CLINICAL RESEARCH ARTICLE

MUSCLE&NERVE WILEY

Visual versus quantitative analysis of muscle ultrasound in neuromuscular disease

Juerd Wijntjes MD ⬛ | Joris van der Hoeven MD | Christiaan G. J. Saris MD PhD |
Jonne Doorduin PhD | Nens van Alfen MD PhD ⬛

Department of Neurology and Clinical
Neurophysiology, Donders Institute for Brain,
Cognition and Behaviour, Radboud University
Medical Center, Nijmegen, The Netherlands

Correspondence
Juerd Wijntjes, Department of Neurology and
Clinical Neurophysiology, Donders Institute for
Brain, Cognition and Behavior, Radboud
University Medical Center, Route 920, PO Box
9101, 6500 HB Nijmegen, The Netherlands.
Email: juerd.wijntjes@radboudumc.nl

Abstract

Introduction/Aims: Visual and quantitative muscle ultrasound are both valid diagnostic tools in neuromuscular diseases. To optimize muscle ultrasound evaluation and facilitate its use in neuromuscular disease, we examined the correlation between visual and quantitative muscle ultrasound analysis and their pitfalls.

Methods: Retrospective data from 994 patients with 13,562 muscle ultrasound images were analyzed. Differences in echogenicity z-score distribution per Heckmatt grade and corresponding correlation coefficients were calculated.

Results: Overall, there was a correlation of 0.60 between the two scoring systems, with a gradual increase in z-score with increasing Heckmatt grades and vice versa. Patients with a neuromuscular disorder had higher Heckmatt grades ($p < 0.001$) and z-scores (median z-score = 0.30, $p < 0.001$) than patients without. The highest Heckmatt grades and z-scores were found in patients with either a dystrophy or inflammatory myopathy (both median Heckmatt grade of 2 and median z score of 0.74 and 1.20, respectively). Discrepant scores were infrequent (<2%), but revealed important pitfalls in both grading systems.

Discussion: Visual and quantitative muscle ultrasound are complementary techniques to evaluate neuromuscular disease and have a moderate positive correlation. Importantly, we identified specific pitfalls for visual and quantitative muscle ultrasound and how to overcome them in clinical practice.

KEYWORDS
echogenicity, Heckmatt grading, muscle ultrasound, neuromuscular disorders, quantitative analysis

1 | INTRODUCTION

Muscle ultrasound (MUS) is gradually becoming more accepted as a diagnostic and longitudinal monitoring modality in neuromuscular diseases (NMDs).[1] To facilitate its use, it is important to have a clear, standardized, and accurate method to evaluate images.[2] Both visual

and quantitative assessments can be used to measure the degree of muscle pathology. Heckmatt et al. introduced a four-point grading scale for visual assessment of MUS images.[3] Previous studies showed that the Heckmatt scale has moderate to good diagnostic values for detecting NMD.[4,5] In contrast, quantitative MUS (QMUS) measures the mean grayscale value of the muscle region of interest (ROI), and compares this muscle echogenicity to a reference value. Subsequently, a standardized echogenicity (z-score) can be calculated, which denotes the number of SDs of the measured echogenicity from the predicted echogenicity.[6] Compared to visual Heckmatt grading this improves the diagnostic value from ~70% to ~90%.[7,8] Also the inter-rater reliability is greater for QMUS than for the visual assessment.[4] Unfortunately, QMUS is critically dependent on the hardware of ultrasound systems and post-processing. Differences in each of these alter echogenicity, so that images and reference values from different systems or probes cannot be directly compared. In practice, this means that new reference values need to be obtained for every device and preset, which is time-consuming. This limits the widespread use of QMUS.[2,6]

To improve the value of visual MUS evaluation and make it more useful for neuromuscular practitioners who do not have access to QMUS, knowledge about the relation between QMUS and visual MUS evaluation is valuable. To date, only a few studies have evaluated the relation between both methods in a small population of patients with confirmed facioscapulohumeral muscular dystrophy (FSHD), and reported correlations of around 0.8.[9,10] In the current study, we assessed the correlation between the Heckmatt grade and z-score in an extensive set of muscles of a large sample of patients who were screened for a wide range of different suspected NMD. Finally, we examined discrepancies between visual assessment and QMUS to identify pitfalls for both methods and how to avoid them in general practice.

2 | METHODS

2.1 | Patients

This retrospective observational study was performed at the department of Neurology of the Radboud university medical center Nijmegen, The Netherlands. Our center is a European Reference Centre for NMD, with a focus on FSHD patients. A consecutive sample was collected between May 2017 and August 2020 of patients of every age with symptoms suspected of a neuromuscular disorder, who were referred for a routine diagnostic MUS screening for NMD. When patients underwent multiple MUS studies during this period, only data from their first examination were used.

