are most commonly used to differentiate normal lateral canal function from unilateral lateral canal damage: the canal paresis derived from the caloric test, which quantifies the asymmetry in eye velocity when lateral canal activity in each ear is altered with a thermal stimulus,[5] and the time constant of the vestibulo-ocular reflex (VOR) derived from the rotational test, which quantifies the rate that eye velocity decays during constant-velocity yaw-axis head rotation.[6,7] Both of these measures typically remain abnormal after unilateral vestibular damage[5,7,8]; since the variability of these factors in the large normative set of data previously collected in our vestibular laboratory are known,[9] a quantitative, statistical cutoff point (eg, $P = .05$) can be defined in which the canal paresis and time constant values shift from normal to abnormal. We used these 2 well-characterized test results to segregate all of our patients into 3 categories (**Figure 1**A): unilateral (indicating a unilateral reduction in lateral canal function) if both results were abnormal, normal (indicating normal lateral canal function) if both results were normal, and undetermined if the 2 results conflicted. Patients identified as bilateral (eg, having bilateral

Figure 1. Patient Classification Based on the Time Constant and Canal Paresis Values
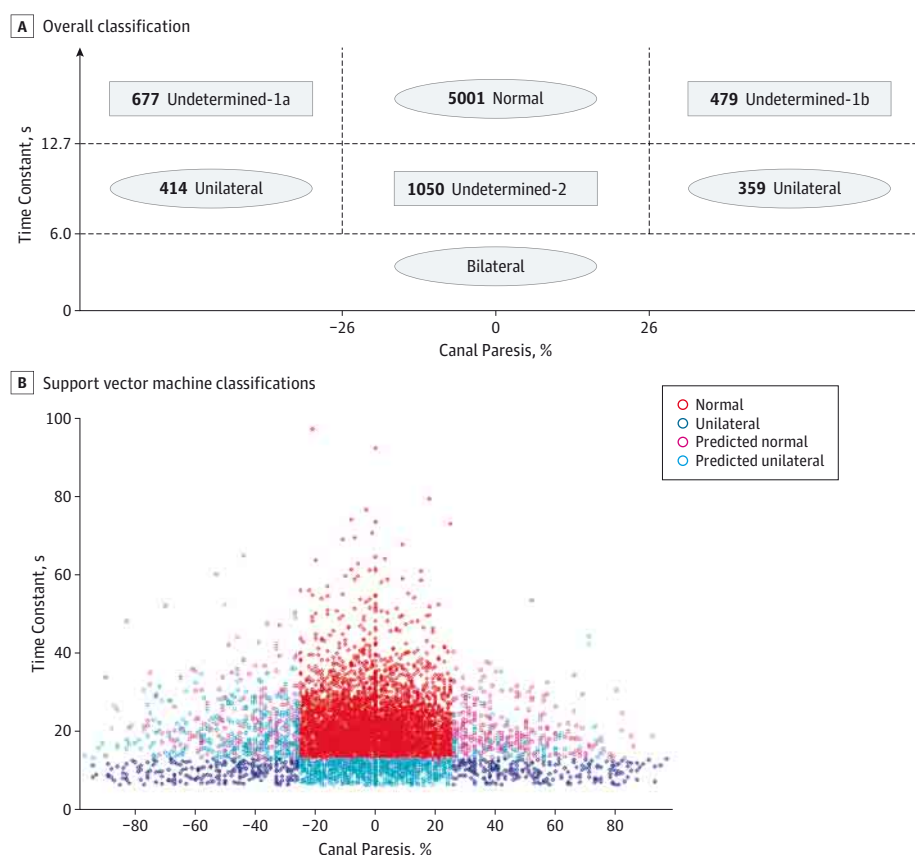
A, Patients were classified as having normal lateral canal function (normal time constant and canal paresis), unilateral reduction in lateral canal function (abnormal time constant and canal paresis), or undetermined function (conflicted time constant and canal paresis). B, Classifications generated by the support vector machine with radial basis kernel. These patients were divided into predicted-to-be-normal and predicted-to-be-unilateral categories. The other patients, classified by the congruence between canal paresis and time constant measurements, were designated as normal or unilateral.

vestibular deficits) in Figure 1A had time constants less than 6.0 seconds[8,10] and were therefore excluded from the data set we analyzed (see the Methods section).

If the canal paresis and time constant always agreed, then accurate diagnosis would be straightforward because the probability of patients being correctly labeled as having normal or unilateral peripheral vestibular hypofunction when both results concur is high. Unfortunately, the 2 results frequently conflict (eg, in approximately 30% of our patients), so providing accurate diagnoses for patients in the undetermined category becomes the crux of the diagnostic issue. This process has been problematic, however, because the vast majority of these undetermined patients cannot receive an accurate diagnosis using clinical criteria. We approached this problem by reviewing the medical records of the undetermined patients (while blinded to their vestibular test data) and manually labeled the relatively small subset of those who could receive an accurate diagnosis of either normal or unilateral peripheral vestibular hypofunction using strict objective criteria. We then used machine-learning classifiers to extrapolate quantitatively from the pattern of vestibular test data in these manually labeled patients to label the much larger population in the undetermined group who could not otherwise be diagnosed. Once all patients were labeled in this manner, the information content of each vestibular test variable was quantified.

