

# Evaluation and construction of diagnostic criteria for inclusion body myositis



Thomas E. Lloyd, MD,  
PhD  
Andrew L. Mammen,  
MD, PhD  
Anthony A. Amato, MD  
Michael D. Weiss, MD  
Merrilee Needham,  
MBBS  
Steven A. Greenberg, MD

Correspondence to  
Dr. Lloyd:  
tlloyd4@jhmi.edu  
or Dr. Greenberg:  
sagreenberg@partners.org

## ABSTRACT

**Objective:** To use patient data to evaluate and construct diagnostic criteria for inclusion body myositis (IBM), a progressive disease of skeletal muscle.

**Methods:** The literature was reviewed to identify all previously proposed IBM diagnostic criteria. These criteria were applied through medical records review to 200 patients diagnosed as having IBM and 171 patients diagnosed as having a muscle disease other than IBM by neuromuscular specialists at 2 institutions, and to a validating set of 66 additional patients with IBM from 2 other institutions. Machine learning techniques were used for unbiased construction of diagnostic criteria.

**Results:** Twenty-four previously proposed IBM diagnostic categories were identified. Twelve categories all performed with high ( $\geq 97\%$ ) specificity but varied substantially in their sensitivities (11%–84%). The best performing category was European Neuromuscular Centre 2013 probable (sensitivity of 84%). Specialized pathologic features and newly introduced strength criteria (comparative knee extension/hip flexion strength) performed poorly. Unbiased data-directed analysis of 20 features in 371 patients resulted in construction of higher-performing data-derived diagnostic criteria (90% sensitivity and 96% specificity).

**Conclusions:** Published expert consensus-derived IBM diagnostic categories have uniformly high specificity but wide-ranging sensitivities. High-performing IBM diagnostic category criteria can be developed directly from principled unbiased analysis of patient data.

**Classification of evidence:** This study provides Class II evidence that published expert consensus-derived IBM diagnostic categories accurately distinguish IBM from other muscle disease with high specificity but wide-ranging sensitivities. *Neurology*® 2014;83:426-433

## GLOSSARY

**ANRI** = Australian Neuromuscular Research Institute; **BWH** = Brigham and Women's Hospital; **CD** = clinically defined; **CPD** = clinicopathologically defined; **ENMC** = European Neuromuscular Centre; **IBM** = inclusion body myositis; **IM** = Institute of Myology; **JHMI** = Johns Hopkins Medical Institutions; **MHC-1** = major histocompatibility complex class I; **MMT** = manual muscle testing; **MRC** = Medical Research Council; **PD** = pathologically defined; **UWASH** = University of Washington Seattle.

Sporadic inclusion body myositis (IBM) is a progressive autoimmune and degenerative disorder of muscle of unknown cause. Research diagnostic categories for IBM are important in the reporting of research on cohorts of patients with IBM and are essential for the conduct of IBM clinical therapeutic trials. Formal research diagnostic criteria and categories for IBM have been proposed since 1987 by individual authors<sup>1-4</sup> and through publications of consensus expert opinions developed in 5 meetings in 1995,<sup>5</sup> 1996,<sup>6,7</sup> 2008,<sup>8</sup> 2009,<sup>9</sup> and 2011.<sup>10</sup> These publications have created features (e.g., “presence of quadriceps weakness”), categories (Boolean algebraic combinations of features; e.g., definite), and schemes (combinations of categories; e.g., definite, probable, and possible). All of these consensus features, categories, and schemes were reported in the absence of published data on their sensitivity and specificity for IBM among muscle diseases.

Supplemental data  
at [Neurology.org](http://Neurology.org)

From the Departments of Neurology (T.E.L., A.L.M.), Neuroscience (T.E.L.), and Medicine (A.L.M.), Johns Hopkins University School of Medicine and Johns Hopkins Bayview Myositis Center, Baltimore, MD; Department of Neurology (A.A.A., S.A.G.), Brigham and Women's Hospital and Harvard Medical School, Boston, MA; Department of Neurology (M.D.W.), University of Washington, Seattle; Department of Neurology (M.N.), Australian Neuromuscular Research Institute, University of Western Australia; and Children's Hospital Informatics Program (S.A.G.), Boston Children's Hospital and Harvard-MIT Division of Health Sciences and Technology, Boston, MA.

Go to [Neurology.org](http://Neurology.org) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

To understand the basis and applicability of existing IBM diagnostic criteria and categories, we separately studied the performance of features, categories, and schemes in a large cohort of patients diagnosed with IBM by their treating physicians and in patients diagnosed with muscle diseases other than IBM at 4 different institutions. In addition, we demonstrate a method for developing empirical data-derived IBM diagnostic criteria.

**METHODS** We examined the literature for all published IBM diagnostic schemes, separating features and categories. We retrospectively reviewed medical records of 200 patients diagnosed during clinical care as definitely having IBM and of 171 patients diagnosed as definitely having a muscle disease other than IBM by neuromuscular specialists at Brigham and Women's Hospital (BWH; n = 182) and Johns Hopkins Medical Institutions (JHMI; n = 189). No uniform criteria for diagnosis had been systematically applied; clinicians made these diagnoses in the course of ordinary clinical practice. Strength criteria were assessed according to Medical Research Council (MRC) manual muscle testing (MMT) scores. Strength criteria were interpreted as satisfied if they were met on one or both sides of the body. No separate review of muscle pathologic slides or additional performance of studies other than what had been performed for clinical care and recorded in medical records was performed. For the research question pertaining to diagnostic sensitivity and specificity of various criteria against the gold standard of clinical diagnosis by treating clinicians, level of evidence Class II applies because this is a case-control study of a broad spectrum of persons with and without the disease, and clinical diagnosis was determined without reference to criteria sets.