2.2 | Ultrasound acquisition

Neurodiagnostic technicians performed the MUS studies using an Esaote MyLabTwice ultrasound system (Esaote SpA, Genova, Italy) with a LA533 3–13 MHz broadband linear transducer with a 53 mm

footprint. A strictly standardized ultrasound preset, probe, and measurement protocol were used in all individuals to ensure comparability between measurements. Depending on the referral question, one of four different screening protocols was used (see Supporting Information Methods, which are available online).[1] A subset of the patients underwent MUS using a research-specific protocol, as described previously.[4]

2.3 | Image analysis

For the visual assessments, one of the observers (C.S., J.W., N.v.A.) graded the ultrasound images for each muscle using the Heckmatt scale before taking note of the quantitative results.[3] All observers were board-certified neurologists and clinical neurophysiologists with neuromuscular expertise. Both C.S. and J.W. are experienced (3–5 years) ultrasound observers, and N.v.A. has extensive experience (20 years) in the assessment of MUS images.

We used a custom software package developed in Matlab (R2013b, Mathworks, Natick, MA) for quantitative analysis, following a previously described protocol.[6] Our center has developed prediction models using linear regression for echogenicity for a large set of muscles in healthy subjects. In these models age, weight, height, body mass index (BMI), sex, and handedness are used as predictors. Measured echogenicity within the ROI is compared to the predicted echogenicity to calculate a z-score (the number of SDs that the measured echogenicity is from the predicted echogenicity). An echogenicity z-score greater than or equal to 2.0 was considered abnormal. The ROI was measured as the maximum size of the muscle that can be included in the image by default, but was adjusted by the observer if needed (in case of severe attenuation).

2.4 | Clinical assessment

The final clinical diagnosis was established by each patient's treating neurologist, based on a combination of the history, physical examination and ancillary diagnostic tests that included the ultrasound results. These diagnoses were first classified in three groups: "no neuromuscular disorder", "uncertain or unknown", or "neuromuscular disorder." A diagnosis was considered uncertain if the clinician had described it as such, or when no final diagnosis had been reached yet at the time of analysis. We subdivided the neuromuscular disorders in four categories based on their localization: myopathic, neurogenic, neuromuscular junction, or a NMD without muscle involvement (eg, sensory neuronopathy). In the latter group, we expected no muscle involvement and, therefore, no abnormalities of the MUS images. Finally, we subcategorized the myopathic and neurogenic disorders into different etiologic subgroups: FSHD, muscular dystrophies other than FSHD, congenital myopathies, metabolic myopathies, inflammatory myopathies/myositis, other myopathies, motor neuron diseases (MNDs), radiculopathies, mononeuropathies, polyneuropathies, and plexopathies.

TABLE 1 Patient characteristics

Total	N = 994
Sex	Male n = 539 (54.2%)
	Female n = 455 (45.8%)
Age (mean ± SD [range])	45.78 y ±21.92 [0–88]
<5 y	50 (5%)
6–18 y	112 (11.3%)
19–75 y	776 (78.1%)
>75 y	56 (5.6%)
BMI (mean ± SD [range])	24.54 kg/m^2 ± 5.26 [11.5–42.7]
Final diagnosis	N (percentage)
No NMD	370 (37.2%)
Uncertain or unknown	91 (9.2%)
Neuromuscular disorder	533 (53.6%)
Myopathic	304 (57.0%)
FSHD	128
Inflammatory myopathies	86
Other myopathies	33
Congenital myopathies	22
Metabolic myopathies	20
Dystrophies other than FSHD	15
Neurogenic	211 (39.6%)
Radiculopathies	64
MNDs	61
Polyneuropathies	44
Plexopathies	31
Mononeuropathies	11
Neuromuscular junction disorders	9 (1.7%)
NMD without muscle involvement	9 (1.7%)
Benign fasciculation syndrome	7
Small fiber neuropathy	2

2.5 | Statistical analysis

Prior to the analysis, double data entries were excluded. Ultrasound measurements from the left and right side were pooled on the patient level per muscle. In case of unexpected findings, specifically for MUS images with a normal z-score and a Heckmatt grade 3 or 4, or vice versa muscles with a z-scores ≥3 and a Heckmatt grade 1, the original images were re-evaluated to seek an explanation for this discrepancy. In rare instances, when the original Heckmatt grade needed adjustment, based on the image characteristics, grading was done by consensus between all observers.

An overall comparison was made between the median z-scores for each Heckmatt grade on all muscles in the dataset. Next, results were analyzed for the subgroups of different diagnostic categories.

Differences in z-score distributions were compared using the Kruskal Wallis H test, because of non-normality. Post-hoc analysis with the Mann–Whitney U test defined which of the Heckmatt grade subgroups had the highest median z-score. A Bonferroni correction was applied for multiple testing. Spearman rank correlation was used to establish the correlation coefficients between Heckmatt grade and z-score.

All statistical analyses were carried out using IBM SPSS v.25.0 (IBM, Armonk, NY). For all analyses the significance level was set at a p-value <0.05, and the p-value after Bonferroni correction for six sub-analyses was set to <0.00833.

2.6 | Ethical considerations

We conducted this study following the national and Helsinki guidelines for medical research. Patients were excluded if they had objected to the use of their de-identified personal information for further research, as noted in our electronic health record system. As this was a retrospective study with data collected during routine non-invasive diagnostic testing, per our institute's policy, no further ethical approval was needed.