Methods

Patient Selection

The Massachusetts Eye and Ear Infirmary institutional review board approved the study and waived the need for informed consent. We examined all data collected at our clinical vestibular laboratory during a decade, included patients who had caloric and rotational tests performed, and excluded patients with a rotational time constant of less than 6.0 seconds since these individuals most likely experienced bilateral vestibular damage (see the Discussion section).[8,10] The remaining data set consisted of 8080 patients. One hundred of these patients were given manual labels based on clinical criteria (see below), so the data set analyzed with machine-learning algorithms consisted of the remaining 7980 patients (Figure 1A). No clinical information was available for these patients other than their presenting symptom, which was invariably either dizziness or imbalance. The interval between symptom onset (which presumably dates the onset of the lesion) and the date vestibular testing was performed was unknown. Based on the referral patterns in the laboratory, these intervals were predicted to cover a large range (eg, between 4 weeks and 1 year). Overall, the canal paresis and time constant were congruent in 72.4% (labeled as *normal* or *unilateral* if both tests were normal or abnormal, respectively) and conflicted in 27.6% of the patients (labeled *undetermined*).

Vestibular Tests

We used a statistical measure based on the variance of the canal paresis and time constant in individuals with normal peripheral vestibular function tested in our laboratory[11,12] to de-

fine the cutoff point between normal and abnormal values. Bithermal caloric testing was performed using a closed-loop water stimulus at temperatures of 44°C or 27°C. The caloric canal paresis was calculated using a standard formula[9] from the asymmetry in eye velocity responses between the ears when warm and cool stimuli were applied to each ear separately; a value of 26% corresponds to a P value of .05 in our laboratory, so any canal paresis greater than or equal to 26% was defined as evidence of unilateral lateral canal damage (eg, a unilateral label). The time constant was calculated from sinusoidal earth-vertical axis rotational testing at 7 frequencies (0.01, 0.02, 0.05, 0.1, 0.2, 0.5, and 1.0 Hz) and a peak velocity of 50° per second using standard methods was defined as $1/(2 \cdot \pi \cdot f_c)$, where f_c is the corner frequency (eg, the frequency where the phase lead is 45°)[6,11] and becomes shortened when damage occurs in the vestibular periphery; a value of 12.7 seconds corresponds to a P value of .05 in our laboratory. Thus, values that were 12.7 seconds or lower were evidence of unilateral lateral canal damage (eg, a unilateral label; individuals with bilateral damage, evidenced by a time constant of <6 seconds, were previously excluded). Because the time constant changes with age, we applied an age adjustment when calculating each participant's value.[11] Aside from the caloric canal paresis and rotational time constant, many other variables were available for analysis (eTable 1 in the Supplement), ranging from 67 if only the caloric and rotational test were performed to 99 if all tests were performed.

Manual Labeling of Undetermined Patients

To obtain a "ground truth" within the undetermined category, medical records of undetermined patients were examined by 2 authors (A.J.P. and R.F.L.), and 100 patients were found who could be accurately labeled as normal or unilateral based on clinical information that was independent of the vestibular test variables to which the authors were blinded. Most reviewed records could not be accurately labeled using clinical criteria. Patients were classified as unilateral if they had at least 2 objective findings indicative of lateral canal damage in the same inner ear. These findings could include an abnormal head thrust test, spontaneous nystagmus, or head-shaking nystagmus[13] observed on clinical examination. Patients were classified as normal only if they had no evidence of lateral canal damage in either inner ear (using the above-described criteria) and additionally had a highly probable diagnosis that explained their symptoms (eg, benign positional vertigo or vestibular migraine) and was not usually associated with lateral canal damage.[14,15]

Machine-Learning Algorithms

Supervised machine-learning methods form a classifier from labeled training data wherein each training data point is represented by a set of features. In our application, the labels were *normal* and *unilateral,* and the vestibular test results on the manually labeled individuals were the training data. The accuracy of the algorithms was tested with standard cross-validation methods.[16] Because no one learning algorithm is best for all data sets, we applied several of the more common methods (decision trees, random forests, logistic regression,

Table 1. Borda Rankings of Test Variables

Test Component	Borda Ranking	Test Component	Borda Ranking
VOR time constant	40	VOR phase lead, 0.2 Hz	13
VOR phase lead		VOR gain, 0.02 Hz	13
0.05 Hz	35	Nystagmus slow-phase velocity, primary position, eyes closed	12
0.02 Hz	34	Nystagmus direction, head and body left	11
0.01 Hz	33	Caloric directional preponderance	10
VOR gain, 0.01 Hz	32	Nystagmus slow-phase velocity, right gaze, eyes closed	10
VOR phase lead, 0.1 Hz	28	Sum of the 4 caloric responses	9
Right cool caloric response	22	Pursuit description, 0.2 Hz	9
Caloric canal paresis	19	VOR gain, 0.1 Hz	8
VOR phase lead		Response right Hallpike maneuver	7
0.5 Hz	16		
1.0 Hz	13		

Abbreviation: VOR, vestibulo-ocular reflex.

