

Development and validation of the Myasthenia Gravis Impairment Index

OPEN

Carolina Barnett, MD,
PhD
Vera Bril, MD, FRCP
Moirá Kapral, MD, MSc
Abhaya Kulkarni, MD,
PhD
Aileen M. Davis, PhD

Correspondence to
Dr. Barnett:
c.barnett@utoronto.ca

ABSTRACT

Objective: We aimed to develop a measure of myasthenia gravis impairment using a previously developed framework and to evaluate reliability and validity, specifically face, content, and construct validity.

Methods: The first draft of the Myasthenia Gravis Impairment Index (MGII) included examination items from available measures enriched with newly developed, patient-reported items, modified after patient input. International neuromuscular specialists evaluated face and content validity via an e-mail survey. Test-retest reliability was assessed in stable patients at a 3-week interval and interrater reliability was evaluated in the same day. Construct validity was assessed through correlations between the MGII and other measures and by comparing scores in different patient groups.

Results: The first draft was assessed by 18 patients, and 72 specialists answered the survey. The second draft had 7 examination and 22 patient-reported items. Field testing included 200 patients, with 54 patients completing the reliability studies. Test-retest reliability of the total score was good (intraclass correlation coefficient 0.92; 95% confidence interval 0.79–0.94), as was interrater reliability of the examination component (intraclass correlation coefficient 0.81; 95% confidence interval 0.79–0.94). The MGII correlated well with comparison measures, with higher correlations with the MG-activities of daily living ($r = 0.91$) and MG-specific quality of life 15-item scale ($r = 0.78$). When assessing different patient groups, the scores followed expected patterns.

Conclusions: The MGII was developed using a patient-centered framework of myasthenia-related impairments and incorporating patient input throughout the development process. It is reliable in an outpatient setting and has demonstrated construct validity. Responsiveness studies are under way. *Neurology*® 2016;87:879–886

GLOSSARY

CI = confidence interval; **COSMIN** = Consensus-Based Standards for the Selection of Health Measurement Instruments; **FDA** = Food and Drug Administration; **ICC** = intraclass correlation coefficient; **IQR** = interquartile range; **MG** = myasthenia gravis; **MG-ADL** = myasthenia gravis-specific activities of daily living; **MGC** = Myasthenia Gravis Composite; **MGFA** = Myasthenia Gravis Foundation of America; **MGII** = Myasthenia Gravis Impairment Index; **MG-QOL15** = myasthenia gravis-specific quality of life 15-item scale; **PR** = patient-reported; **QMGS** = Quantitative Myasthenia Gravis Scale; **QoL** = quality of life.

Myasthenia gravis (MG) is characterized by fatigable weakness and symptom fluctuations.¹ Measurement of disease severity is therefore challenging, since the examination at one point in time might not accurately reflect the patients' clinical status. We previously conducted a qualitative study to assess the experiences of patients with MG and developed a conceptual framework of disease severity in MG.² We conducted semistructured interviews in which patients described their experiences with MG ($n = 20$). The interviews were analyzed through

Editorial, page 858

Supplemental data
at Neurology.org

From the Division of Neurology (C.B., V.B.), Department of Medicine, University of Toronto and University Health Network; Institute of Health Policy, Management and Evaluation (C.B., M.K., A.K., A.M.D.), University of Toronto; Department of Medicine (M.K.), University of Toronto and University Health Network; Department of Neurosurgery (A.K.), Sick Kids Hospital, Toronto; Division of Health Care and Outcomes Research (A.M.D.), Krembil Institute, University Health Network, Toronto; and Department of Physical Therapy and Graduate Department of Rehabilitation Science (A.M.D.), University of Toronto, Canada.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article. The Article Processing charge was paid by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially.

content analysis to develop a framework of disease severity based on the impairments. In the framework, not only the severity of individual impairments but also fatigability, defined as the worsening or triggering of the impairments with activities or through the day, accounted for overall MG severity.² This highlights the importance of adding content related to fatigability to outcome measures.

Several tools have been developed to measure MG-related impairments, variably including different impairments, fatigability, and patient-reported (PR) and performance-based items.^{3–8} Some of these measures include items not responsive to interventions possibly because of floor effects.^{3–11} In addition, most measures were developed using expert opinion, before current standards for measure development were defined by bodies such as the Food and Drug Administration (FDA)¹² and the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) Group.¹³ The FDA and COSMIN guidelines require patient input into the process of development of PR measures. We aimed to develop a measure of MG impairment that can distinguish between patients with different degrees of severity and that is responsive to change in status. Conforming with current guidelines, we incorporated patient input based on the previously developed framework,² ensuring content validity. This report describes item generation and reduction, content and face validity, reliability and construct validity of the Myasthenia Gravis Impairment Index (MGII).