Non-IBM diagnoses (appendix e-1 on the *Neurology*<sup>®</sup> Web site at [Neurology.org](http://Neurology.org)) included inflammatory myopathies (n = 87; 51%) and inherited myopathies (n = 43; 25%), of which 20 patients had myopathies with prominent rimmed vacuoles (e.g., hereditary inclusion body myopathies or myofibrillar myopathy). Medical records from 2 other cohorts of patients diagnosed as having IBM by neuromuscular specialists at the Australian Neuromuscular Research Institute (ANRI; n = 42) and University of Washington Seattle (UWASH; n = 24) had more limited review, confined to MMT scores in selected muscles. Machine learning algorithms were implemented using the software package Orange (appendix e-2).<sup>11</sup>

**Standard protocol approvals, registrations, and patient consents.** All patient data were studied under protocols approved by institutional review boards.

**RESULTS** **Applicability of existing IBM research diagnostic criteria to patients.** We studied the applicability of 10 different schemes of IBM diagnostic criteria<sup>1–10</sup> comprising 24 categories. We found that 12 categories could be applied to patients (table), while 12 other categories were not further studied for the following reasons. Two European Neuromuscular Centre (ENMC 1997 probable and ENMC 1997 definite) categories<sup>6</sup> were subsequently replaced with revised versions in the year 2000<sup>7</sup>; only the revised categories were analyzed here. Four categories lacked

specificity for IBM and were not further analyzed: 3 1987 definite, probable, and possible<sup>1</sup> categories lack a requirement for the microscopic detection of inflammatory cells in muscle biopsy specimens, allowing patients with genetic disorders with rimmed vacuoles to be categorized as IBM; and a 1991 criteria set<sup>2</sup> indicated that filamentous inclusions were pathognomonic of IBM, although these were known to be present in a number of hereditary distal myopathies.<sup>12</sup> The category Griggs probable was of uncertain existence, having been reported in a single review article where it had been referred to in text as Griggs possible.<sup>4</sup> Five categories have criteria we were unable to unambiguously apply to patients for the following reasons: ENMC 1997 possible (uncertainty about application of the required criterion: “IBM is most often a disease of middle-aged or elderly men”)<sup>6</sup>; Western Australia 2002 definite, probable, and possible (uncertainty about minimal pathologic features required)<sup>3</sup>; and Institute of Myology 2010 pathologically defined (circularly defined clinical criteria of “consistent with IBM”).<sup>9</sup>

#### **Sensitivity and specificity of IBM categories and features.**

Next, we determined the sensitivity (proportion of 200 patients diagnosed with IBM meeting diagnostic category criteria) and specificity (proportion of 171 patients diagnosed as not having IBM not meeting IBM diagnostic category criteria) of these 12 diagnostic categories from 5 different schemes (figure 1, A and B). All categories had very high specificity (range 98%–100%) but differed substantially in sensitivities (range 11%–84%). The best performing categories were ENMC 2013 probable and Institute of Myology (IM) 2010 possible (sensitivities of 84% and 77%, respectively). The least sensitive categories were Griggs definite/MRC 2010 PD (whose criteria are identical) and ENMC 2013 clinicopathologically defined (CPD) (sensitivities of 11% and 15%, respectively). Only 45% of patients fulfilled either the MRC 2010 PD or clinically defined (CD) categories, and 60% fulfilled either the ENMC 2013 CPD or CD categories that evolved from refinements in the MRC 2010 scheme.

To understand the substantial differences in the sensitivities of these categories, we studied the sensitivity and specificity of the 25 individual features used among these categories (figure 1, C and D). Features pertaining to demographics (e.g., age at onset), temporal course, and serum creatine kinase limits generally had high sensitivities but very low specificities. In contrast, certain strength and pathologic criteria had high specificities but very low sensitivities. The least sensitive strength feature “weakness of knee extension greater than hip flexion,” a requirement for the MRC 2010 CD and IM 2010 CD categories, had 47% sensitivity. Similarly, certain pathologic features were also highly specific but of very low sensitivity (e.g.,