AdaBoost applied with decision trees, and support vector machines [SVMs]).[17-19] We used 4 data sets to train the classifiers—the 100 manually labeled patients or these patients plus the 5774 individuals labeled as normal or unilateral based on the canal paresis/time constant congruence (Figure 1A)—with the canal paresis and time constant values included in the data or withheld. For our data set, SVMs performed the best (see the Results section).

Statistical Analysis
The predictive power of the classifiers was compared with random classification using the McNemar test.[20] The distribution of patients with weak caloric responses was compared between the normal and unilateral categories using the Mann-Whitney test.

Results

Manually Labeled Subjects
Fifty-nine undetermined-1 subjects (abnormal canal paresis, normal time constant) were labeled manually: 31 as normal, and 28 as unilateral. Forty-one undetermined-2 patients (abnormal time constant, normal canal paresis) were labeled manually: 15 as normal and 26 as unilateral. Although we included individuals whose net caloric response (eg, the sum of warm and cool responses from both ears) was less than 20° per second, which may be a less reliable measure than stronger caloric responses, we found that the distribution of subjects with weaker total responses did not differ significantly ($P = .26$, Mann-Whitey test) between the normal and unilateral categories; therefore, their inclusion did not bias the analysis.

Accuracy of the Classifiers
The classifiers trained on only manually labeled data were generally better than those trained on data from all patients, but the accuracy for both approaches was similar. The classifiers trained on all data were generally less accurate than those trained on data that withheld the time constant and canal paresis. Because all manually labeled patients were under-

mined (the time constant and canal paresis conflicted), it is logical that removing these conflicting variables would improve the classifier's accuracy. An SVM with a radial basis kernel[19] trained on manually labeled data in which the canal paresis and time constant were excluded performed best, with an accuracy of 76%. The McNemar test,[20] which determines whether 2 classifiers have the same predictive power, indicated that the accuracy of the best classifier was significantly better than random ($P < .001$). The other classifiers yielded accuracies ranging from 57% to 75% (eTable 2 in the Supplement).

Applying the Best Classifier to All Undetermined Patients
The canal paresis and time constant values (without machine learning) allow labeling of 72.4% of the patient population, with 62.7% categorized as normal (normal canal paresis and time constant) and 9.7% as unilateral (abnormal canal paresis and time constant). Assuming that the 100 manually labeled patients approximate a random sample of the undetermined group, we can apply the 76% accuracy of the SVM classifier to the entire undetermined population, which implies that 93.4% of all patients could be correctly labeled using this approach. The SVM predicted that 61% of the undetermined patients were unilateral and 39% were normal. Overall, this analysis indicates that 26.5% of the entire population had unilateral lateral canal hypofunction.

Information Content of the Different Test Elements
To examine the relative contributions of all 99 features in the data, we applied 2 approaches.[21] We ranked the features with respect to the mutual information[22] between the feature and the class label and also ranked the features by the magnitude of the weight of a linear SVM. **Table 1** reports the 20 data elements with the highest Borda rankings,[23] which combines both methods. The value of the time constant was the most important variable for determining whether a patient is labeled as unilateral or normal, and the value of the canal paresis was the eighth most important variable. Five of the 6 variables ranked between 2 and 7 were VOR features that are closely related to the time constant via velocity storage (see the Discussion section); the second through fourth most accurate measures were

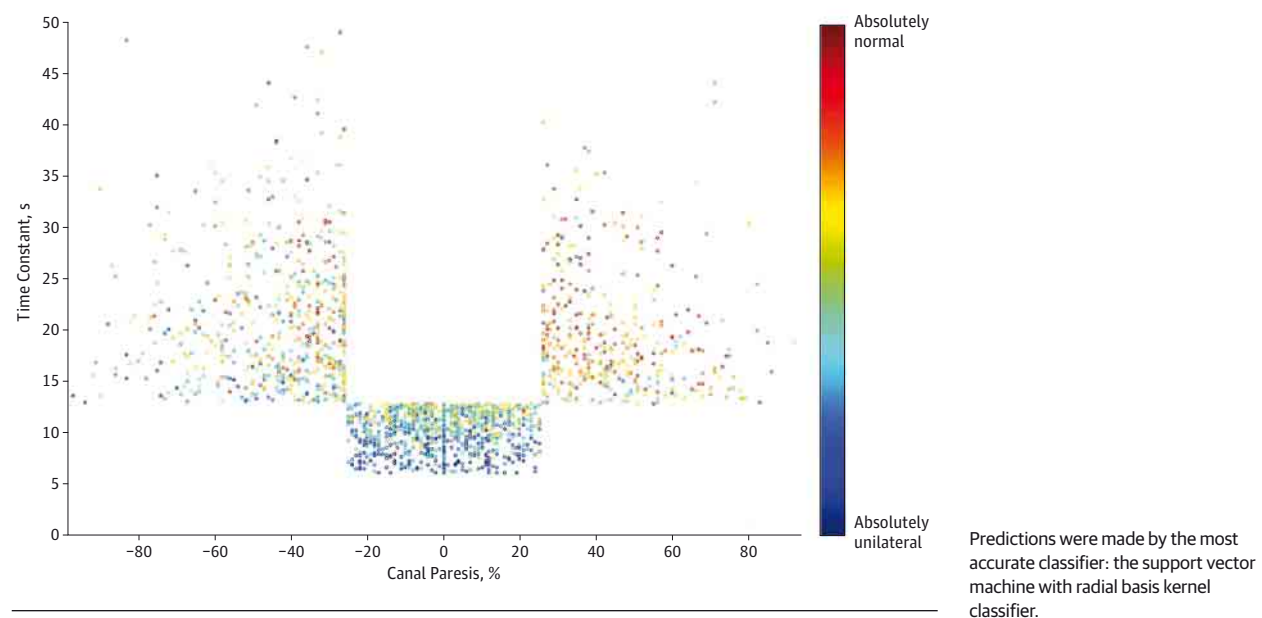Figure 2. Color-Coded Probability for Each Patient's Classification

