Systematic Review

Muscle Sonography in Inclusion Body Myositis: A Systematic Review and Meta-Analysis of 944 Measurements

Ramy Abdelnaby 1, Khaled Ashraf Mohamed 2, Anas Elgenidy 2, Yousef Tarek Sonbol 2, Mahmoud Mostafa Bedewy 2, Aya Moustafa Aboutaleb 3, Mohamed Ayman Ebrahim 2, Imene Maallem 4, Khaled Tarek Dardeer 2, Hamed Amr Heikal 2, Hazem Maher Gawish 2 and Jana Zschüntzsch 5,*

1 Department of Neurology, RWTH Aachen University, Pauwels Street 30, 52074 Aachen, Germany; rabdelnaby@ukaachen.de
2 Faculty of Medicine, Cairo University, 1 Gamaa Street, Cairo 12613, Egypt; Khaled74ashraf@gmail.com (K.A.M.); anas.elgenidy@gmail.com (A.E.); yousetarek15@gmail.com (Y.T.S.); mahmoudbedewy11@gmail.com (M.M.B.); mohdaymanebrahim@gmail.com (M.A.E.); khaledegy99@gmail.com (K.T.D.); Hamedhikel18@gmail.com (H.A.H.);
3 Faculty of Medicine, Zagazig University, Zagazig 44519, Egypt; ayam78801@gmail.com
4 Faculty of Medicine, Pharmacy Department, University Badji Mokhtar Annaba, Zaafrania Street, Annaba 23000, Algeria; imene.maallem@yahoo.com
5 Clinic for Neurology, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany
* Correspondence: j.zschuentzsch@med.uni-goettingen.de

Abstract: Inclusion body myositis (IBM) is a slowly progressive muscle weakness of distal and proximal muscles, which is diagnosed by clinical and histopathological criteria. Imaging biomarkers are inconsistently used and do not follow international standardized criteria. We conducted a systematic review and meta-analysis to investigate the diagnostic value of muscle ultrasound (US) in IBM compared to healthy controls. A systematic search of PubMed/MEDLINE, Scopus and Web of Science was performed. Articles reporting the use of muscle ultrasound in IBM, and published in peer-reviewed journals until 11 September 2021, were included in our study. Seven studies were included, with a total of 108 IBM and 171 healthy controls. Echogenicity between IBM and healthy controls, which was assessed by three studies, demonstrated a significant mean difference in the flexor digitorum profundus (FDP) muscle, which had a grey scale value (GSV) of 36.55 (95% CI, 28.65–44.45, \( p < 0.001 \)), and in the gastrocnemius (GC), which had a GSV of 27.90 (95% CI 16.32–39.48, \( p < 0.001 \)). Muscle thickness in the FDP showed no significant difference between the groups. The pooled sensitivity and specificity of US in the differentiation between IBM and the controls were 82% and 98%, respectively, and the area under the curve was 0.612. IBM is a rare disease, which is reflected in the low numbers of patients included in each of the studies and thus there was high heterogeneity in the results. Nevertheless, the selected studies conclusively demonstrated significant differences in echogenicity of the FDP and GC in IBM, compared to controls. Further high-quality studies, using standardized operating procedures, are needed to implement muscle ultrasound in the diagnostic criteria.

Keywords: muscle ultrasound; inclusion body myositis; meta-analysis; idiopathic inflammatory myopathy; neuromuscular disorder

1. Introduction

Inclusion body myositis (IBM) is one of the most common subtypes of idiopathic inflammatory myopathies (IIM) [1], primarily affecting those above 45 years of age. It has a progressive course and affects skeletal muscles with a distinct pattern, [2] causing asymmetric muscle weakness [3–5] mainly in finger flexors and/or quadriceps muscles [6].
This combination of weakness often results in loss of ambulation and independence, as well as the need for assertive devices and increased supportive care over the duration of the disease. Dysphagia is also frequent in patients, which leads to swallowing disorders and an increased risk of aspiration pneumonia [7,8]. The prevalence of IBM was estimated to be between 24.8–46 patients per 1,000,000, with a male to female ratio of 2:1 [9,10]. Overall life expectancy is still shorter in IBM patients compared to the general population [11]. Although the pathophysiology of IBM has not yet been clearly elucidated [12], several factors, e.g., genetic, aging, immunologic and mitochondrial dysfunction, have been suggested to play a role [13–20].