METHODS **Standard protocol approvals, registrations, and patient consents.** The study was approved by the University Health Network Research Ethics Board and all participants provided written informed consent.

Item generation. This followed a formative structure,^{14–16} whereby the combined items define the construct of interest and items are not necessarily correlated. We reviewed the literature to incorporate items from available measures, representing the main themes and subthemes of the guiding framework. The search was conducted in Embase (1947–2013) and MEDLINE (1946–2013) using a validated filter¹⁷ for detecting publications related to outcome measures, adding the terms “myasthenia gravis” and “myasthenia.” Items from measures identified in the search were reviewed for content and face validity, reliability, and responsiveness, and those in keeping with the framework and with adequate measurement properties were retained. When no suitable items were found, we developed new items using as anchors themes obtained through the qualitative study. The

measure was structured to incorporate fatigability and severity items for each body function/structure defined in the framework.² Exceptions were swallowing and breathing, where fatigability is difficult to isolate from severity. Two measurement experts, 3 neuromuscular physicians, a speech therapist, and a neuro-ophthalmologist assessed the preliminary items for input on content, wording, relevance, and missing items. Modifications were made to develop a first draft.

Pilot testing and user input. Patients who were previously interviewed evaluated the draft through cognitive debriefing, assessing each item and commenting on wording and irrelevant or missing items. Interviewing stopped after no new comments were elicited. International MG experts provided input on the items through an electronic survey, in which they rated each item using a 5-point Likert scale for relevance and feasibility. They also commented on wording and suggested additional items. Medians and interquartile range (IQR) were calculated and items with a median ≤ 4 on either question were reviewed. Based on the opinions of patients and experts, further modifications were made. Patients reassessed the new draft for wording, clarity, and formatting. This version was field tested.

Field testing. All patients with confirmed MG attending the Neuromuscular Clinic, Toronto General Hospital, were invited to participate, except those participating in the qualitative study. The patients were assessed with the new measure, hereafter the MGII, and answered demographic questions as well as PR measures including the PROMIS-Fatigue,¹⁸ and an MG-specific activities of daily living (MG-ADL) questionnaire.^{5,9} Quality of life (QoL) measures included the MG-specific quality of life 15-item scale (MG-QOL15),¹⁹ the generic 36-Item Short Form Health Survey,²⁰ the INQoL (Individualized Neuromuscular Disease Quality of Life),²¹ and the EuroQoL 5 Dimensions.²² Patients completed the questionnaires in clinic or at home according to their preference. A trained examiner (clinical fellow or physician assistant) obtained the MGII examination items, the Quantitative Myasthenia Gravis Scale (QMGS),³ and the MG Composite (MGC) score.⁴ We also recorded clinical variables, including disease duration, MG type, thymoma, antibody status, medications, and Myasthenia Gravis Foundation of America (MGFA) class.

Reliability. Patients were eligible if they had no change in clinical status or MG medications for the last 3 months. We tested interrater reliability for the examination items on the same day, with a rest period of 30 to 60 minutes between the 2 raters who were blinded to each other's scores, and test–retest reliability for all items 2 to 3 weeks after the first visit. Patients were asked on visit 2 whether they felt better, worse, or unchanged, and only stable patients were included in the test–retest calculations. Reliability was tested with intraclass correlation coefficients (ICCs) for total score and subscales using a random-effects model (ICC 2, 1).²³ ICC values ≥ 0.8 are recommended for group and ≥ 0.9 for individual use.²⁴ We also calculated test–retest weighted kappas for all items, and interrater weighted kappas and absolute agreement for the examination items. There is no universal consensus as to the interpretation of kappa, but usually values between 0.6 and 0.8 are considered substantial and >0.8 excellent agreement.²⁵ Finally, we calculated the standard error of measurement.²⁶ With an estimated ICC of 0.8 and a lower 95% confidence interval (CI) of 0.7, 48 participants are required.²⁷ We aimed to enroll 60 patients allowing for loss to follow-up or change in status.

Item reduction and final measure. We calculated the percentage of missing responses and score distributions for each item.

Interitem correlations were calculated to identify redundant items. In a formative model, item-total correlations and Cronbach α are not informative^{14,15,28} and were not calculated. Considerations for item reduction were as follows: >10% missing responses, interitem correlations ≥ 0.9 (redundancy), and low reliability (weighted kappa ≤ 0.50). We analyzed items for floor effects (% score = 0) excluding patients in whom scores of 0 are expected (remission), and items with >70% of score = 0 were candidates for reduction. However, the final decision on item retention was based on the conceptual framework, as less prevalent impairments are still relevant to ensure content validity.