Predictions were made by the most accurate classifier: the support vector machine with radial basis kernel classifier.

Table 2. Sensitivity and Specificity of the Canal Paresis and Time Constant

Characteristic	Findings, %	
	Abnormal	Normal
Test Sensitivity		
Caloric canal paresis		
Patient unilateral	59.6	40.4
Patient normal	11.5	88.5
Rotational time constant		
Patient unilateral	77.2	22.8
Patient normal	3.4	96.6
Test Specificity	**Unilateral**	**Normal**
Caloric		
Abnormal	64.9	35.1
Normal	14.1	85.9
Time constant		
Abnormal	89.0	11.0
Normal	7.8	92.2

the VOR low-frequency phases, which were expected to have accuracy similar to that of the VOR time constant since the time constant is derived from these phase values. Of the 12 variables ranked beneath the canal paresis (variables 9-20), 5 were VOR features and 2 were caloric factors (directional preponderance ranked 15th, and the sum of the 4 caloric responses ranked 17th).

Sensitivity and Specificity of the Canal Paresis and Time Constant
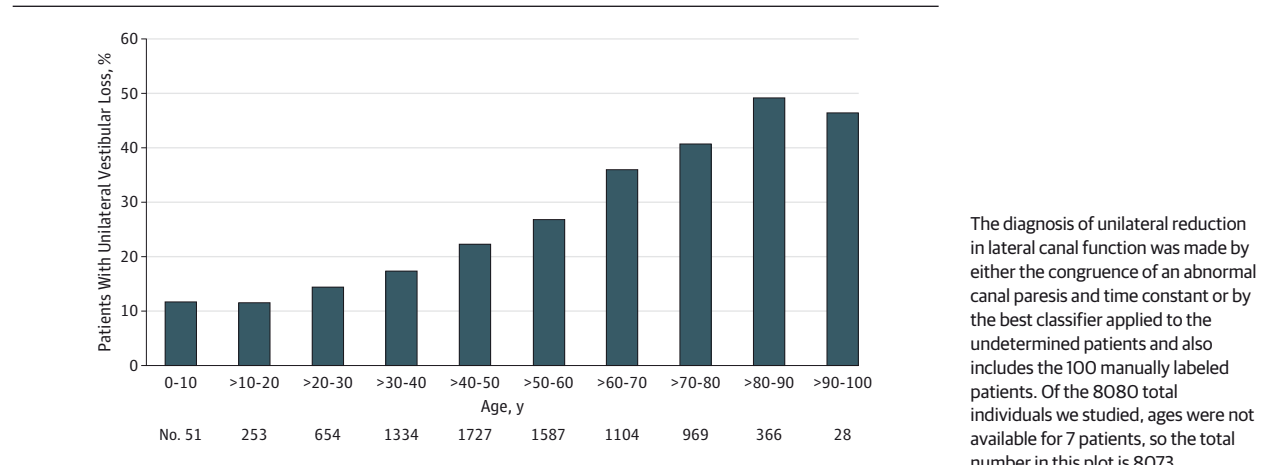
Figure 1B illustrates the SVM classifier's predictions for the undetermined subjects as well as for the normal and unilateral groups defined by the canal paresis–time constant congruence. In the undetermined group, when the time constant was

abnormal (undetermined-2), 81% of the patients were labeled unilateral, but when the canal paresis was abnormal (undetermined-1), only 29.2% were labeled unilateral. These results provide strong evidence that when canal paresis and time constant conflict, the time constant is a much more reliable indicator of the correct patient classification than is the canal paresis. Furthermore, there is a strong trend in both undetermined groups toward the unilateral label as the time constant shortens, but a weaker trend is evident as the canal paresis increases (**Figure 2** and eTable 3 in the Supplement). The sensitivity of the canal paresis for diagnosing unilateral lateral canal loss is only 59.6%, but the sensitivity of the time constant is 77.2% (**Table 2**). Because random classification would be correct in 50% of the subjects, the time constant adds nearly 3 times more information to the sensitivity measure than does the canal paresis. In a similar manner, when the canal paresis is abnormal, the specificity of this result is only 64.9% (35.1% of individuals with a normal peripheral vestibular system are misclassified as abnormal), but the specificity of an abnormal time constant is 89.0% (11% of the normal group are misclassified as abnormal), so again the information added to random classification by the time constant is nearly 3 times that of the canal paresis.