Due to the insidious character and slow progression of the disease, there is often an estimated delay between 3–5 years in reaching a diagnosis of IBM, [13–21], which makes the diagnosis difficult in the early stages [22]. Currently, the diagnosis is based on clinical presentation, CK-values and histopathological findings, such as simultaneous presence of inflammatory and degenerative changes [23]. Antibodies against cytosolic 5’-nucleotidase (cN) -1A (anti-5NT1A/5NTC1A; alternatively: muscle protein 44; Mup44) can be present in 30–80% of all IBM patients and is often associated with more severe dysphagia [24].

Nevertheless, the diagnosis of IBM may be challenging as it might interfere with other subtypes of myopathies [21]. Hence, developing alternative and efficient diagnostic techniques became a necessity for ensuring early diagnosis of the disease. Muscle sonography has been described as an emerging tool for diagnosis of many muscle affections and their characteristic patterns [25,26], as well as for evaluating several inflammatory neuropathies [27,28].

Therefore, the aim of this study was to evaluate the current evidence and effectiveness of the muscle ultrasound (US) as a reliable method to investigate and diagnose IBM, as well as validating its sensitivity and specificity regarding its performance in improving patient management.

2. Materials and Methods

We performed a systematic search on the following databases: Medline (through PubMed), Scopus and Web of Science, up until 11 of September 2021. We used the following search strategy and we searched on the previously mentioned databases using the “title and abstract” domain in order to reach all the studies discussing the use of ultrasound to measure the muscle parameters in inclusion body myositis (IBM). The synonyms of our search strategy were retrieved from the MeSH database and were revised by a senior author (R.A). The terms were combined by “OR” and “AND” Boolean operators according to the method described in the Cochrane Handbook for Systematic Reviews of Interventions (Chapter 4.4.4) [29] as follows: (Ultrasonography OR ultrasound OR Ultrasonic OR Echotomography OR Sonography OR Sonographic OR Ultrasonographic OR Echography OR Ultrasonic) AND (Inclusion Body Myositis OR Inclusion Body Myositis OR Inclusion Body Myopathy). Our study was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [30]. The details of the search process, and the included studies, are shown in the following PRISMA diagram (Figure 1).

2.1. Selection Criteria

We included the full-text articles published in peer-reviewed journals that measured muscle parameters, such as muscle thickness and muscle echo intensity by ultrasound, in IBM patients diagnosed by muscle biopsy, and compared them to healthy controls. We also included studies that calculated the sensitivity and specificity of ultrasound in the diagnosis of IBM for the diagnostic test accuracy analysis. We excluded case reports, letters to the editor and studies that did not provide numerical data for muscle echo intensity or muscle thickness.
2.2. Selection and Screening

Four authors (A.E., A.M., M.B., Y.T.S.) screened the articles by title and abstract, two authors (A.M., M.B.) independently screened the articles by reading their full text and a third author (Y.T.S.) was referred to in case of a disagreement. Two authors (A.M. and M.B.) extracted data regarding patient characteristics, US devices used and outcome measures (muscle echo intensity and muscle thickness).

2.3. Quality Assessment

Two authors (M.B. and Y.T.S.) evaluated the quality of the included studies using the National Institute of Health (NIH) quality assessment tool for observational cohort and cross-sectional studies [31]. This tool consists of 14 questions regarding the sample size and selection and exposure and outcome assessment. Studies scoring $\geq 9$ points
were considered of good quality, 5–8 points of fair quality and 1–4 points of poor quality. Diagnostic accuracy studies were assessed using Cochrane’s QUADAS-2 Tool.