Construct validity. We formulated several hypotheses regarding the correlations between the MGII scores and other measures and for the difference in MGII scores among known patient groups. Confirming $\geq 75\%$ of the predefined hypotheses supports construct validity.¹³ We compared mean MGII scores in patients with ocular and generalized disease, hypothesizing that ocular patients would have lower scores (assessed by *t* test) than those with generalized disease. Furthermore, we expected that pure ocular patients would have low scores on generalized items. We also compared mean MGII scores by different MGFA classes, expecting scores near 0 in patients in remission, with subsequently higher scores with increasing MGFA class (tested through analysis of variance). In addition, we expected high correlations ($r = 0.6\text{--}0.8$) between the MGII and other impairment measures (e.g., QMGs, MGC). Because QoL reflects other dimensions such as mental health and social factors, we expected lower correlations ($r = 0.4\text{--}0.7$) with QoL measures. Overall, we expected positive correlations, except with the 36-Item Short Form Health Survey and EuroQol 5 Dimensions in which lower scores represent worse QoL. Significance was considered to be $p < 0.05$. Based on the correlational analyses, with $\alpha = 0.05$, $\beta = 0.2$, and minimum correlation of 0.4, 40 patients are required.²⁹ However, we aimed to enroll at least 150 patients, following published recommendations.¹⁴ Statistical analyses were performed using R statistical software version 3.1.2.

RESULTS Figure 1 illustrates the item-generation process described below.

Item generation. Sixty-two unique items were identified from published measures. Of these, 7 examination items fit the framework and had evidence of validity, reliability, and/or responsiveness. These were 4 items from the QMGs³: ptosis and arm, leg, and neck endurance, modified based on previous data¹⁰; one item (neck strength) from the MGC⁴; and 2 items from the MG Impairment Scale⁶: diplopia and facialis inferior (modified for clarity). We developed 12 PR items, which were expanded after input from local experts to 18 PR items: diplopia (double vision) and ptosis (droopy eyelids) fatigability (2 each); swallowing (1); chewing (2); breathing (1); fatigability of arms, legs, and neck (1 each); generalized fatigability (1); and voice and speech articulation fatigability (3 each).

Pilot testing and user input. Thirteen patients participated in the first round of cognitive debriefing. Major suggestions were adding items on ptosis and diplopia severity and to modify the anchors for ocular

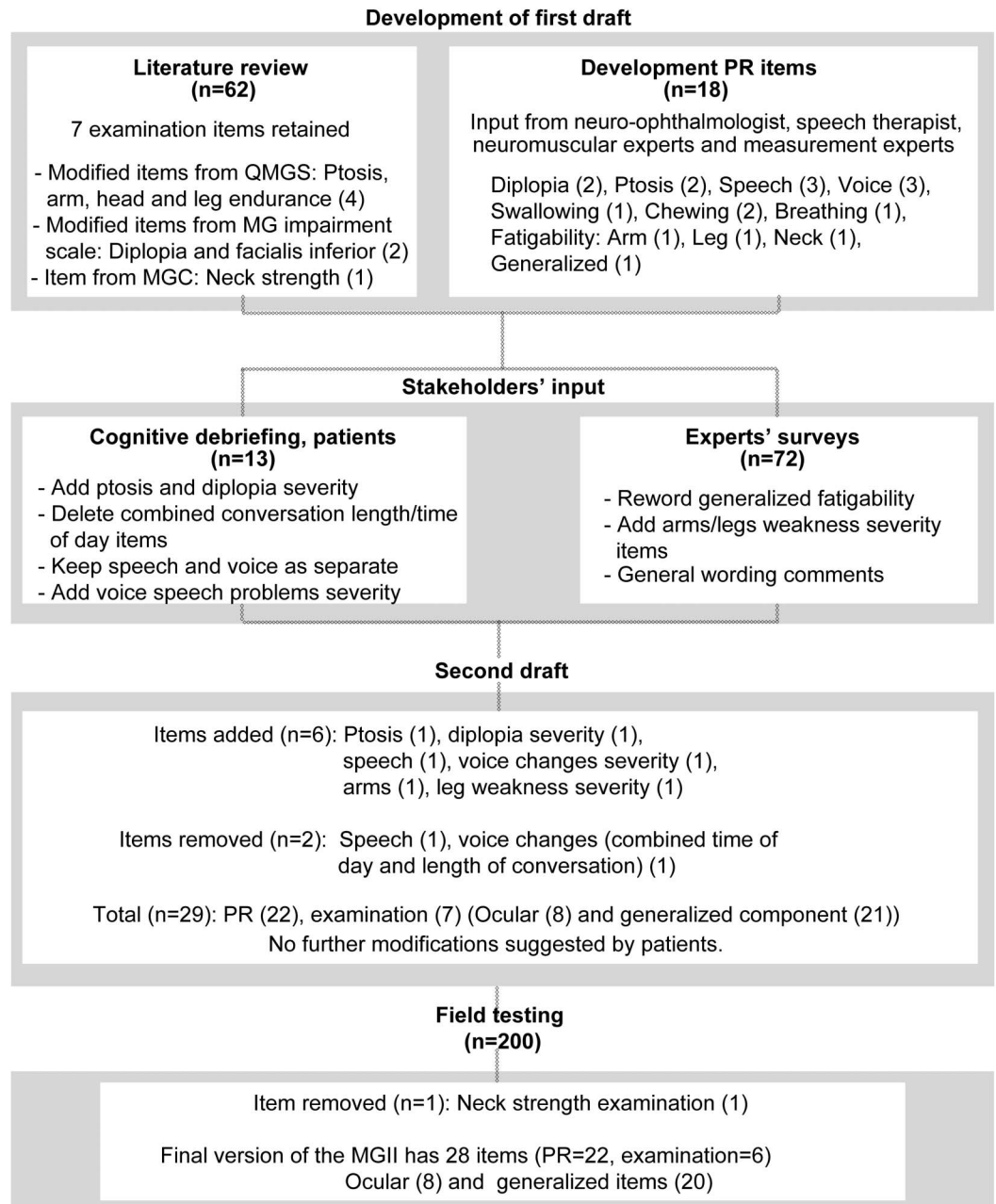
fatigability. Patients also suggested keeping separate items for speech (i.e., slurred speech) and voice changes (i.e., nasal voice, hypophonia), with different items for fatigability through the day and with long conversations, because combined items were confusing. Patients suggested asking about the severity of voice and speech changes. All patients thought the generalized fatigability item was relevant and suggested wording changes. One hundred sixty-eight MG experts were invited to evaluate the resultant draft and 72 (42.3%) answered the survey. The median response for relevance and feasibility was ≥ 4 for all items, with exception of the PR item of generalized fatigability, with a median feasibility of 3 (IQR 2–4). Some experts thought that patients could not distinguish MG fatigability from other causes of fatigue. Since all patients reported that generalized fatigability was relevant, the item was retained with modified wording. The inclusion of items on weakness severity of legs and arms was suggested in addition to those on fatigability. The modified MGII draft, incorporating input from patients and experts, had 22 PR items and 7 examination items. Another 5 patients evaluated the second draft, finding it clear, relevant, and easy to complete.