Prevalence of Unilateral Lateral Canal Damage

Because all patients tested in our vestibular laboratory had symptoms of dizziness or imbalance, we were able to determine the prevalence of unilateral canal damage in patients with these symptoms. Overall, of the 7980 individuals whose data we analyzed, 1955 patients (24.5%) were classified as having unilateral reduction in vestibular (lateral canal) function; the other 75.5% were classified as having normal peripheral (lateral canal) function. Since most disorders that damage the labyrinthine or eighth cranial nerve do not spare the lateral canal (see the Discussion section), we can estimate from these re-

Figure 3. Distribution of Unilateral Diagnoses vs Age

The diagnosis of unilateral reduction in lateral canal function was made by either the congruence of an abnormal canal paresis and time constant or by the best classifier applied to the undetermined patients and also includes the 100 manually labeled patients. Of the 8080 total individuals we studied, ages were not available for 7 patients, so the total number in this plot is 8073.

sults that approximately one-fourth of people with dizziness or imbalance have unilateral inner ear damage that contributes to or is responsible for these symptoms, and three-fourths of patients appear to have an extralabyrinthine source for their symptoms. The prevalence of vestibular damage was not uniform across all ages (**Figure 3**); the percentage of individuals who had vestibular deficits increased monotonically up to the eighth to ninth decades of life and then decreased.

Discussion

There are 3 principal findings of the present study. First, the most informative variables in the test battery are measures derived from the rotational test. Second, the rotational time constant is a much more accurate way to diagnose unilateral vestibular damage than is the caloric canal paresis. Third, machine-learning analysis improves the diagnostic accuracy of clinical vestibular testing. Below we discuss the implications of these results as well the limitations that are inherent to the approach we used to examine this important clinical problem.

Accuracy of Patient Labels
For the 2206 patients in the undetermined group, labels were based on the SVM algorithm's classification, which was trained with manually labeled patients in the undetermined group. These manual labels were expected to be highly accurate because a unilateral manual label required 2 or more objective abnormalities in the same ear. All of these measures have high specificities,[24] so it was extremely unlikely that these individuals had normal peripheral vestibular function (eg, abnormal head thrust and head-shaking nystagmus indicates a 99.6% likelihood of a unilateral vestibular deficit). However, the sensitivity of these tests is lower,[24] so a manual label of normal also required a clear diagnosis that explained the patients' symptoms but was unassociated with peripheral vestibular damage.

The best classifier trained with these manually labeled data had an error rate of 24%, which is applicable to the classifications of the undetermined group. A larger group of patients

(5774 total) was labeled as normal or unilateral based on the congruence of normal or abnormal time constant and canal paresis values. There is a 5% probability that either the time constant or the canal paresis results were misclassified (based on the .05 *P* value used for each test). The likelihood, therefore, that a normal or unilateral label was erroneous is 10% (5% for each of the 2 tests). Combining a 10% error rate for 5774 patients with a 24% rate for 2206 patients yields an overall error rate for patient labels of 13.9%. It seems most probable that these errors were not biased but rather were distributed randomly; therefore, although the errors increase variability, they should not shift our results in a particular direction. Finally, because complete unilateral loss of lateral canal function results in a VOR time constant near 6.0 seconds,[8,10] we excluded individuals with time constants less than 6.0 seconds since they must have bilateral vestibular damage. If bilateral vestibular damage was mild, these people may not have been excluded, but since unilateral vestibular damage is much more common than bilateral damage (estimated prevalence of 170:1 calculated from previously published studies[25,26]), these misclassifications would be rare.

Diagnostic Information in the Vestibular Test Data
The rotational test clearly has the highest information content in the vestibular test battery when the objective is separating patients with normal lateral canal function from those with a unilateral lateral canal deficit. Eight of the top 10 variables were derived from rotational measurements and, following the VOR time constant, the most important variables were the low-frequency phase lead (used to calculate the time constant) and the low-frequency gain of the VOR. All of these measurements are primarily determined by the velocity storage integrator in the brain,[27] implying that the status of velocity storage is the most important physiological factor that identifies the integrity of the peripheral vestibular system. The caloric canal paresis ranked eighth in information content, and other important variables (ie, spontaneous, positional, and gaze-evoked nystagmus) all reflect an imbalance in vestibular tone that would occur with an acute or subacute unilateral vestibular deficit but may resolve during the process of ves-

tibular compensation. Finally, other caloric-based measurements (directional preponderance, sum of responses) and 1 eye movement measurement (pursuit classification) were among the top 20 variables.