2.4. Statistical Analysis

We conducted two types of analyses: double arm meta-analysis and diagnostic test accuracy analysis. We used review manager software version 5.4 (The Cochrane Collaboration, UK) [32] and OpenMetaAnalyst software [33] for the double arm analysis. This analysis pooled the mean difference between IBM patients and healthy controls in terms of muscle echo intensity and muscle thickness. The random effects model of the DerSimonian Laird method was applied to account for heterogeneity between the studies [34]; \( p < 0.05 \) was considered statistically significant. Heterogeneity was assessed using the Chi-squared test and the \( I^2 \) statistic; \( p < 0.05 \) proved significant heterogeneity and \( I^2 > 50\% \) indicated substantial heterogeneity. Leave-one-out sensitivity analysis was carried out to examine the effect of omitting each study on the overall result. Publication bias could not be assessed using funnel plots due to the small number of included studies [29].

We used Meta-DiSc software [35] for the diagnostic test accuracy analysis to pool the sensitivity and specificity, as well as the likelihood ratios of ultrasonography, in the diagnosis of IBM compared to the other reference tests (muscle biopsy and magnetic resonance imaging (MRI)). A receiver–operator curve (ROC) was created and the area under the curve (AUC) was calculated to evaluate the performance of US as a diagnostic test for IBM.

3. Results

3.1. Study Characteristics

A total of 7 studies [25,36–41], with 108 IBM patients and 171 healthy controls, were included in the analysis. Four studies [25,36,38,41] reported the echogenicity of the flexor digitorum profundus (FDP) and gastrocnemius (GC) muscle groups with 168 participants (76 IBM patients and 92 controls). Two studies [38,39] reported the muscle thickness of the FDP with 128 participants (46 IBM patients and 82 controls). Three studies [36,37,40] were involved in diagnostic test accuracy meta-analysis. The characteristics of the included studies are presented in Table 1.

3.2. Quality Assessment

The quality of included studies was assessed using the NIH scale and QUADAS-2 tool. Using the NIH scale, one study scored 9 and was considered of high quality, while four studies were considered to be of fair quality (score 5–8) (Table 2). For diagnostic test accuracy studies, some concerns regarding the patient selection methodology were found for two studies [36,37], while no concern regarding applicability was found for other domains (Figure 2).
| Author               | Year  | Country          | Study Design | US Device Type                                                      | Evaluated Muscles (Echogenicity +/- Thickness)                                                                 | IBM             | Control             |
|---------------------|-------|------------------|--------------|---------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|------------------|---------------------|
| Albayda             | 2018  | USA              | Prospective  | GE Logiq e with 12L linearphased array transducer                   | Muscle Echointensity: FDP, flexor carpi ulnaris (FCU), GC, deltoids, biceps, rectus femoris (RF) and tibialis anterior (TA) | Yes              | 18                  | 64.8 (9.8)          | 9.9                  | 138.0 (93.2)        | 28                | 48.5 (15.5)         | 13:15               |
| Nodera              | 2015  | Japan            | Prospective  | LOGIQ™ with a fixed 11-MHz linear-array transducer                  | Muscle Echointensity: medial head of the GC, soleus, FDP and FCU                                                | No, only right side chosen in order to avoid potential selection bias of highly correlated, bilateral data from the same individual | 11               | 74.5 (6.8)          | 7.4                  | 48.0 (40.2)         | 11                | 73.5 (9.9)          | 6:5                 |
| Paramalingam        | 2021  | Australia        | Prospective  | Canon Aplio 500 machine with a 14-5 linear probe set at 14 MHz      | Muscle Echointensity and thickness: FDP, FCU, vastus lateralis (VL), TA and deltoid                                | No (on the participant’s right-hand side)                                                                       | 5                | 70.40 (6.50)        | 5.0                  | 29                  | 46.60 (16.10)       | 16:13              |
| Leeuwenberg         | 2020  | USA and Netherlands | Retrospective | GE Logiq e with 12L linearphased array transducer                   | Muscle Echointensity and thickness: FDP, medial head of the GC, RF and/or VL                                    | Yes              | 25                  | 65.1 (52-80)        | 12:13               | 116.9 (30-360)      | 25                | 65.9 (51-80)        | 9:16                |
**Table 1. Cont.**