Field testing. Two hundred patients completed the assessments (118 sent the questionnaires by mail). Demographic and clinical characteristics are in table 1. All items had $\leq 6\%$ missing data and only 9 patients (4.5%) had >10% missing items. Score distributions and missing data for each item are in table e-1 on the *Neurology*[®] Web site at Neurology.org.

Reliability. Sixty-three patients were enrolled and the 54 returning for visit 2 were assessed for interrater reliability. Of the returning patients, 42 were unchanged from baseline and were included in test–retest reliability. All items had test–retest weighted kappa values between 0.57 and 0.90. Exceptions were examination items for lower face strength (weighted kappa = 0.54) and neck weakness (undefined kappa, a case in which kappa values are not meaningful).^{30,31} Table e-2 presents the reliability statistics of all items. For the raw sum of the items (PR and examination), test–retest reliability was excellent with an ICC of 0.92 (95% CI 0.86–0.96) for the total score and ICC of 0.91 (95% CI 0.84–0.95) for the PR items. The sum of the examination items had good interrater reliability with ICC of 0.81 (95% CI 0.79–0.94).

Development of the final measure and scoring. Two items had high floor effect in symptomatic patients: lower facial weakness (74%) and neck strength examination (77%). There were no interitem correlations > 0.9 . The neck strength item was eliminated based on floor effect and the presence of 2 other items

Figure 1 Flowchart of the item-generation process



MG = myasthenia gravis; MGC = Myasthenia Gravis Composite; MGII = Myasthenia Gravis Impairment Index; PR = patient-reported; QMGS = Quantitative Myasthenia Gravis Scale.

reflecting neck impairment. The lower facial weakness item was retained because it is the only bulbar examination item, is easy to implement, and has high interrater agreement. All other items had adequate reliability and were considered necessary for content validity based on the framework.²

The final version of the MGII has 28 items, 6 examination and 22 PR. The items are unweighted and a total score is obtained by the sum of all the examination and PR items, reflecting overall MG disease severity. The total score has a possible range of 0 to 84, where higher scores indicate more severe

impairments. We also developed an ocular subscore with 8 items (2 examination and 6 PR items) and a generalized subscore with 20 items (4 examination and 16 PR items). Table e-2 specifies which items belong to each subscore. The possible score ranges are 0 to 23 for the ocular and 0 to 61 for the generalized subscore. Missing items were imputed with the mean score of the corresponding item subscore (ocular or generalized).

Reliability statistics for the final version of the MGII were recalculated for the total scores and subscores, as well as for the examination and PR items.