**Table IBM diagnostic categories and underlying feature criteria**

Category	Feature criteria
Griggs definite <sup>5</sup>	AND(14a,15,OR(16,17))
Griggs possible <sup>5</sup>	AND(1a,3a,5a,6,13,14a,OR(10,11,12),NOT(OR(15,16,17)))
ENMC 2000 definite <sup>7</sup>	AND(2,4,14a,15,OR(10,11,16))
ENMC 2000 probable <sup>7</sup>	AND(2,4,14a,OR(10,11,15))
MRC 2010 PD <sup>8</sup>	AND(14a,15,OR(16,17))
MRC 2010 CD <sup>8</sup>	AND(1b,3b,7,9,OR(14a,15,21),NOT(OR(16,17)))
MRC 2010 possible <sup>8</sup>	AND(1b,3b,OR(7,9),OR(14a,15,21),NOT(OR(16,17)))
IM 2010 CD <sup>9</sup>	AND(1a,3a,6,7,9,OR(14b,21),NOT(OR(16,17,18)))
IM 2010 possible <sup>9</sup>	AND(1a,3a,6,OR(7,9),OR(14b,21),NOT(OR(16,17,18)))
ENMC 2013 CPD <sup>10</sup>	AND(1c,3b,5b,14b,15,21,OR(8,9),OR(16,17,18,19,20))
ENMC 2013 CD <sup>10</sup>	AND(1c,3b,5b,8,9,OR(14b,15,16,17,18,19,20,21),NOT(AND(14b,15,21,OR(16,17,18,19,20))))
ENMC 2013 probable <sup>10</sup>	AND(1c,3b,5b,OR(8,9),OR(14b,15,16,17,18,19,20,21),NOT(AND(14b,15,21,OR(16,17,18,19,20))))
<b>Features</b>	
<b>Demographic</b>	<b>Strength</b>
1. Age lower limit	7. KE weaker than HF
a. Age >30 y	8. KE equal to or weaker than HF
b. Age >35 y	9. FF weaker than SA
c. Age >45 y	10. WF weaker than WE
2. Sporadic	11. FF weakness
<b>Temporal aspect</b>	12. Quads ≤MRC 4
3. Duration minimum	13. Proximal/distal weakness arms/legs
a. >6 mo	<b>Muscle pathology</b>
b. >12 mo	14. Inflammation
4. Slow progression	a. Invasion of nonnecrotic fibers
<b>Laboratory studies</b>	b. Endomysial inflammation
5. Serum CK maximum	15. RVs
a. 12× ULN	16. EM tubulofilaments
b. 15× ULN	17. Amyloid
6. EMG consistent	18. SMI31
	19. p62
	20. TDP-43
	21. MHC-1

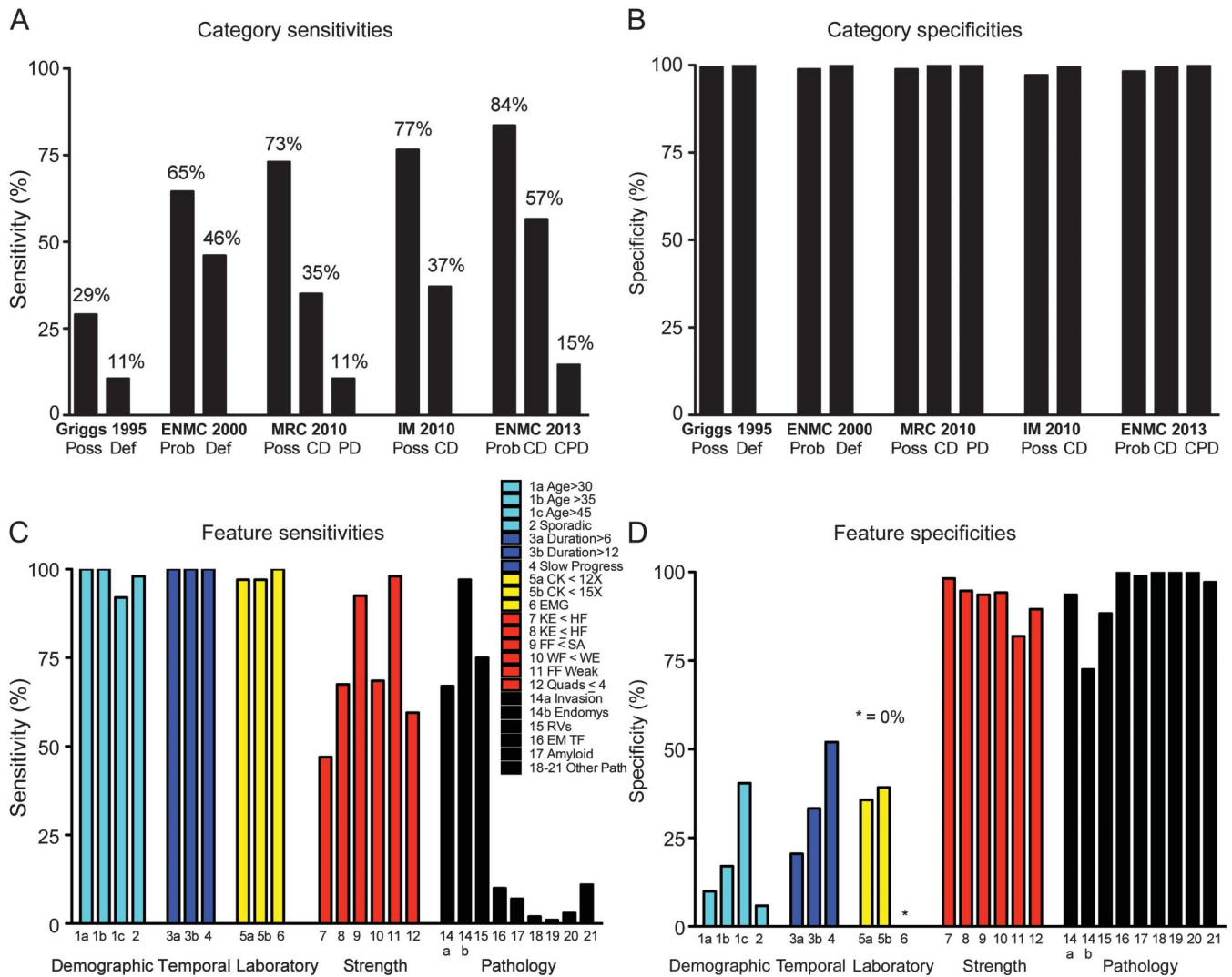
Abbreviations: CD = clinically defined; CK = creatine kinase; CPD = clinicopathologically defined; EM = electron microscopy; ENMC = European Neuromuscular Center; FF = finger flexion; HF = hip flexion; IBM = inclusion body myositis; IM = Institute of Myology; KE = knee extension; MHC-1 = major histocompatibility complex class I; MRC = Medical Research Council; PD = pathologically defined; RV = rimmed vacuole; SA = shoulder abduction; SMI31 = phosphorylated neurofilament antigen; TDP-43 = TAR DNA-binding protein-43; ULN = upper limit of normal; WE = wrist extension; WF = wrist flexion. Categories analyzed and their criteria. Years (2000, 2010, and 2013) refer to the year of publication rather than year of conference/workshop. For example, MRC 2010 CD category was published in 2010 and requires features 1b and 3b and 7 and 9 and either (14a, 15, or 21) and not (16 or 17).