| Author                     | Year | Country            | Study Design | US Device Type                                                                 | Evaluated Muscles (Echogenicity +/− Thickness) | Bilateral Muscle Evaluation | No. of Patients | Mean Age in Years (SD) | Sex M:F | Duration of Disease, Mean (SD) | No. of Patients | Mean Age in Years (SD) | Sex M:F |
|----------------------------|------|--------------------|--------------|--------------------------------------------------------------------------------|-----------------------------------------------|------------------------------|----------------|------------------------|---------|-----------------------------|----------------|------------------------|---------|
| Leeuwenberg (Radbourdumc)  | 2020 | USA and Netherlands| Retrospective| Esaote Mylab Twice machine (Esaote SpA, Genoa, Italy) with an LA533 linear 3-13-MHz transducer | Muscle Echointensity and thickness: FDP, medial head of the GC, RF and/or VL | Mostly but unilateral results were used as representative of both sides | 16              | 70.5 (54–84)      | 9.7     | 67.2 (12–228)            | 63              | 63.4 (50–82)         | 28:35   |
| Karvelas                   | 2019 | USA                | Prospective, blinded | GE Logiq S8 (GE Healthcare, Chicago, Illinois) with a 15-MHz linear array transducer; an Esaote MyLabGamma (Esaote S.p.A, Genoa, Italy) with an 18-MHz linear array transducer; and a Biosound MyLab25 (Esaote) with an 18-MHz linear array transducer. | FDP and FCU | Yes | 15 | 73.2 (5.78) (61–83) | 13:2   | 55.93 (13.82) (34–79) | 7:8   |
| Guimaraes                  | 2021 | Brazil             | Prospective | HD II XE (Philips Medical Systems, Nederland B.V., Best, Netherlands) US system with a 12-MHz linear array | Quadricep muscle group (RF, vastus medialis, vastus intermedius, VL), GC and FDP | Yes | 12 | 63.3 (9.1) (49–84) | 11:1   | 90.6 (35.1)            |                   |
| Noto                       | 2013 | Japan              | Prospective | GE Logiq P5 system with a 10-MHz linear-array probe (GE Healthcare Japan) | Echogenicity in FDP–FCU | no | 6 | 71.5 (68–79)         | 5:1     | 56.7 (14–120)          |                   |
Table 2. Quality assessment of the included studies according to the National Institute of Health (NIH) quality assessment tool for observational cohort and cross-sectional studies.

| Study                  | C 1 | C 2 | C 3 | C 4 | C 5 | C 6 | C 7 | C 8 | C 9 | C 10 | C 11 | C 12 | C 13 | C 14 |
|------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|
| Paramalingam 2021      | Yes | Yes | CD  | Yes | No  | NA  | Yes | NA  | Yes | Yes  | Yes  | Yes  | Yes  | Yes  |
| Leeuwenberg 2020       | Yes | Yes | Yes | Yes | No  | NA  | Yes | No  | Yes | No   | Yes  | No   | Yes  | No   |
| Albayda 2018           | Yes | No  | CD  | Yes | No  | NA  | Yes | No  | Yes | No   | Yes  | No   | Yes  | No   |
| Nodera 2015            | Yes | No  | CD  | Yes | No  | NA  | Yes | No  | Yes | No   | Yes  | Yes  | Yes  | No   |
| Noto 2013              | Yes | Yes | CD  | Yes | No  | NA  | Yes | NA  | Yes | NA   | Yes  | No   | Yes  | No   |

C—criterion; CD—cannot be determined; NA—not applicable; Criterion 1—clear statement of research question and objective; Criterion 2—clear specification and definition of the study population; Criterion 3—participation rate of 50% of eligible persons; Criterion 4—selection of subjects from the same population and time period and application of selection criteria on all subjects uniformly; Criterion 5—justification of sample size; Criterion 6—measurement of exposure before measurement of outcome; Criterion 7—sufficient timeframe to predict an association between exposure and outcome; Criterion 8—examination of different levels of the exposure in relation to the outcome; Criterion 9—clear and valid definition of exposure measures and its consistent implementation on study subjects; Criterion 10—assessment of exposure more than once over time; Criterion 11—clear and valid definition of outcome measures and its consistent implementation on study subjects; Criterion 12—blinding of outcome assessors to exposure status of participants; Criterion 13—loss to follow up being 20% or less; Criterion 14—measurement or statistical adjustment of confounding variables.