Sensitivity and Specificity of Caloric and Rotational Tests

The older literature on the accuracy of different vestibular tests was limited by the assumption that the caloric test was the criterion standard, so the sensitivity and specificity of other vestibular tests were determined by comparing them with caloric results.[2,7] However, unless patients had surgically induced vestibular lesions,[8] these studies were marred by the absence of well-defined patient diagnoses. The small groups of patients studied with surgical lesions usually had severe peripheral deficits, so their results provide no information about the accuracy of vestibular testing when peripheral damage is milder. In several more recent studies,[3,4,28] clinical labels were assigned to relatively small populations of patients with dizziness, but these diagnoses were subjective and not well characterized, and the results of these studies were conflicting. Despite the absence of conclusive data, an expert panel[29] indicated that caloric tests were the most useful to conduct when evaluating patients for possible unilateral vestibular damage. This conclusion was based primarily on studies[2] that showed that many individuals with abnormal caloric tests had normal rotational results. Our results demonstrate, however, that most of these patients have normal peripheral vestibular function; that is, the abnormal caloric result is a false-positive. All of the vestibular test data we analyzed were acquired in our laboratory at the Massachusetts Eye and Ear Infirmary and therefore may differ subtly from results in other laboratories that do not use closed-loop caloric irrigation to calculate the canal paresis or sinusoidal rotation to calculate the time constant. However, our results convincingly demonstrate that to distinguish normal peripheral (lateral canal) function from unilateral vestibular (lateral canal) loss, the rotational test is almost 3 times as sensitive and specific as the caloric test.

Machine-Learning Algorithms Applied to Clinical Vestibular Data

No available diagnostic tests provide absolute evidence that peripheral vestibular function is normal or deficient. However, standard vestibular test batteries generate large volumes of data, but human analysts often have difficulty observing patterns within these extensive data sets. Machine-learning algorithms that are based on quantitative extrapolation of diagnostic labels from patients with clinically designated diagnoses to those who lack a clear diagnosis are an attractive way to assess and improve clinical vestibular diagnostics. Although our best classifier had an accuracy of 76%, which introduced variability into the labels it generated, applying machine learning to vestibular testing clearly improved the diagnostic accuracy of the test battery. Machine-learning approaches could be improved by including other types of vestibular test data as they become available (see below), such as head impulse tests (HITs)[30] and vestibular-evoked myogenic potentials,[31] which were not included in our analysis because of the relatively small size of our database for these newer tests.

Vestibular Disorders Associated With Dizziness and Imbalance

Twenty-five percent of the patients with dizziness or imbalance had unilateral peripheral (lateral canal) deficits that presumably contributed to or caused their symptoms, and the prevalence of vestibular damage varied in an age-dependent manner (Figure 3). It is reasonable to speculate that younger patients had a lower prevalence of peripheral vestibular damage because they were more prone to experience vestibular migraine, the most common cause of dizziness in children.[14] Many disorders that damage the inner ear, such as labyrinthitis, vestibular neuritis, or traumatic vestibular concussion, can be considered as essentially random events that are more likely to occur as one ages, and such disorders could underlie the monotonic increase in peripheral vestibular damage that occurs with aging and peaks in the eighth decade of life. Furthermore, chronic ear diseases, such as Meniere disease, are generally progressive, so vestibular damage is more likely to be observable in older patients. For patients older than 90 years, our data suggest that other disorders, many of which are presumably degenerative, become more heavily represented in the dizzy/unsteady population.

Limitations of the Study

Several issues must be considered when interpreting the results described above. The first of these issues involves the validity of the machine-learning approach that we used. At a superficial level, the method appears to be circular since vestibular test data were used to label the large undetermined group of subjects as normal or unilateral and the information content of the test data then was assessed by evaluating these same individuals. This approach is not circular, however, because the labels applied to the entire undetermined population were defined by the characteristics of the 100 manually labeled undetermined patients, and these manual labels were based on clinical data that excluded the vestibular test results (to which we were blinded). This was a several-step process in which the intermediate steps (training of the machine-learning classifier and applying the classifier to the entire undetermined group) used a mathematical template derived from the pattern of vestibular test variables. The vestibular data did not add information to the analysis but rather served as a conduit that allowed us to use the characteristics of the manually labeled patients (whose labels were not based on vestibular testing) to label the entire undetermined group.

This study focused on lateral canal function, which is the only suborgan in the vestibular labyrinth that is tested by the caloric and rotational tests. If lateral canal function is abnormal, the patient clearly has peripheral vestibular damage. However, the converse is not true since normal lateral canal function does not exclude possible damage in the vertical canals or the otolith organs. Therefore, some of the individuals who we labeled as normal based on the normalcy of their caloric and rotational test results may have had damage in other peripheral vestibular end organs, implying that our analysis may have overestimated the number of normal labels and underestimated the number of unilateral labels. We suggest, how-

ever, that the frequency of these errors was most likely very low; with a few uncommon exceptions, such as vestibular neuritis limited to the inferior vestibular nerve,[32] most labyrinthine and eighth cranial nerve disorders do not selectively spare the lateral canal.[33]