3 | RESULTS

3.1 | Demographics

Data from the MUS examinations of 1100 patients were available for this study. Of these, 40 patients were excluded because they objected to the use of their data for further research, and 66 duplicate entries were removed. This resulted in a total of 994 MUS examinations from unique patients (84% adults) with 13,562 MUS images available for analysis, of whom 53.6% were diagnosed with a NMD. Demographic and diagnostic details are shown in Table 1. An overview of the MUS images per muscle is shown in Supporting Information Table S1.

3.2 | Images with discrepant z-scores versus Heckmatt grade

We revised 184 MUS images (1.4%) with a z-score lower than 2, that were scored Heckmatt grade 3 or 4. Most of these images (86) were from FSHD patients with severely dystrophic muscles. In 41 images, an inadequate ROI placement was found that was not adjusted for attenuation. Another discrepancy was encountered in 36 images, consisting of patchy or inhomogeneous echogenicity abnormalities. Less frequently, we noticed observer grading errors. These discrepancies were most often found (144/184 = 78.3%) in lower limb muscles, especially the tibialis anterior muscle, vastus lateralis, and the medial head of the gastrocnemius.

The reverse situation of a z-score ≥3 and a normal Heckmatt grade 1, was seen in 50 (0.4%) MUS images. This discrepancy was

most commonly seen in upper extremity muscles (40/50), of which 30 images were from either the flexor carpi radialis (FCR), biceps brachii, or deltoid muscle. In 16 of 19 images of the FCR, we found a so-called "background effect" optical illusion (also see Figure 4E and an explanation of the background effect in the Discussion section). In the remaining images, the discrepant score was attributed to a visual underestimation of the grade by the observer.

As stated in the Methods section, in 12 different images (0.09%), the original Heckmatt grade was adjusted from 1 to 2 based on the image characteristics, prior to further analysis.

3.3 | Overall comparison of visual versus QMUS

The distribution of z-scores per Heckmatt grade in the total population is shown in Figure 1. Using the visual assessment, 72.2% of the MUS images were scored a Heckmatt grade 1 and had a median z-score of −0.50 (interquartile range [IQR] −1.30–0.40). In 21.1% a Heckmatt grade 2 was found, with a median z-score of 1.50 (IQR 0.50–2.60). Heckmatt grade 3 was found in 6.2% with a median z-score of 3.90 (IQR 2.30–5.60) and Heckmatt grade 4 was rarely found in only 76 muscles (0.6%) with a median z-score of 2.80 (IQR 1.70–6.20). There was a significant positive correlation of 0.60 (p < 0.001) between the z-score and Heckmatt grade. Z-scores increased significantly with increasing Heckmatt grades (all pairwise comparisons p < 0.001), except for the z-scores in Heckmatt grade 3 and 4, between which no significant difference were found.

3.4 | Heckmatt grade versus z-score in patients with and without a NMD

The relation between z-scores and Heckmatt grades of muscles from patients with a final diagnosis of NMD, no-NMD, or an uncertain diagnosis is shown in Figure 2 and Supporting Information Table S2.

The majority of MUS images in patients without a NMD were Heckmatt grade 1 and had a z-score lower than 2 (median z = −0.40). In this group, a significant difference was seen in the distribution of z-scores for the different Heckmatt grades (χ^2 [3] = 440.084, p < 0.001), in which the median z-score increased significantly with increasing Heckmatt grade (p < 0.001). However, similar to the overall group, the z-scores for Heckmatt grade 4 did not increase significantly compared to those of lower Heckmatt grades. A weak but significant positive correlation of 0.32 was found between Heckmatt grade and z-score in patients with no-NMD.

For patients diagnosed with a NMD, MUS images were scored a significantly higher Heckmatt grade (p < 0.001) and z-score (median z = 0.30, p < 0.001) compared to patients without a NMD. The distribution of z-scores was different between Heckmatt grades 1–4 in this group too (χ^2 [3] = 3688.486, p < 0.001), with a significant strong correlation coefficient between z-score and Heckmatt grade of 0.67. Pairwise analysis showed a significant increase in median z-score with every incremental Heckmatt grade (p < 0.001), but the z-scores for Heckmatt grade 3 and 4 did not differ (p = 0.036, Bonferroni correction applied).

In patients with an uncertain diagnosis, the median Heckmatt grade and median z-score were between those of patients with and without a NMD, with similarly distributed z-scores between Heckmatt grades as non-NMD patients.