Table 1 Demographic and clinical characteristics of field-testing participants (n = 200)

Demographic characteristics	Mean ± SD or n (%)	Clinical characteristics	Mean ± SD or n (%)
Age, y	59.8 ± 14.5	Disease duration, mo	121 ± 124
Sex, female	100 (50)	Generalized disease	164 (82)
Marital status (n = 195)		Thymoma	45 (22)
Single	25 (13)	Thymectomy	90 (45)
Married/common-law	137 (70)	AChRAb (tested n = 106)	70 (60)
Divorced/separated	19 (10)	MuSK (tested n = 30)	4 (2)
Widowed	15 (7)	MGFA class ^a	
Highest level of education (n = 195)		Remission	7 (3)
Some/completed elementary school	12 (6)	Minimal manifestations	13 (6)
Some/completed secondary school	46 (23)	I	47 (24)
Some/completed college or community school	32 (16)	II	91 (46)
Some or completed university	99 (51)	III	36 (18)
Other	7 (4)	IV	6 (3)
Employment status (n = 198)		Current treatments	
Employed (full/part time)	74 (38)	Pyridostigmine	135 (68)
On disability	24 (12)	Prednisone	110 (55)
Unemployed/not employed	5 (2)	Azathioprine	56 (28)
Retired	80 (40)	Mycophenolate	36 (18)
Other	15 (8)	No medications	16 (8)
Race/ethnicity (n = 196, more than 1 option possible)		QMGS	7.9 ± 5.6
White	126 (64)	MGC	6.0 ± 6.0
Black	17 (9)	MG-ADL	4.1 ± 3.9
Aboriginal people of North America	2 (1)	MG-QOL15	16.1 ± 14.3
Hispanic/Latino	2 (1)		
Filipino	14 (7)		
Chinese	21 (11)		
Other	16 (8)		
English as first language	144 (75)		

Abbreviations: AChRAb = acetylcholine receptor autoantibody; MG-ADL = myasthenia gravis-specific activities of daily living; MGC = Myasthenia Gravis Composite; MGFA = Myasthenia Gravis Foundation of America; MG-QOL15 = myasthenia gravis-specific quality of life 15-item scale; MuSK = muscle-specific kinase; QMGS = Quantitative Myasthenia Gravis Score.

^aMGFA classes: remission = no symptoms for ≥1 year; minimal manifestation = minimal symptoms for ≤1 year; I = pure ocular; II = mild generalized; III = moderate generalized; IV = severe generalized.

These can be found in table 2 with the corresponding standard error of measurement values.

Construct validity. The mean total MGII score was 19.6 ± 16.0 (range 0–69), with mean ocular subscore of 7.6 ± 6.9 and mean MGII generalized subscore of 6.9 ± 6.6 . The MGII had a broader distribution of scores and low floor effect (5%) compared to the MGC (16%) and MG-ADL (22%, figure e-1). As hypothesized, the MGII total score was lower in patients with ocular compared to generalized disease (11.4 ± 9.8 and 21.4 ± 16.6 , $p < 0.00001$) and ocular patients had minimal scores in the generalized subscore (mean 2.9 ± 5.0 ; figure 2A). Patients in

remission had very low scores (mean 1.0 ± 1.2), and scores increased with progressively higher MGFA class ($p < 0.0001$; figure 2B). Seventy-eight percent of the correlations tested were in the expected directions and ranges, except higher correlations with the MG-ADL ($r = 0.91$) and the MG-QOL15 ($r = 0.78$). Table 3 summarizes the results of these associations. When the PR component was analyzed alone and the sample was split based on age, the correlational studies had similar values (data not shown).

DISCUSSION The MGII is a new measure of MG disease severity aimed at measuring MG-related impairments as defined by the International

Table 2 Reliability statistics for the final version of the MGII and its components

	Visit 1, mean ± SD (n = 54)	Visit 2, mean ± SD (n = 54)	ICC	SEM
Total MGII score (28 items)	14.1 ± 11.6	13.2 ± 12.4	0.92 (0.86-0.96)	3.3
Ocular subscore (8 items)	5.8 ± 6.1	4.8 ± 5.9	0.87 (0.76-0.93)	2.2
Generalized subscore (20 items)	8.3 ± 7.9	8.4 ± 5.6	0.92 (0.86-0.96)	2.2
Patient-reported items (22 items)	11.5 ± 10.0	10.7 ± 10.5	0.91 (0.84-0.95)	3.0
Examination, 6 items (interrater) ^a	2.7 ± 2.6	3.1 ± 3.0	0.80 (0.68-0.88)	1.2
Examination, 6 items (test-retest)	2.6 ± 2.5	2.5 ± 2.6	0.90 (0.82-0.94)	0.8

Abbreviations: ICC = intraclass correlation coefficient; MGII = Myasthenia Gravis Impairment Index; SEM = standard error of measurement.