Another possible limitation of the study involves the effects of vestibular compensation. Our analysis shows that the rotational time constant is a more sensitive and specific measure than the caloric canal paresis when the aim is to identify unilateral vestibular (lateral canal) damage. However, rotational testing is not perfect; we found that 22.8% of the participants with unilateral vestibular damage had normal rotational time constants (Table 2, sensitivity of the rotational test). Because no information was available on the interval between symptom onset and performance of vestibular tests, it is likely that many patients were tested at least several months after symptoms began. This time frame would allow vestibular compensatory processes[34,35] to correct or suppress many of the abnormalities that would be present soon after labyrinthine damage occurred. The mixture of acute and chronic lesions and the effects of compensation on the latter group would have the effect of reducing the sensitivity of some test variables. However, the rotational time constant and caloric canal paresis usually remain abnormal for at least several years after peripheral damage,[34,35] and it is unlikely that many patients were tested this long after symptom onset. Vestibular compensation, therefore, should have a minimal effect on the accuracy of the rotational time constant or caloric canal paresis but would be expected to reduce the information content of findings that are known to resolve during compensation, such as spontaneous and gaze-evoked nystagmus.[34] In addition, the normal range for the time constant is large, and some patients in this 22.8% group may have had relatively long time constants before their peripheral vestibular insult (eg, as associated with migraine[36]) so that the peripheral lesion reduced the time constant while remaining above the normal cutoff of 12.7 seconds.

As noted above, we did not include vestibular-evoked myogenic potentials or HIT[31] data in our analysis since our database was too small for these newer tests. However, the HIT is of particular interest and deserves additional comment. The qualitative head-thrust test, which is the basis of the quantitative HIT, was described 27 years ago[37] and has been a standard part of the clinical vestibular examination for many years. Although the HIT will almost certainly improve the accuracy of vestibular diagnostics, it tests something different from the standard low-frequency rotational testing that we evaluated in this study. The low-frequency test focuses on velocity storage, quantified by the rotational time constant, and has been shown[7] to be very sensitive to damage in the peripheral vestibular system. The HIT, in contrast, measures the high-frequency amplitude (or gain) of the VOR response, and experience with the qualitative head-thrust test suggests that significant peripheral damage may be required before the gain becomes abnormal.[24] This result implies that the HIT is likely to be a very specific but possibly a less-sensitive way to assess peripheral vestibular function. Despite the current enthusiasm for the HIT, we suggest that it will not prove to be the perfect diagnostic tool for peripheral vestibular damage but rather will serve to complement the caloric and low-frequency rotational tests. Both the HIT and the vestibular-evoked myogenic potential test should be included in future iterative machine-learning approaches to vestibular diagnostics since the information they provide will almost certainly improve the accuracy of these computer algorithms.

In addition to providing evidence concerning peripheral vestibular damage, vestibular testing is useful to gauge the severity of peripheral damage and help determine which ear is abnormal. However, central lesions (eg, those that damage the vestibular nuclei and their interconnections or the vestibulo-cerebellum) rarely mimic the test results that are characteristic of peripheral vestibular damage. Furthermore, some inner ear disorders that cause dizziness are mechanical and are not associated with reduced peripheral vestibular function (eg, benign positional vertigo and superior canal dehiscence). Benign positional vertigo involving the posterior canal, in particular, causes positional nystagmus that is primarily torsional, and this nystagmus is best documented with direct visualization of the eyes. More generally, vestibular testing is most valuable when performed in the appropriate clinical context and serves to augment, rather than replace, a detailed history and physical examination.

Conclusions

Our results demonstrate that, by training classifiers using manually labeled patients, we can substantially improve the ability of vestibular test data to classify patients as having normal peripheral (lateral canal) function or damage in one vestibular labyrinth or its afferent innervation. The measures derived from the rotational test, particularly the VOR time constant, provide the most valuable information; the caloric tests are much less useful and should no longer be considered the criterion standard when the issue is the presence or absence of unilateral lateral canal damage. Because this question usually dominates the process of clinical diagnosis in patients with dizziness and imbalance, we recommend a significant paradigm shift in the field since most vestibular laboratories and physicians continue to rely almost exclusively on caloric testing.

REFERENCES

1. Agrawal Y, Carey JP, Della Santina CC, Schubert MC, Minor LB. Disorders of balance and vestibular function in US adults: data from the National Health and Nutrition Examination Survey, 2001-2004. *Arch Intern Med*. 2009;169(10):938-944.

2. Baloh RW, Sills AW, Honrubia V. Impulsive and sinusoidal rotatory testing: a comparison with results of caloric testing. *Laryngoscope*. 1979;89(4):646-654.

3. Ahmed MF, Goebel JA, Sinks BC. Caloric test versus rotational sinusoidal harmonic acceleration and step-velocity tests in patients with and without suspected peripheral vestibulopathy. *Otol Neurotol*. 2009;30(6):800-805.

4. Kaplan DM, Marais J, Ogawa T, Kraus M, Rutka JA, Bance ML. Does high-frequency pseudo-random rotational chair testing increase the diagnostic yield of the ENG caloric test in detecting bilateral vestibular loss in the dizzy patient? *Laryngoscope*. 2001;111(6):959-963.

5. Lieberman MC, Rosowski JJ, Lewis RF. Physiology and pathophysiology. In: Merchant S, Nadol JB, eds. *Schuknecht's Pathology of the Ear*. Shelton, CT: People's Medical Publishing House; 2012: 98-134.

6. Wall C III. The sinusoidal harmonic acceleration rotary chair test: theoretical and clinical basis. *Neurol Clin*. 1990;8(2):269-285.