3.5 | Heckmatt grade and z-score in myopathic and neurogenic NMDs

In patients with a myopathic NMD, the median Heckmatt grade and the median z-score were significantly higher (median z = 0.80) than in patients with a neurogenic disorder (median z = −0.16, all p < 0.001). For both patients with myopathic and neurogenic NMDs, the distribution of z-scores was statistically different for each Heckmatt grade (χ^2 (3) = 2144.438, p < 0.001 and χ^2 (3) = 1116.873, p < 0.001,
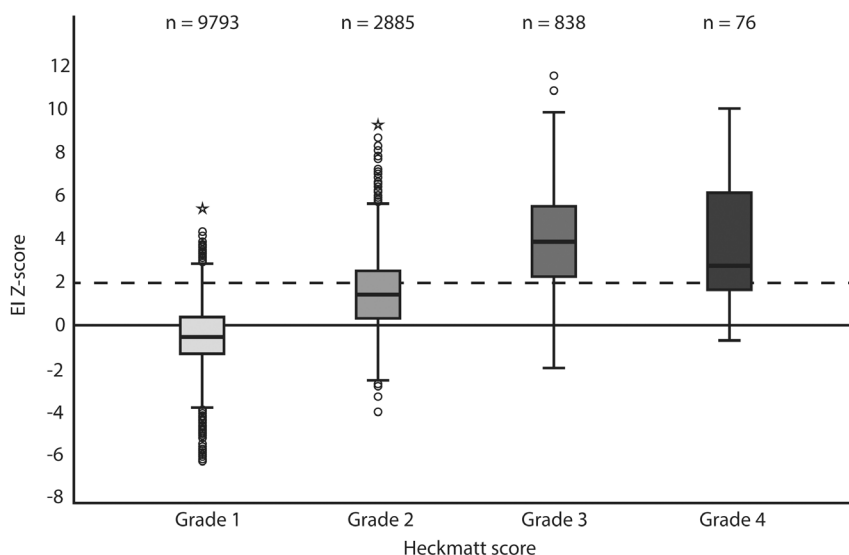

FIGURE 1 Z-scores per Heckmatt grade. The horizontal bar inside the boxes indicate the median and the lower and upper ends represent the first and third quartiles. The whiskers indicate values within 1.5× the IQR from the upper or lower quartile (or minimum and maximum if within 1.5× IQR of the quartiles). Circles and stars show values respectively greater than 1.5× or 3.0× IQR from the upper or lower quartile. The broken line refers to the threshold for abnormal z-scores. The number of muscles per Heckmatt score is shown

FIGURE 2 Z-score per Heckmatt grades in the no NMD, uncertain or unknown, and NMD category. The number of muscles per Heckmatt score is shown
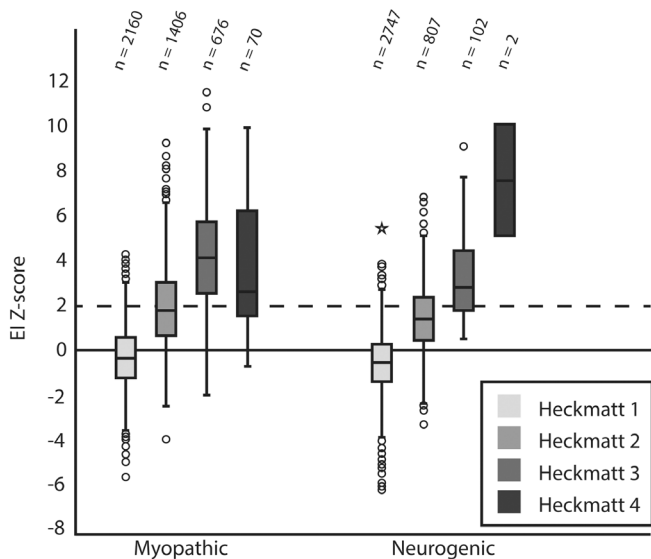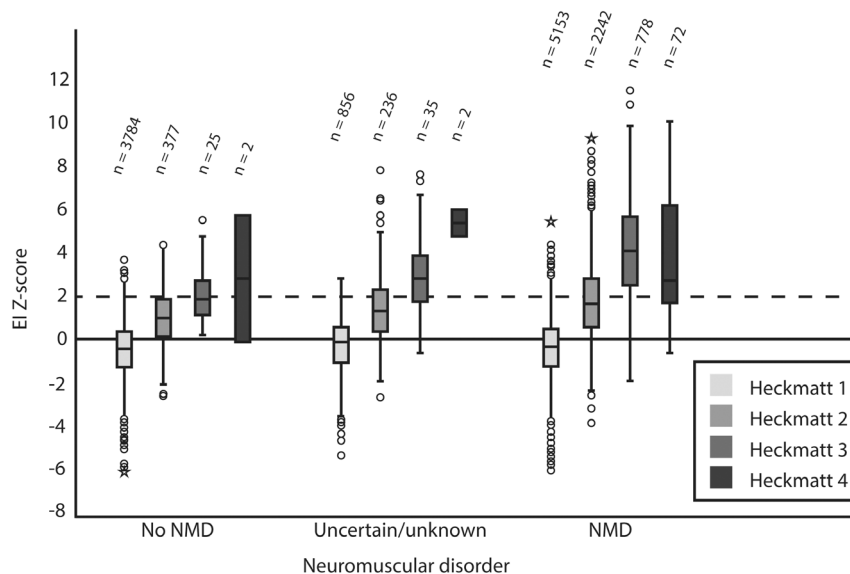