the requirement for electron microscopy tubulofilaments had only 10% sensitivity).

**Category failure rates and reasons.** To understand whether the low sensitivities of certain categories were attributable to the low sensitivities of certain features,

we examined the specific features that resulted in failure to meet specific category criteria. This enabled us to determine the proportion of patients failing to meet that category's requirements (failure rates) and the single most common feature responsible for that failure (figure 2A).

**Figure 1** Sensitivities and specificities of IBM categories and features



(A, B) IBM categories. (C, D) IBM features. CD = clinically defined; CK = creatine kinase; CPD = clinicopathologically defined; Def = definite; EM = electron microscopy; ENMC = European Neuromuscular Center; FF = finger flexion; HF = hip flexion; IBM = inclusion body myositis; IM = Institute of Myology; KE = knee extension; MRC = Medical Research Council; PD = pathologically defined; Poss = possible; Prob = probable; RV = rimmed vacuoles; SA = shoulder abduction; TF = tubulofilaments; WE = wrist extension; WF = wrist flexion.

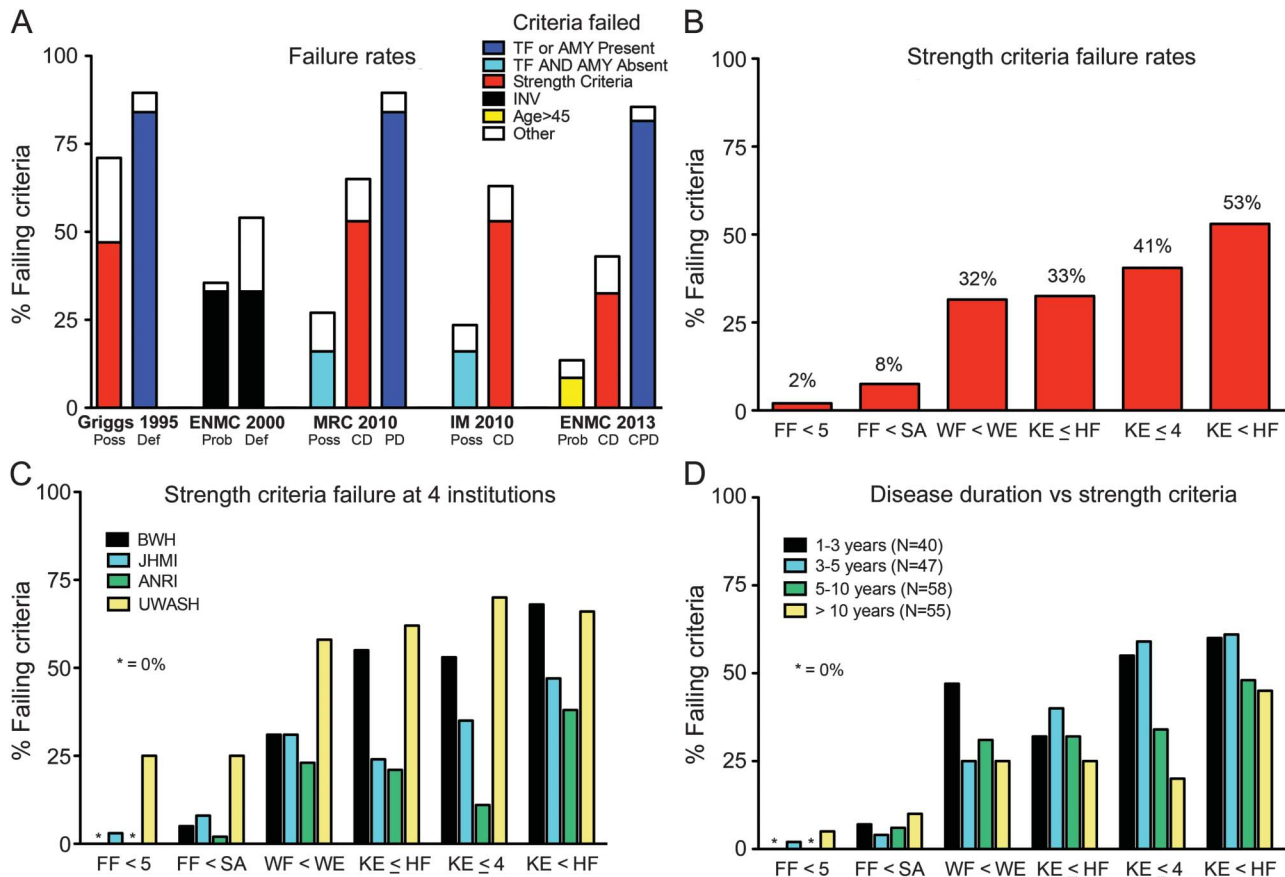
Certain categories performed poorly in part not because patients lacked typical features of IBM but paradoxically because patients indeed had such features. For example, the MRC 2010 CD category (failure rate of 65%), as well as the MRC 2010 possible, IM 2010 possible, and IM 2010 CD categories, exclude patients if they have tubulofilaments or amyloid present. The ENMC 2013 CD and ENMC 2013 probable categories exclude patients for having too many pathologic features, although such patients instead meet ENMC 2013 CPD criteria.