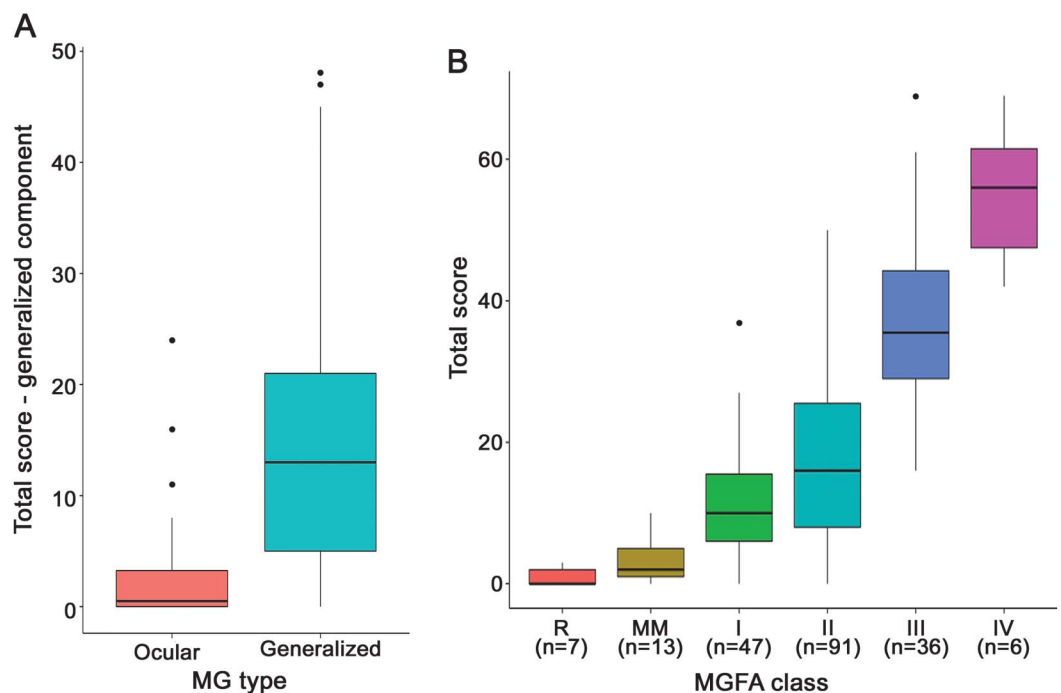
For ICC, values >0.8 are good for groups and >0.9 good for individual use. SEM = SD × sqrt(1 - ICC); this is the error around a single measurement.

^aFor interrater reliability, scores are for rater 1 and rater 2.

Classification of Functioning, Disability and Health.³² Following current guidelines, we used a patient-centered approach, aimed at measuring impairments relevant to patients. Therefore, we used a predominantly PR format, designed to capture those impairments that typically fluctuate and that are triggered by activities not easily observed in a clinical visit. The questions have a 2-week recall period to reduce recall error and because some interventions can show efficacy within 2 weeks (e.g., IV immunoglobulin). We also incorporated

some examination items because these are important for clinical decision-making for physicians. However, the PR component has excellent reliability and construct validity so it can be used as a stand-alone instrument.

Content validity was ensured by developing a conceptual framework of MG impairments based on patients' experiences,² and using the framework for item generation. For example, the framework has speech and voice impairments as separate entities and this was consistent with field-testing data in which voice

Figure 2 Known-groups validity

(A) Generalized subscores in patients with ocular and generalized disease. As expected, pure ocular patients had very low scores on the generalized subscore and lower than the generalized patients (median 0.5, interquartile range 3; mean 2.9 ± 5.0 and 21.4 ± 16.6 , $p < 0.00001$). (B) Total scores according to different MGFA classes. As expected, mean scores were near 0 for patients in remission and increased with increasing MGFA class (remission [R] = 1 ± 1.3 ; minimal manifestations [MM] = 3.2 ± 3.1 ; class I = 11.5 ± 7.4 ; class II = 18.4 ± 12.5 ; class III = 37.1 ± 13.3 ; class IV = 55.2 ± 10.3). Analysis of variance, $p < 0.0001$. MG = myasthenia gravis; MGFA = Myasthenia Gravis Foundation of America.

Table 3 Construct validity, correlation studies for the total Myasthenia Gravis Impairment Index scores

Comparison measure	Pearson <i>r</i> (95% CI)	<i>p</i> Value
QMGS	0.79 (0.73 to 0.84)	<0.0001
MGC	0.81 (0.76 to 0.86)	<0.0001
MG-ADL ^a	0.91 (0.88 to 0.93)	<0.0001
PROMIS-Fatigue short form	0.69 (0.60 to 0.75)	<0.0001
MG-QOL15 ^a	0.79 (0.73 to 0.84)	<0.0001
SF-36, physical health	-0.61 (-0.69 to -0.51)	<0.0001
SF-36, mental health	-0.51 (-0.61 to -0.40)	<0.0001
EQ-5D	-0.62 (-0.7 to -0.56)	<0.0001
INQoL, overall QoL score	0.63 (0.54 to 0.71)	<0.0001

Abbreviations: CI = confidence interval; EQ-5D = EuroQol 5 Dimensions; INQoL = Individualized Neuromuscular Quality of Life; MG-ADL = myasthenia gravis-specific activities of daily living; MGC = Myasthenia Gravis Composite; MG-QOL15 = myasthenia gravis-specific quality of life 15-item scale; PROMIS = Patient Reported Outcomes Measurement Information System; QMGS = Quantitative Myasthenia Gravis Score; QoL = quality of life; SF-36 = 36-Item Short Form Health Survey.