3.6 | Relation of Heckmatt grade and z-score in different myopathies

MUS images of 55.3% of the patients with FSHD and 52.8% of the patients with inflammatory myopathies were scored a Heckmatt grade 2 or higher. Muscles from patients with inflammatory myopathies had the highest group median z-score of 1.20, while patients with FSHD had the next highest group median z-score of 0.74. In metabolic myopathies, the median z-score was found to be lower (0.50). For all myopathies, the differences in z-score distribution between Heckmatt grades and post-hoc comparison were all highly significant, but the correlation coefficients varied (Supporting Information Table S2). However, in FSHD patients, pairwise comparison showed a significantly higher median z-score in muscles with Heckmatt grade 3 compared to Heckmatt grade 4 ($p < 0.0001$). For inflammatory myopathies and other myopathies no significant difference in z-score distribution could be established for Heckmatt grade 4 compared to grade 2 and grade 3. See Supporting Information Figure S1.

FIGURE 3 Z-score per Heckmatt grade specified to myopathic or neurogenic NMD. The number of muscles per Heckmatt score is shown

3.7 | Relation of Heckmatt grade and z-score in neurogenic NMD

Only 2 of 3658 MUS images were scored a Heckmatt grade 4 in patients with neurogenic disorders. The distribution of z-scores across the Heckmatt grades differed significantly for each Heckmatt grade in MNDs, radiculopathies, and polyneuropathies, with a significant increase in z-scores for each Heckmatt grade compared to the lower grades ($p < 0.001$), with moderately positive correlation coefficients. See Supporting Information Figure S2.

respectively). Patients with a myopathic NMD showed a strong positive correlation of 0.70 ($p < 0.001$) with significantly increased median z-scores with increasing Heckmatt grades, except for muscles with a Heckmatt grade 4, in which the median z-score was significantly lower than in Heckmatt grade 3 ($p = 0.006$). In patients with a neurogenic disorder, a moderate positive correlation was found between z-scores and Heckmatt grades ($r_s = 0.55$, $p < 0.001$) with significantly increased median z-scores with increasing Heckmatt grades, except that no significant difference between Heckmatt grade 4 and lower Heckmatt grades was found. See Figure 3.

4 | DISCUSSION

In this study, we found a moderate positive correlation of 0.60 between the visual and the quantitative scoring system. We found a small percentage of muscles (<2%) with a clearly discrepant Heckmatt grade compared to their z-scores. Overall, there was a gradual increase of the z-score with increasing Heckmatt grades and vice versa. This correlation increased to a strong correlation of 0.7 in patients with a myopathic NMD.

4.1 | Correlation between Heckmatt grade and z-score

In contrast to the overall correlation between z-scores and Heckmatt grades, we found no difference in z-scores between Heckmatt grade 3 and 4. Of note, even in this sizable sample, muscles were rarely scored a Heckmatt grade 4. This was likely at least in part caused by an inclusion bias, as most patients in our cohort were scanned as part of their diagnostic workup and were, thus, less likely to have long-standing or severe disease. This presumption is further supported by the finding that most Heckmatt grade 4 MUS images were seen in the patients with FSHD, in whom the diagnosis was already established. This predominance of dystrophic muscles in the Heckmatt grade 4 category may also be the cause of the lack of a difference in z-scores between Heckmatt grade 3 and 4, and even the apparent decrease in z-score in grade 4 compared to grade 3. This finding is known and can be attributed to the phenomenon that fully fatty-degenerated muscles again have a low echogenicity and, thus, lower z-scores.[2] It is very likely also inherent to the Heckmatt grading system, because the strong attenuation that defines grade 4 automatically implies that echogenicity will progressively decrease in the deeper parts of the muscle.[2]

4.2 | Patients categorized as no neuromuscular disorder but high Heckmatt grade

In two patients in the "no NMD" category, the medial gastrocnemius muscle scored a Heckmatt grade 4. The first patient was referred for exercise intolerance. The medical history included a lumbosacral radiculopathy from which she had recovered. This could explain the abnormal MUS finding. The second patient was referred because of exercise intolerance of the proximal leg muscles. Medical history included a non-Hodgkin lymphoma and prostate cancer for which he received chemotherapy. Neurological examination showed decreased vibration sensation in both feet but intact tendon reflexes. Even though the clinical diagnosis of polyneuropathy was not made, we hypothesize a subclinical polyneuropathy could explain the abnormal MUS finding. In both patients, no final clinical diagnosis of a NMD was established by their treating neurologist. The final clinical diagnosis used for our categorization was based on a combination of the history, physical examination, and ancillary diagnostic tests. Although this reflects clinical practice, it also led to the inclusion of these patients in the "no NMD" group. The neuromuscular involvement in both patients was considered clinically irrelevant, but most likely explains the MUS abnormalities. Both cases emphasize the need to take the medical history into account when interpreting muscle imaging results.