Certain categories rely on specialized pathologic features, often not performed or performed and found to be absent, even at tertiary referral centers. The Griggs definite/MRC 2010 PD categories (failure rates 89%), and ENMC 2013 CPD (failure rate 85%) category both rely on low sensitivity features, such as electron

microscopy–demonstrated tubulofilaments, or Congo red–demonstrated amyloid. Electron microscopy was performed in 24% of our patients with IBM and showed tubulofilaments in only 35% of those in whom it was done. Congo red staining was performed in 36% of our patients and revealed amyloid in only 19% of those in whom it was done.

Other categories performed poorly because of the low sensitivity of certain strength criteria. The best performing strength criteria was “finger flexion weakness” (2% failure rate) and the worst performing was “weakness of knee extension greater than hip flexion” (53% failure rate; figure 2B). The Griggs possible criterion of “proximal and distal weakness of arms and legs” was also quite limiting (47% failure rate). The recently proposed categories of MRC 2010 CD (65% failure rate) and IM 2010 CD (63% failure rate) both fail

**Figure 2** Failure rates and reasons for IBM diagnostic criteria



(A) Failure rates for published IBM diagnostic categories according to most common criteria failed. (B) Failure rates of strength criteria. Ambiguity results from meeting strength criteria on one side of the body but failing on the other side. (C) Consistency of strength criteria failure rates at 4 institutions. The criterion of “finger flexion weakness” (FF < 5) performed best and that of “weakness of knee extension greater than hip flexion” performed worst at all of 4 different institutions. (D) Failure rates by disease duration. Disease duration has little impact on finger flexion criteria failure but a consistent impact on all knee extension-related criteria, with greater failure rates for 1–3 and 3–5 years than 5–10 and >10 years for KE ≤ HF, KE < HF, and KE ≤ 4. AMY = amyloid; ANRI = Australian Neuromuscular Research Institute; BWH = Brigham and Women’s Hospital; CD = clinically defined; CPD = clinicopathologically defined; Def = definite; ENMC = European Neuromuscular Center; FF = finger flexion; HF = hip flexion; IBM = inclusion body myositis; IM = Institute of Myology; INV = invasion of non-necrotic muscle fibers; JHMI = Johns Hopkins Medical Institutions; KE = knee extension; MRC = Medical Research Council; PD = pathologically defined; Poss = possible; Prob = probable; SA = shoulder abduction; TF = tubulofilaments; UWASH = University of Washington Seattle; WE = wrist extension; WF = wrist flexion.

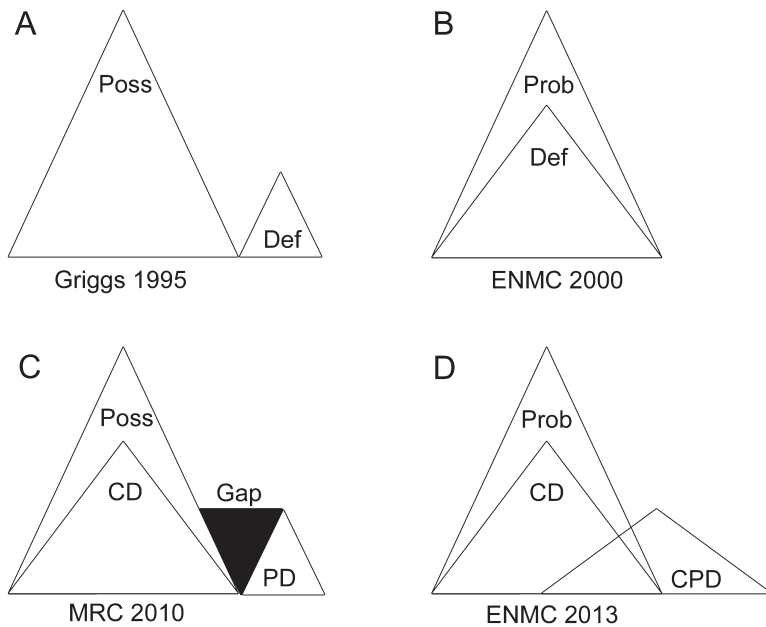
primarily because of the “weakness of knee extension greater than hip flexion” criterion, while the ENMC 2013 CD (failure rate 43%) primarily fails because of the similar “weakness of knee extension equal to or greater than hip flexion” criterion.