Figure 2. Final methodological quality summary.

3.3. Echogenicity of FDP

Three studies [25,36,38] assessed the FDP muscle echogenicity using ultrasound between 70 IBM (129 US measurements) cases and 92 controls (173 US measurements). Gray scale value (GSV) was graded using an arbitrary unit that ranged from 0 to 255, from black to white [36,38]. Echogenicity index mean difference was 36.55 GSV (95% CI, 28.65–44.45, \( p < 0.001 \)), thus the echogenicity index was significant and higher in IBM patients than in controls. Between-study heterogeneity was high and significant (\( I^2 = 70\%, \ p = 0.02 \)) (Figure 3).
Figure 3. Forest plot of the echogenicity of FDP in grey scale value (GSV) of the included studies.

A study by Noto et al. in 2013 (six patients), assessed muscle echo intensity using the Heckmatt rating scale (1–4) for two muscles: the FCU and FDP. The Heckmatt rating scale mean (range) was 1.3 (1-3) for the FCU and 2.7 (2-3) for the FDP. These data were not included in the GSV analysis.

3.4. Echogenicity of GC

Ultrasound analysis between 70 IBM patients (129 US measurements) and 98 controls (185 US measurements) in three studies found that the mean difference in echogenicity index was 27.90 GSV (95% CI 16.32–39.48, \( p < 0.001 \)), denoting significance and a higher value in IBM patients than in healthy controls in the GC. Significant heterogeneity was found between study data (\( I^2 = 84\% \), \( p < 0.001 \)) (Figure 4).

Figure 4. Forest plot of the echogenicity of the GC muscle in GSV in the included studies.

Nodera et al. established higher echoic signal in the GC and FDP of IBM patients [36]. Subsequently, Albayda et al. found a significant difference in echogenicity between IBM patients and the control group, with FDP being the most discerning [25]. Furthermore, Leeuwenberg et al. assessed the echogenicity and muscle thickness in a group of 41 IBM patients with variable disease duration; Radboudumc 67.2 (12–228) months and Johns Hopkins 116.9 (30–360) months, as reported per median and range [38]. It showed a drastic surge in echogenicity in IBM patients.

3.5. Muscle Thickness of FDP

Analysis of two studies between 46 IBM cases (87 US measurements) and 82 controls (135 US measurements) revealed no significant difference in muscle thickness between IBM patients and healthy controls (MD = −1.75, 95% CI −5.01–1.51, \( p = 0.29 \)). Heterogeneity was significant and high (Figure S1).

Regarding the muscle thickness of FDP, the study from John Hopkins presented a significant increase in the thickness compared to the control group. However, Radboudumc’s data included the point of no difference, denoting no statistical difference. Paramalingam et al. compared a group of five IBM patients to a group of age- and sex-matched healthy
controls. The results showed no significant association between the muscle thickness of the FDP in IBM patients and in the control [39].

Noto et al., in 2013, also assessed muscle atrophy by measuring the muscle CSA. The mean CSA (range) was 80.5 mm$^2$ (63.0–117.4) for the FDP muscle and 131.8 mm$^2$ (109.0–149.5) for the FCU muscle.

3.6. Ultrasound Diagnostic Accuracy

Three studies assessed the diagnostic performance of ultrasound, differentiating between IBM (65 US measurements) and healthy controls (41 US measurements). The pooled sensitivity and specificity of US were 0.82, 95% CI 0.75–0.88 and 0.98, 95% CI 0.89–1.00, respectively (Figures 5 and 6). Between-study heterogeneity was not significant for either measurement. The pooled positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio were 16.89, 95% CI 4.32–66.01, 0.21, 95% CI 0.13–0.33 and 73.88, 95% CI 15.58–350.48, respectively. The Cochran’s Q and I$^2$ tests showed no significant heterogeneity between studies (Figures 7–9). As for the summary ROC curve (SROC), the area under the curve (AUC) was 0.612 and the Q* index was 0.5848 (Figure S2).