4.3 | Neurogenic versus myopathic NMDs

Muscles from patients with neurogenic NMDs had on average lower Heckmatt grades and z-scores compared to muscles from patients with myopathic disorders, and less strong correlations between the Heckmatt grades and z-scores, with the exception of MND. This is a relevant observation for the use of MUS in practice, as it implies that the sensitivity for neurogenic disorders other than MND will be less than for myopathic changes. A reason for this decreased sensitivity could be that, in monophasic axonal injury, collateral reinnervation will result in a normal appearance of the muscle again.[2] Also, changes in muscle texture in neurogenic disorders often have a more heterogenous distribution of abnormalities throughout the muscle as motor unit territories become affected.[7] We confirmed that the diagnostic value of MUS is limited in disorders with little structural damage in the muscle, such as metabolic myopathies.[11,12] A drawback of our study design focusing on individual muscles is that no disease severity measure could be reported for patients. We realize that the lack of clinical correlations with MUS findings limits the evaluation and clinical interpretation of the analysis of the different neurogenic and myopathic subcategories.

TABLE 2 Pitfalls and suggested solutions for different MUS analyzing methods

Pitfall	Muscles	Solution
Visual assessment		
Background optical illusion effect	Upper extremity, mainly FCR	Combine visual with quantitative assessment for objective measurements and prevent underestimation of EI
Observer estimation error	Tibialis anterior, rectus abdominis	Familiarize the observer with different effects of age, BMI, and sex on MUS images; elderly, higher BMI, and female sex will result in higher EI
Quantitative assessment		
Fatty degeneration	Tibialis anterior, medial gastrocnemius	Exclude muscle parts with fatty degeneration in ROI and note visual finding in final results
Inadequate ROI placement	Vastus lateralis, tibialis anterior	Exclude the (deeper) parts of muscles with lower EI due to attenuation artifact
Inhomogeneous EI structure	All	Note this finding as a remark in final results and/or use visual assessment

Abbreviations: EI, echogenicity.

4.4 | Discrepancies between Heckmatt grade and z-score

Revision of the images with a clear discrepancy between the Heckmatt grade and z-score indicated that there were a few major areas of error that should lead to caution. We provide an overview of pitfalls and how to overcome them in Table 2 and Figure 4.

A spuriously low z-score was most often seen in patients with FSHD with severely dystrophic muscles that appeared to be completely fatty degenerated, which resulted in a low image echogenicity within the normal range, akin to that of the subcutaneous fat layer (Figure 4A).[2] Incorrect drawing of too large an ROI in muscles with severe attenuation of their deeper parts and, hence, reduced echogenicity (Figure 4B) resulted in a discrepant low z-score. Patchy or inhomogeneous echogenicity abnormalities resulted in a visually clearly abnormal muscle image, but with a less pronounced increase of the overall muscle z-score (Figure 4C). Observer estimation errors

were seen in incorrectly assessing the influence of age or sex on a particular muscle (Figure 4D) and patients with a high BMI, in whom the increased intramuscular fat content leads to a higher echogenicity, but a normal z-score because of correction for weight. For any visual analysis method, this type of error can only be overcome by the observer familiarizing themself with how images from different muscles in different patients can look. However, this type of error was encountered in only 0.5% of the images evaluated, which indicates that the diagnostic performance of a single observer will likely be adequate for the use of MUS in clinical practice and research settings.

A spurious low z-score was most often found in the lower limb muscles. Because of this, we suggest that z-scores of the lower limb muscles should be interpreted with caution, always taking the visual evaluation of their muscle texture into account.