The comparative knee extension/hip flexion criteria we found to be of low sensitivity were developed at 3 workshops (2008 MRC meeting in Oxford,<sup>8</sup> 2009 IM meeting in Paris,<sup>9</sup> and 2011 ENMC workshop in Naarden<sup>10</sup>) and have been incorporated into 7 of the most recently developed categories studied. We therefore wondered whether data from 2 of our institutions (BWH and JHMI) were representative of other institutions’ experience. Accordingly, we examined MMT data from 2 other institutions, ANRI and UWASH, in an additional 66 patients with IBM, for a total of 266 patients with IBM. We found consistent strength criteria failure rates at all 4 institutions

(figure 2C). Thus, “finger flexion weakness” had the lowest failure rates (BWH/JHMI/ANRI/UWASH: 0%/3%/0%/25%, respectively) while “weakness of knee extension greater than hip flexion” had the highest failure rates (BWH/JHMI/ANRI/UWASH: 68%/47%/38%/66%, respectively).

We also studied whether disease duration might affect the proportion of patients failing various strength criteria (figure 2D). For our BWH/JHMI cohort of 200 patients with IBM, the mean duration of IBM symptoms was 6.7 years. The failure rate for the best performing strength criterion “weakness of finger flexion” was low in both early disease (0% for duration 1–3 years) and late disease (5% for duration >10 years). In contrast, the failure rate for “weakness of knee extension greater than hip flexion” was higher early in disease (60% for duration 1–3 years) than in late disease (45% for duration >10 years). All of the

**Figure 3** Hierarchical structure of 4 schemes represented as Venn diagrams



(A) Griggs 1995 categories are nonoverlapping and nonhierarchical. (B) ENMC 2000 categories probable and definite are a nested hierarchy: all patients who meet definite criteria also meet probable criteria, and no gaps are present if a patient with probable gains additional features. (C) MRC 2010 categories are a hybrid between nested hierarchy (all CD patients meet Poss criteria) and nonoverlapping nonhierarchical (no PD patients meet CD or Poss criteria). This scheme contains a gap (both CD and Poss patients can gain an inclusion body myositis feature, such as tubulofilaments, resulting in loss of CD or Poss status, but yet not graduate to PD). (D) ENMC 2013 criteria are a hybrid between nested hierarchy (all CD patients meet Prob criteria) and an overlapping nonhierarchical category (CPD patients can also meet Prob or CD criteria, or they can fail to meet either Prob or CD criteria). CD = clinically defined; CPD = clinicopathologically defined; Def = definite; ENMC = European Neuromuscular Center; MRC = Medical Research Council; PD = pathologically defined; Poss = possible; Prob = probable.

knee extension–based criteria tended toward higher failure rates in early disease than in late disease.

**Structure of IBM classification schemes.** We studied the manner in which 5 published classification schemes (Griggs 1995, ENMC 2000, MRC 2010, IM 2010, and ENMC 2013) have created or avoided overlap or gaps of their categories (figure 3). The ENMC 2000 scheme is a strictly nested hierarchy, whereas the Griggs classification system is nonoverlapping nonhierarchical. Thus, all patients who satisfy ENMC 2000 definite criteria also satisfy ENMC possible criteria, whereas patients only satisfy Griggs possible or Griggs definite criteria, but not both.

All other schemes had a mixture of hierarchical and nonhierarchical structures, some with gaps. Thus, the MRC 2010 (figure 3C) scheme has a hierarchical nested category (all patients meeting CD also meet possible criteria) but also a nonoverlapping category (PD, some of whose inclusion criteria are exclusion criteria for CD and possible). However, these criteria have a gap, in that a patient without rimmed vacuoles can meet MRC 2010 CD criteria by the presence of

major histocompatibility complex class I (MHC-1) upregulation, but if that patient additionally had tubulofilaments, they would no longer meet CD criteria, yet would not graduate to the CPD category. Such a patient would be eliminated from the scheme entirely, failing to meet any MRC 2010 category. A similar phenomenon occurs with IM 2010 categories.

Lastly, the most recently proposed ENMC 2013 scheme is similar to the MRC 2010 scheme with both hierarchical and nonhierarchical categories, but eliminates any gaps (figure 3D). However, this scheme contains partial overlap of categories. Thus, patients satisfying CPD criteria can satisfy all 3 (CPD, CD, and probable) categories, only 2 categories (CPD and probable), or only 1 category (CPD alone).

**Application of machine learning to construct data-derived IBM diagnostic criteria.** We constructed simple IBM diagnostic criteria by using only highly performing features. A range of machine learning algorithms performed classification with higher diagnostic accuracy (98%–99%) than currently used diagnostic criteria (appendix e-2). These approaches are probably not suited to routine clinical practice use, but their unbiased selection of features suggests best performing combinations. Thus, classification trees (appendix e-2) indicated that finger flexion weakness, rimmed vacuoles, and invasion of nonnecrotic muscle fibers used in combination were highly performing features. Indeed, the following data-derived criteria had 90% sensitivity and 96% specificity among 371 patients: (1) finger flexor or quadriceps weakness, and (2) endomyosial inflammation, and (3) either invasion of nonnecrotic muscle fibers or rimmed vacuoles.