In the SF-36 and EQ-5D, higher scores indicate better QoL; therefore, negative correlations were expected.

^a These were the only correlations that were above the hypothesized range.

and speech items were shown to represent different phenomena based on their moderate correlations ($r = 0.5-0.6$). This is in keeping with the different muscles involved, since voice impairments reflect mostly laryngeal or palatal weakness and speech changes are due to tongue and labial weakness. This is in contrast to other measures in which these impairments are measured together. We also added an item on generalized physical fatigability even though some experts thought patients would have problems separating MG fatigability from general fatigue. However, field testing showed minimal missing data and, in fact, it was the item with the smallest floor effect, highlighting the importance of physical fatigability for patients with MG. We incorporated several items measuring fatigability of specific muscles or functions using different triggers (e.g., time of day) when appropriate to better capture the heterogeneity of the impairments. This was supported by patients through cognitive debriefing and during field testing and might explain the broader distribution of scores and limited floor effect compared to other measures.

The MGII scores can be presented as a total score (overall MG severity) and 2 subscores, reflecting ocular and generalized impairments. The subscores have excellent reliability, and patients with pure ocular disease had very low scores in the generalized component, supporting the validity of the subscores. Finally, individual items are unweighted since, with >20 items, weighting has minimal effect on the total scores.^{24,28} However, the total score is implicitly weighted toward the generalized component since 20 of the 28 items reflect generalized impairments.

The MGII has excellent reliability and is easy to implement. The only equipment required is a watch that can time in seconds. Patients can likely complete the questionnaire in the waiting room, minimizing the effect on a clinical appointment. We developed and tested the MGII in outpatients who were able to complete the PR component, and thus, it is not aimed at patients who are currently in myasthenic crisis (intubated). Arguably, in myasthenic crisis, outcomes such as time to extubation are more relevant. Therefore, the validation studies are generalizable to similar populations. In this setting, the MGII can discriminate between patient groups, including pure ocular and generalized, as well as different MGFA classes.

The correlations with the QMGS and MGC were within the expected ranges; however, the MGC has more floor effect (16%). The QMGS requires instrumentation, takes longer to complete, and several items are unresponsive to change.¹⁰ The correlation with the MG-ADL was higher than expected ($r = 0.9$). The MG-ADL is shorter, but it has a marked floor effect (22%) compared to the MGII (5%), which could affect responsiveness. Overall, the main advantages of the MGII are that it is easy to use, does not take long to complete, and at the same time provides a comprehensive assessment of the myasthenic state in patients. The correlation with the MG-QOL15 was higher than expected although the correlations with other QoL measures were within the hypothesized range.

The MGII reliability and construct validity studies were conducted in a single academic center and it is possible that some patients with milder disease are followed in the community. While this can affect generalizability, our cohort has broad clinical characteristics similar to published studies.⁴ In addition, the MGII was developed and tested in Canada and cross-cultural studies are required to confirm the measurement properties in other contexts. Studies assessing the responsiveness and minimal important difference of the MGII are under way.

The MGII is a measure of impairments in MG that has strong content validity based on a conceptual framework developed from patients' experiences. Therefore, the MGII follows current FDA guidelines for outcome measure development. While studies on responsiveness are under way, the MGII has demonstrated feasibility, reliability, and construct validity in an outpatient setting.

AUTHOR CONTRIBUTIONS

Carolina Barnett participated in the design of the study, data collection and analysis, and writing the manuscript. Vera Brill participated in the design of the study, data collection, and review of the data and manuscript. Moira Kapral participated in the design of the study and review of the data and manuscript. Abhaya Kulkarni participated in the design of the study and review of the data and manuscript. Aileen Davis participated in the design of the study, data analysis review, and review of the manuscript.

STUDY FUNDING

Carolina Barnett received salary support through a clinical research training award by the American Academy of Neurology and American Brain Foundation.

DISCLOSURE

C. Barnett received a clinical research training award by the American Academy of Neurology and American Brain Foundation. V. Bril has acted as consultant for Grifols, CSL, BioNevia, Lilly, Pfizer, Dainippon, Sumitomo, and Eisai; she has received research grant support from all of these. M. Kapral, A. Kulkarni, and A. Davis report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

Received November 2, 2015. Accepted in final form April 7, 2016.