![Figure 5. Sensitivity of the US in the diagnosis of IBM.](image1)

![Figure 6. Specificity of the US in the diagnosis of IBM.](image2)
3.7. Sensitivity and Subgroup Analysis

According to the leave-one-out sensitivity analysis, none of the outcomes were affected following the removal of a single study (Figures 10–12). A relatively large, but insignificant, change was found in muscle thickness difference outcome following removal of Leeuwenberg (the John Hopkins population) [38]; the overall mean difference following its removal was still insignificant between patients and controls (Figure 12). We also performed sensitivity analyses to examine the effect of the study design (retrospective or prospective) on the heterogeneity in the meta-analyses. The study by Leeuwenberg et al. [38] was found to have a high impact on heterogeneity, as heterogeneity dropped substantially after eliminating it from the analysis, especially the Leeuwenberg (Radboudumc) population in the FDP echogenicity analysis (Figure S3) or the Leeuwenberg (John Hopkins) population in muscle thickness meta-analysis (Figure S4), and there was no major impact on the pooled effect after their elimination. However, there was only a minimal change in the pooled
effect and heterogeneity in the GC echogenicity meta-analysis after leaving it out of that study [38].

Figure 10. Leave-one-out analysis of the studies commented on the echogenicity of the FDP muscle.

Figure 11. Leave-one-out analysis of the studies commented on the muscle thickness of the GC muscle.

Figure 12. Leave-one-out analysis of the studies commented on the muscle thickness of the FDP muscle.

To assess the impact of disease duration on the overall effect and heterogeneity, we conducted subgroup analysis of the FDP and GC echogenicity meta-analysis across studies by mean duration of disease, either < 70 months or ≥70 months. The result of the subgroup analysis showed that there was no statistically significant difference in overall FDP and GC echogenicity between the subgroups (p = 0.77, I² = 0%; p = 0.59, I² = 0% respectively) and the pooled effects of FDP and GC echogenicity analysis for each subgroup were consistent with the primary analysis (Figures S5–S6).
3.8. Heterogeneity

Substantial heterogeneity was observed among the studies \( (I^2 = 70\% \text{ and } I^2 = 84\%) \) with regards to the echogenicity of the FDP and GC muscle group set, respectively. Moreover, muscle thickness demonstrated higher heterogeneity \( (I^2 = 91\%) \). Regarding the diagnostic test accuracy meta-analysis, the heterogeneity in the sensitivity and the specificity analysis recorded \( X^2 = 1.7; \ p = 0.42 \) and \( X^2 = 2.42; \ p = 0.49 \), respectively. The threshold analysis for different cut-off values was made using Spearmann’s Test in an attempt to identify the reason for heterogeneity, and yielded no effect. Only three \([36,37,39]\) out of the seven studies \([25,36–41]\) evaluated the inter-rater or intra-rater reliability of the measurements. Accordingly, publication bias assessment could not be performed due to the small number of included studies.

Following the strategies described by Part Two of the Cochrane Handbook 9.5.3 for addressing the heterogeneity, we performed a meta-regression analysis for covariates that may have potentially caused heterogeneity. The meta-regression showed that the US device could be a source of heterogeneity in the meta-analysis, comparing the FDP muscle echogenicity between IBM patients and controls (\( p \)-value 0.01 and 0.03), as shown in Table S1. However, the meta-regression analyses for other analyses showed that the device was not a source of heterogeneity. Meta-regression may be not conclusive when there are few studies included in the meta-analysis.

In terms of the laterality of muscle evaluation, four studies \([25,37,38,40]\) used the average of bilateral examination. It is worth noting that Leeuwenberg (John Hopkins) \([38]\) used the average of the bilateral examination while Leeuwenberg (Radboudumc) \([38]\) mostly used the average of the bilateral examination, and in cases of unilateral examination, these results were used as representative for both sides. Five studies \([25,36,38,39,41]\) used the quantitative method for the scoring system. Three of them \([36,39,41]\) also reported using the Heckmatt rating scale (which is a semi-quantitative scale), which included the use of a modified Heckmatt rating scale by one study \([36]\), whereas, two studies \([37,40]\) used only the Heckmatt rating scale.