**DISCUSSION** Diagnostic criteria for IBM are necessary to define inclusion criteria in clinical trials, and are therefore a prerequisite for eventual regulatory approval of an IBM therapeutic. From 24 IBM diagnostic categories that have been proposed since 1987, we found that 12 had been updated, had poor specificity, or had sufficient ambiguities precluding their application. Of the remaining 12 categories, we found that although all had very high specificities of 97% or greater, some had poor sensitivities, as low as 11%, while the best performing category (ENMC 2013 probable) had a sensitivity of 84%. We traced the poor performance of many IBM categories to either highly specific but very insensitive pathologic criteria or recently introduced comparison strength criteria with low sensitivity. Poor performance of certain specialized pathologic features is at least partially attributable to not being frequently performed in clinical practice. Although these pathologic features, in addition to recently identified biomarkers such as p62 and TDP-43 (TAR DNA-binding protein-43)

immunoreactivity, might be more frequently present in specialized research laboratories, their performance by clinical pathology laboratories in the course of clinical care is uncommon. Nevertheless, the use of these specialized pathologic features in the MRC 2010 PD category, which has been advocated for use in clinical trials,<sup>9,13</sup> and the ENMC 2013 CPD may be limiting in that only 11% and 15% of 200 patients with IBM fulfill these categories' criteria, respectively.

In addition, we found that recently introduced comparative strength criteria pertaining to knee extension and hip flexion, adopted into 7 new categories, were of low sensitivity. Of 266 patients with IBM at 4 institutions in the United States and Australia, a mean of 55% (range 38%–69%) fail to meet “weakness of knee extension greater than hip flexion” and a mean of 41% (range 21%–63%) fail to meet “weakness of knee extension equal to or greater than hip flexion” criteria. The low sensitivity of these criteria across our 4 institutions is otherwise consistent with existing literature that has emphasized substantial weakness of hip flexion in IBM.<sup>12,14,15</sup> We also note that in our cohort, the lower proportion of shorter-duration patients than longer-duration patients meeting the “weakness of knee extension equal to or greater than hip flexion” criterion would bias clinical trials using it toward longer-duration, more severely affected patients, which is likely not ideal for development of therapeutics aimed at treating patients with IBM before extensive muscle loss and fibrosis are present.

Some of our results can be contrasted with a recent single-institution study of 67 patients with IBM.<sup>13</sup> This study used different criteria for several categories than we did. Their Griggs possible category<sup>5</sup> required rimmed vacuoles as an inclusion criterion, which we did not require. The ENMC 2013 CPD criteria used in this previous study required the presence of MHC-1, which is not listed as an ENMC 2013 requirement as recently published.<sup>10</sup> Only 18% of their patients failed the criterion “weakness of knee extension equal to or greater than hip flexion” in contrast to 21% to 63% of patients across our 4 institutions. This previous study and ours both found low sensitivity of ENMC 2013 CPD (16% previous study, 15% our study) and ENMC 2013 CD (39% previous study, 57% our study) categories.

More generally, several considerations for future IBM criteria development applicable to clinical trial enrollment can be noted. The use of pathologic methods not routinely performed might be avoided. Criteria should not exclude patients because of the presence of typical features. Ambiguous strength criteria that might be met on one side but not the other side of the body should be avoided. Comparison strength criteria (comparing one muscle group with another) may result in patients previously meeting

criteria subsequently failing to meet such criteria (e.g., knee extension weaker than hip flexion at one point in disease, but not as hip flexion weakness progresses). Circular definitions (e.g., IM 2010 PD having clinical criterion of “consistent with IBM”) should be avoided, because they make category membership undeterminable. Lastly, because all the current “probable” and “possible” categories have high specificity ( $\geq 97\%$ ), these could in fact be called “definite” categories.

Schemes that have nonhierarchical natures should be used cautiously because these may result in paradoxical classification (e.g., meeting stringent MRC 2010 PD criteria but failing the weaker MRC 2010 possible criteria). In contrast, all patients meeting ENMC 2000 definite automatically meet the ENMC 2000 probable category. Overlapping nonhierarchical approaches can also create unusual behavior: some patients meeting the ENMC 2013 CPD category meet ENMC 2013 probable or both ENMC 2013 probable and ENMC 2013 CD, while others meet neither. Nonhierarchical structures can also create unintended gaps, in which adding an additional typical feature of IBM can cause a patient who satisfies one category to be eliminated from that category as well as all of a scheme's categories. For example, a patient who has rimmed vacuoles but not myofiber invasion can meet the MRC 2010 CD criteria, but if that same patient has an electron microscopy study showing tubulofilaments (a feature that should suggest the diagnosis more strongly), they will then fail to meet criteria for any of the categories.