REFERENCES

1. Kuks JBM, Oosterhuis HJ. Clinical presentation and epidemiology of myasthenia gravis. In: Kaminski HJ, editor. *Myasthenia Gravis and Related Disorders*, 1st ed. Totowa, NJ: Humana Press; 2003:93–148.
2. Barnett C, Bril V, Kapral M, Kulkarni A, Davis AM. A conceptual framework for evaluating impairments in myasthenia gravis. *PLoS One* 2014;9:e98089.
3. Barohn RJ, McIntire D, Herbelin L, Wolfe GI, Nations S, Bryan WW. Reliability testing of the quantitative myasthenia gravis score. *Ann NY Acad Sci* 1998;841:769–772.
4. Burns TM, Conaway MR, Cutter GR, Sanders DB; Muscle Study Group. Construction of an efficient evaluative instrument for myasthenia gravis: the MG composite. *Muscle Nerve* 2008;38:1553–1562.
5. Wolfe GI, Herbelin L, Nations SP, Foster B, Bryan WW, Barohn R. Myasthenia gravis activities of daily living profile. *Neurology* 1999;52:1487–1489.
6. Romani A, Piccolo G, Bergamaschi R, Versino M, Cosi V. A reliability study of impairment and disability scales for myasthenia gravis patients. *Funct Neurol* 2002;17:137–144.
7. Sharshar T, Chevret S, Mazighi M, et al. Validity and reliability of two muscle strength scores commonly used as endpoints in assessing treatment of myasthenia gravis. *J Neurol* 2000;247:286–290.
8. Sanders DB, Tucker-Lipscomb B, Massey J. A simple manual muscle test for myasthenia gravis: validation and comparison with the QMG score. *Ann NY Acad Sci* 2003;998:440–444.
9. Mantegazza R, Antozzi C, Peluchetti D. Azathioprine as a single drug or in combination with steroids in the treatment of myasthenia gravis. *J Neurol* 1988;235:449–453.
10. Wolfe GI, Barohn RJ, Sanders DB, McDermott MP; Muscle Study Group. Comparison of outcome measures from a trial of mycophenolate mofetil in myasthenia gravis. *Muscle Nerve* 2008;38:1429–1433.
11. Barnett TC, Bril V, Davis AM. Performance of individual items of the quantitative myasthenia gravis score. *Neuromuscul Disord* 2013;23:413–417.
12. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U. S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;4:79. DOI: 10.1186/1477-7525-4-79.
13. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
14. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*, 1st ed. Cambridge, UK: Cambridge University Press; 2011.
15. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
16. Ellwart T, Konradt U. Formative versus reflective measurement: an illustration using work-family balance. *J Psychol* 2011;145:391–417.
17. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–1123.
18. Lai JS, Cella D, Choi S, et al. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Arch Phys Med Rehabil* 2011;92:S20–S27.
19. Burns TM, Wolfe GI, Conaway MR, Sanders DB. The MG-QOL15 for following the health-related quality of life of patients with myasthenia gravis. *Muscle Nerve* 2011;43:14–18.
20. Ware JE, Snow KK, Kosinski M, Gandek B; New England Medical Center Hospital. Health Institute. SF-36 Health Survey. Boston: Health Institute, New England Medical Center; 1993.
21. Vincent KA, Carr AJ, Walburn J, Scott DL, Rose MR. Construction and validation of a quality of life questionnaire for neuromuscular disease (INQoL). *Neurology* 2007;68:1051–1057.
22. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
23. Portney LG, Watkins MP. *Foundations of Clinical Research*, 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2009.
24. Nunnally JC Jr. *Psychometric Theory*, 2nd ed. New York: McGraw Hill; 1978.
25. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–268.
26. Crosby RD, Kolotkin RL, Williams GR. An integrated method to determine meaningful changes in health-related quality of life. *J Clin Epidemiol* 2004;57:1153–1160.
27. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med* 1987;6:441–448.
28. Streiner DL, Norman GR. *Health Measurement Scales*, 4th ed. New York: Oxford University Press; 2008.
29. Geoffrey R, Streiner DLN. *Biostatistics*. In: *The Bare Essentials*, 2nd ed. Hamilton, ON: B.C. Decker; 2000.
30. Feinstein AR, Cicchetti DV. High agreement but low kappa: I: the problems of two paradoxes. *J Clin Epidemiol* 1990;43:543–549.
31. O'Leary S, Lund M, Ytre-Hauge TJ, et al. Pitfalls in the use of kappa when interpreting agreement between multiple raters in reliability studies. *Physiotherapy* 2014;100:27–35.
32. World Health Organization. *International Classification of Functioning, Disability and Health (ICF)*, 1st ed. Geneva: World Health Organization; 2001.