Muscles with a high z-score but normal Heckmatt grade were most often seen in the upper limbs. The main pitfall in these images was the background optical illusion effect (Figure 4E). This effect
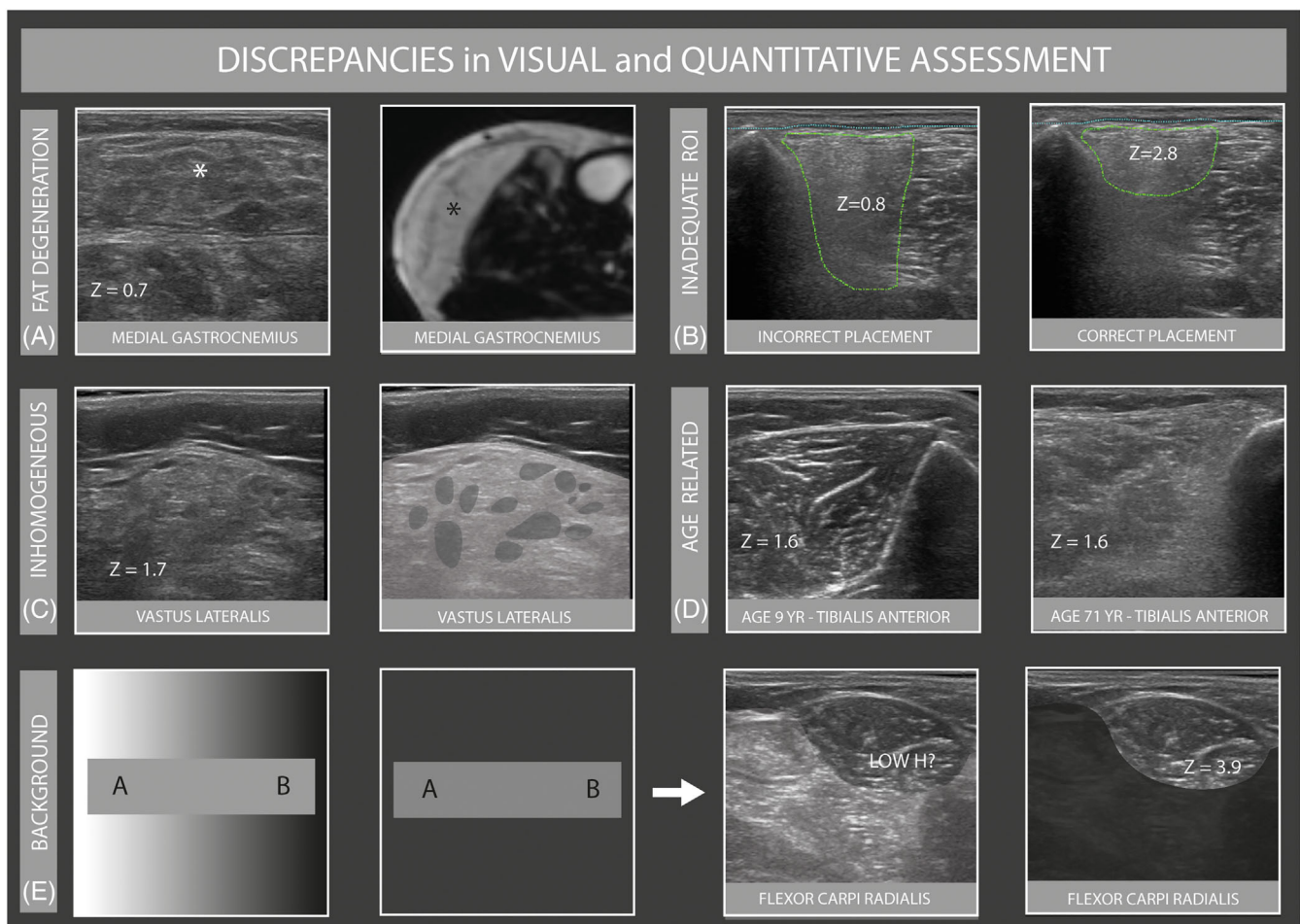

FIGURE 4 Discrepancies in visual and quantitative assessment. A, Ultrasound image of the medial gastrocnemius (* asterisk) with normal z score of a patient with FSHD. Corresponding MR image shows complete fatty infiltration. B, Inadequate placement of ROI in tibialis anterior with attenuation, with increase of z-score after redrawing. C, Inhomogeneous muscle texture in the vastus lateralis of a patient with MND resulting in a normal z-score. D, Age related texture changes of the tibialis anterior in a 9 and 71-year-old patient with same echogenicity z-score but clearly different Heckmatt grades. E, "Background effect optical illusion" displayed schematically in the first two boxes (both gray areas A and B in the first box have the same shade of gray; this can be seen in the exact same gray bar in the second box) and in FCR in the last two boxes (before and after removal of the transition with bright contrast)

makes it challenging for a human observer to assess if the echogenicity is abnormal when the muscle looks blacker than everything else around it. We think this effect is not a feature of the upper limb per se, but will be most evident when multiple smaller muscles in the forearm, some normal and others abnormal, are examined in the same image. To overcome this limitation, quantitative analysis provides a more objective approach to identify abnormality.

4.5 | Future directions

Visual and QMUS are useful and complementary, but it seems worthwhile to investigate the use of other methods to facilitate more accurate MUS image interpretation by the clinician. For visual evaluation, the difference between Heckmatt grade 1 and 2 can be very subtle. A slightly adapted and more straightforward visual grading tool might be more effective.[2] One method could be to use a simple three-point scale that scores the MUS images as "normal", "abnormal" or "uncertain." In this case, the observer can use all available information, such as the Heckmatt parameters of overall echogenicity, attenuation, and texture, but also focal image abnormalities, patient characteristics, and even the a priori chance of having a NMD to evaluate the image. For pure diagnostic purposes, we expect such a scoring system to give higher accuracy, especially for relatively inexperienced observers and across different observers. A drawback of this simpler scoring system may be the loss of grading progressive muscle abnormalities that would be useful in a follow-up setting. To optimize the use of visual analysis as a clinical outcome measure for specific NMDs, Rasch analysis could be used to transform the ordinal scoring system into quantitative interval scores.[13,14] Of course, such a scoring system should be validated in further studies and the diagnostic accuracy should be compared with the current quantitative and semi-quantitative Heckmatt grading method. To optimize quantitative scoring and overcome device dependency, the use of deep learning tools for automated muscle segmentation and detection of abnormalities could be an important step forward.[15–17]

5 | CONCLUSIONS

We found a moderate but significant positive correlation between visual and quantitative analysis of MUS in NMD. We identified rare but important pitfalls when using either method and described how to overcome them in clinical practice. We recommend using both techniques when evaluating MUS images for optimal sensitivity and accuracy.