Over 18 years during which 5 workshops have produced publications of IBM diagnostic criteria,<sup>5,6,8–10</sup> no studies of the performance of these criteria have been published, until this year.<sup>13</sup> Herein, we have shown that such criteria can be empirically derived in a principled fashion directly from patient data through the use of unbiased machine learning approaches. Such data-derived criteria performed well in our cohort with 90% sensitivity and 96% specificity, although we recognize that some patients with severe inflammatory myopathies or inherited vacuolar myopathy may meet these criteria. The use of data-derived criteria is an approach that can be used in future IBM criteria development.

#### AUTHOR CONTRIBUTIONS

T.E.L. and S.A.G. designed the study, analyzed and interpreted data, and drafted and revised the manuscript. A.L.M., A.A.A., M.D.W., and M.N. analyzed patient data and revised the manuscript.

#### ACKNOWLEDGMENT

T.E.L. and A.L.M. thank Kristie Owoyemi (Johns Hopkins) and Kalpana Prasad (National Neuroscience Institute, Singapore) for entering information into a database.

#### STUDY FUNDING

No targeted funding reported.

## DISCLOSURE

T. Lloyd has consulted for Novartis. A. Mammen is on the medical advisory board for Biogen and aTyr Pharma. A. Amato is on the medical advisory board for Biogen. M. Weiss has consulted for CSL-Behring, Questcor Pharmaceuticals, Baxter, and Grifols and has been on the speakers bureau for Grifols and Walgreens. He has received honoraria for speaking from the AANEM. He has funding support from the ALS Therapy Alliance and Northeast ALS Consortium. M. Needham has consulted for Novartis. S. Greenberg is an inventor on intellectual property owned and managed by Brigham and Women's Hospital, and has had sponsored research by MedImmune, LLC, and consulted for aTyr Pharma. Go to [Neurology.org](http://Neurology.org) for full disclosures.

Received November 9, 2013. Accepted in final form April 22, 2014.

## REFERENCES

1. Calabrese LH, Mitsumoto H, Chou SM. Inclusion body myositis presenting as treatment-resistant polymyositis. *Arthritis Rheum* 1987;30:397–403.
2. Dalakas MC. Polymyositis, dermatomyositis and inclusion-body myositis. *N Engl J Med* 1991;325:1487–1498.
3. Mastaglia FL, Phillips BA. Idiopathic inflammatory myopathies: epidemiology, classification, and diagnostic criteria. *Rheum Dis Clin North Am* 2002;28:723–741.
4. Tawil R, Griggs RC. Inclusion body myositis. *Curr Opin Rheumatol* 2002;14:653–657.
5. Griggs RC, Askanas V, DiMauro S, et al. Inclusion body myositis and myopathies. *Ann Neurol* 1995;38:705–713.
6. Verschuuren JJ, Badrising UA, Wintzen AR, van Engelen BG, van der Hoeven H, Hoogendijk JE. Inclusion body myositis. In: Emery AEH, editor. *Diagnostic Criteria for Neuromuscular Disorders*. London: Royal Society of Medicine Press, European Neuromuscular Centre; 1997:81–84. Available at: <http://ibmmyositis.com/emery81.pdf>. Accessed January 26, 2014.
7. Badrising UA, Maat-Schieman M, van Duinen SG, et al. Epidemiology of inclusion body myositis in the Netherlands: a nationwide study. *Neurology* 2000;55:1385–1387.
8. Hilton-Jones D, Miller A, Parton M, Holton J, Sewry C, Hanna MG. Inclusion body myositis: MRC Centre for Neuromuscular Diseases, IBM Workshop, London, June 13, 2008. *Neuromuscul Disord* 2010;20:142–147.
9. Benveniste O, Hilton-Jones D. International Workshop on Inclusion Body Myositis held at the Institute of Myology, Paris, on May 29, 2009. *Neuromuscul Disord* 2010;20:414–421.
10. Rose MR. 188th ENMC International Workshop: Inclusion Body Myositis, 2–4 December, 2011, Naarden, The Netherlands. *Neuromuscul Disord* 2013;23:1044–1055.
11. Curk T, Demsar J, Xu Q, et al. Microarray data mining with visual programming. *Bioinformatics* 2005;21:396–398.
12. Lotz BP, Engel AG, Nishino H, Stevens JC, Litchy WJ. Inclusion body myositis: observations in 40 patients. *Brain* 1989;112:727–747.
13. Brady S, Squier W, Hilton-Jones D. Clinical assessment determines the diagnosis of inclusion body myositis independently of pathological features. *J Neurol Neurosurg Psychiatry* 2013;84:1240–1246.
14. Cox FM, Titulaer MJ, Sont JK, Wintzen AR, Verschuuren JJ, Badrising UA. A 12-year follow-up in sporadic inclusion body myositis: an end stage with major disabilities. *Brain* 2011;134:3167–3175.
15. Needham M, James I, Corbett A, et al. Sporadic inclusion body myositis: phenotypic variability and influence of HLA-DR3 in a cohort of 57 Australian cases. *J Neurol Neurosurg Psychiatry* 2008;79:1056–1060.

## 20 Minutes Pack a Punch

### *Neurology*<sup>®</sup> Podcasts

- Interviews with top experts on new clinical research in neurology
- Editorial comments on selected articles
- Convenient—listen during your commute, at your desk, or even at the gym
- On demand—it's there when you want it
- Fun and engaging
- New topic each week
- FREE

Listen now at [www.aan.com/podcast](http://www.aan.com/podcast)