7. Baloh RW, Honrubia V, Yee RD, Hess K. Changes in the human vestibulo-ocular reflex after loss of peripheral sensitivity. *Ann Neurol*. 1984;16(2):222-228.

8. Black FO, Shupert CL, Peterka RJ, Nashner LM. Effects of unilateral loss of vestibular function on the vestibulo-ocular reflex and postural control. *Ann Otol Rhinol Laryngol*. 1989;98(11):884-889.

9. Secretariat: Acoustical Society of America. *American National Standard on Procedures for Testing Basic Vestibular Function*. New York, NY: American National Standards Institute; 1999:S3.45.

10. Brantberg K, Fransson PA, Magnusson M, Johansson R, Bergenius J. Short vestibulo-ocular reflex time-constant in complete unilateral vestibular lesions. *Am J Otol*. 1995;16(6):787-792.

11. Dimitri PS, Wall C III, Oas JG. Classification of human rotation test results using parametric modeling and multivariate statistics. *Acta Otolaryngol*. 1996;116(4):497-506.

12. Dimitri PS, Wall C III, Rauch SD. Multivariate vestibular testing: thresholds for bilateral Ménière's disease and aminoglycoside ototoxicity. *J Vestib Res*. 2001-2002;11(6):391-404.

13. Asawavichianginda S, Fujimoto M, Mai M, Rutka J. Prevalence of head-shaking nystagmus in patients according to their diagnostic classification in a dizziness unit. *J Otolaryngol*. 1997;26(1):20-25.

14. Furman JM, Marcus DA, Balaban CD. Vestibular migraine: clinical aspects and pathophysiology. *Lancet Neurol*. 2013;12(7):706-715.

15. Parnes LS, Agrawal SK, Atlas J. Diagnosis and management of benign paroxysmal positional vertigo (BPPV). *CMAJ*. 2003;169(7):681-693.

16. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. San Francisco, CA: Morgan Kaufmann; 2005.

17. Quinlan JS. *C4.5: Programs for Machine Learning*. Burlington, MA: Morgan Kaufmann; 1993.

18. Breiman L. Random forests. *Mach Learn*. 2001; 45:5-32.

19. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, England: Cambridge University Press; 2000.

20. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12(2):153-157.

21. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;3: 1157-1182.

22. Cover TM, Thomas JA. *Elements of Information Theory*. New York, NY: John Wiley & Sons; 2012.

23. Lin S. Rank aggregation methods. *Comput Stat*. 2010;2(5):555-570.

24. Harvey SA, Wood DJ, Feroah TR. Relationship of the head impulse test and head-shake nystagmus in reference to caloric testing. *Am J Otol*. 1997;18(2):207-213.

25. Neuhauser HK. Epidemiology of vertigo. *Curr Opin Neurol*. 2007;20(1):40-46.

26. Ward BK, Agrawal Y, Hoffman HJ, Carey JP, Della Santina CC. Prevalence and impact of bilateral vestibular hypofunction: results from the 2008 US National Health Interview Survey. *JAMA Otolaryngol Head Neck Surg*. 2013;139(8):803-810.

27. Raphan T, Cohen B. Velocity storage and the ocular response to multidimensional vestibular stimuli. *Rev Oculomot Res*. 1985;1:123-143.

28. Arriaga MA, Chen DA, Cenci KA. Rotational chair (ROTO) instead of electronystagmography (ENG) as the primary vestibular test. *Otolaryngol Head Neck Surg*. 2005;133(3):329-333.

29. Fife TD, Tusa RJ, Furman JM, et al. Assessment: vestibular testing techniques in adults and children: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology*. 2000;55(10): 1431-1441.

30. MacDougall HG, Weber KP, McGarvie LA, Halmagyi GM, Curthoys IS. The video head impulse test: diagnostic accuracy in peripheral vestibulopathy. *Neurology*. 2009;73(14):1134-1141.

31. Shin BS, Oh SY, Kim JS, et al. Cervical and ocular vestibular-evoked myogenic potentials in acute vestibular neuritis. *Clin Neurophysiol*. 2012;123(2): 369-375.

32. Chihara Y, Iwasaki S, Murofushi T, et al. Clinical characteristics of inferior vestibular neuritis. *Acta Otolaryngol*. 2012;132(12):1288-1294.

33. Rauch SD. Vestibular histopathology of the human temporal bone: what can we learn? *Ann N Y Acad Sci*. 2001;942:25-33.

34. Deveze A, Bernard-Demanze L, Xavier F, Lavieille JP, Elziere M. Vestibular compensation and vestibular rehabilitation: current concepts and new trends. *Neurophysiol Clin*. 2014;44(1):49-57.

35. Fetter M, Zee DS. Recovery from unilateral labyrinthectomy in rhesus monkey. *J Neurophysiol*. 1988;59(2):370-393.

36. Jeong SH, Oh SY, Kim HJ, Koo JW, Kim JS. Vestibular dysfunction in migraine: effects of associated vertigo and motion sickness. *J Neurol*. 2010;257(6):905-912.

37. Halmagyi GM, Curthoys IS. A clinical sign of canal paresis. *Arch Neurol*. 1988;45(7):737-739.