CONFLICT OF INTEREST

Nens van Alfen acts as a muscle ultrasound consultant for Dynacure and performs editorial services for Wiley Publishing; all payments go to their employer. Juerd Wijntjes, Joris van der Hoeven, Christiaan Saris, and Jonne Doorduin all have nothing to declare.

ETHICS STATEMENT

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Juerd Wijntjes https://orcid.org/0000-0002-5140-0683
Nens van Alfen https://orcid.org/0000-0001-7839-8125

REFERENCES

1. Mah JK, van Alfen N. Neuromuscular ultrasound: clinical applications and diagnostic values. *Can J Neurol Sci*. 2018;45:605-619. doi:10.1017/cjn.2018.314
2. Wijntjes J, Alfen N. Muscle ultrasound: present state and future opportunities. *Muscle Nerve*. 2021;63:455-466. doi:10.1002/mus.27081
3. Heckmatt JZ, Leeman S, Dubowitz V. Ultrasound imaging in the diagnosis of muscle disease. *J Pediatr*. 1982;101:656-660. doi:10.1016/S0022-3476(82)80286-2
4. Pillen S, van Keimpema M, Nievelstein RAJ, Verrips A, van Kruijsbergen-Raijmann W, Zwarts MJ. Skeletal muscle ultrasonography: Visual versus quantitative evaluation. *Ultrasound Med Biol*. 2006;32:1315-1321. doi:10.1016/j.ultrasmedbio.2006.05.028
5. Zuberi SM, Matta N, Nawaz S, Stephenson JBP, McWilliam RC, Hollman A. Muscle ultrasound in the assessment of suspected neuromuscular disease in childhood. *Neuromuscul Disord*. 1999;9:203-207. doi:10.1016/S0960-8966(99)00002-4
6. van Alfen N, Mah JK. Neuromuscular ultrasound: a new tool in your toolbox. *Can J Neurol Sci*. 2018;45:504-515. doi:10.1017/cjn.2018.269
7. Zaidman CM, Van Alfen N. Ultrasound in the assessment of myopathic disorders. *J Clin Neurophysiol*. 2016;33:103-111. doi:10.1097/WNP.0000000000000245
8. Boon AJ, Wijntjes J, O'Brien TG, Sorenson EJ, Cazares Gonzalez ML, Alfen N. Diagnostic accuracy of gray scale muscle ultrasound screening for pediatric neuromuscular disease. *Muscle Nerve*. 2021;1–9:50-58. doi:10.1002/mus.27211
9. Mul K, Horlings CGC, Vincenten SCC, Voermans NC, van Engelen BGM, van Alfen N. Quantitative muscle MRI and ultrasound for facioscapulohumeral muscular dystrophy: complementary imaging biomarkers. *J Neurol*. 2018;265:2646-2655. doi:10.1007/s00415-018-9037-y
10. Goselink RJM, Schreuder THA, Mul K, et al. Muscle ultrasound is a responsive biomarker in facioscapulohumeral dystrophy. *Neurology*. 2020;94:e1488-e1494. doi:10.1212/WNL.0000000000009211
11. Brockmann K, Becker P, Schreiber G, Neubert K, Brunner E, Bönnemann C. Sensitivity and specificity of qualitative muscle ultrasound in assessment of suspected neuromuscular disease in childhood. *Neuromuscul Disord*. 2007;17:517-523. doi:10.1016/j.nmd.2007.03.015
12. Pillen S, Morava E, Van Keimpema M, et al. Skeletal muscle ultrasonography in children with a dysfunction in the oxidative phosphorylation system. *Neuropediatrics*. 2006;37:142-147. doi:10.1055/s-2006-924512
13. Rasch G. *Probabilistic Models for some Intelligence and Attainment Tests*. Danish Institute for Educational Research; 1960.
14. Mul K, Horlings CGC, Faber CG, van Engelen BGM, Merkies ISJ. Rasch analysis to evaluate the motor function measure for patients with facioscapulohumeral muscular dystrophy. *Int J Rehabil Res*. 2021;44:38-44. doi:10.1097/MRR.0000000000000444

15. Burlina P, Billings S, Joshi N, Albayda J. Automated diagnosis of myositis from muscle ultrasound: exploring the use of machine learning and deep learning methods. *PLoS One*. 2017;12:1-15. doi:10.1371/journal.pone.0184059

16. Gijsbertse K, Bakker M, Sprengers A, et al. Computer-aided detection of fasciculations and other movements in muscle with ultrasound: development and clinical application. *Clin Neurophysiol*. 2018;129:2567-2576. doi:10.1016/j.clinph.2018.09.022

17. Marzola F, van Alfen N, Doorduin J, Meiburger KM. Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment. *Comput Biol Med*. 2021;135:104623. doi:10.1016/j.compbiomed.2021.104623

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wijntjes J, van der Hoeven J, Saris CGJ, Doorduin J, van Alfen N. Visual versus quantitative analysis of muscle ultrasound in neuromuscular disease. *Muscle & Nerve*. 2022;66(3):253-261. doi:10.1002/mus.27669