4. Discussion

The results of our meta-analyses demonstrated an increased echogenicity in certain muscle groups compared to controls and could therefore be used as a supportive criterion in the challenge of an early diagnosis for IBM. In detail, the echogenicity in IBM patients was increased with 36.55 GSV, 95% CI 28.65–44.45 and 27.9 GSV 95% CI, 16.32 – 39.48 for the FDP and GC muscle groups, respectively, and the test of overall effect was significant for both measurements (\( p < 0.001 \)). In contrast, results of muscle thickness in IBM patients showed no difference with \(-1.75, 95\% \text{ CI, } -5.01 \text{ to } 1.51 \). No potential outliers were identified by conducting leave-one-out sensitivity analysis. The heterogeneity testing was significant in the measures of echogenicity in the FDP and GC muscle groups, muscle thickness in the FDP and diagnostic test accuracy. Results of diagnostic test accuracy meta-analysis revealed a collective specificity of 0.98, 95% CI, 0.89–1 and sensitivity of 0.82 CI 95%, 0.75–0.88. This agreed with the data of a previous study that showed significant accuracy of US compared to MRI in detecting muscle abnormalities and fat infiltration of IBM patients \([40]\). On the basis of a literature comparison, US provided a higher sensitivity than anti-cN1A-antibody testing, a less invasive technique than electromyography (EMG) and muscle biopsy, as well as a less expensive method than MRI.

The anti-cN1A-antibody sensitivity greatly varied in several published studies; ranging from 33% to 76% \([24,42]\), while specificity was indicated from 87% to 100%. The anti-cN1A-antibody was also detected in other IIM and autoimmune diseases \([43,44]\). Needle EMG is a well-established and historically used neurophysiological test, for which a sensitivity of 89% was determined for IIM in the work of Bohan and Peter \([45]\) and abnormalities of IIM were described in more detail by \([46]\). For IBM, the data were even less frequent and showed, e.g., spontaneous activity in 62.5% \([47]\) or classical myopathic
changes [48] in EMG examinations. Recently, diagnostic accuracy has been evaluated for IIM, demonstrating a sensitivity of 85.2% in IBM [49].

Muscle biopsy is the diagnostic gold standard for identifying the typical pathological features with a limited sensitivity in early disease phases, which may necessitate multiple muscle biopsies to establish a diagnosis [50]. However, the full pathological picture (rimmed vacuoles, p62 aggregates, increased major histocompatibility complex class I expression and endomysial T cells) reached a sensitivity of 93% and specificity of 100% [50].

To increase diagnostic certainty as early as possible, a non-invasive imaging technique could be used as a supportive diagnostic measure with the potential of repeated application and without potential harm. Both MRI and US fulfil these criteria. Until now, MRI remains the most widely used technique for muscle imaging as such modality can delicately visualize the distribution of affected muscles and surrounding tissues (fascia and skin), disease activity and permanent muscle damage, such as muscle atrophy and/or replacement by fatty tissue [51]. For IBM, the sensitivity of thigh muscle MRI was evaluated in 2002 and showed 72% [52]. Whole-body MRI also contributed to specific recognition patterns in IBM [10], e.g., with a sensitivity of 80% and specificity of 100% for morphological/degenerative features of the quadriceps femoris muscle [53]. Thus, these two statistical parameters are comparable to US, whereby the MRI application has the advantage of less operator dependency but also some potential drawbacks, such as its high cost, time-consuming nature (whole-body MRI requires approximately 1 h), lack of widespread availability and exclusion of subjects with metal implants, pacemakers or claustrophobia. The more cost-effective, faster, geographically widespread (e.g., in rural areas and for patients with restricted mobility), bedside and alternative method is US, with its excellent image resolution and the ability to detect morphological changes in the muscle, such as edema, atrophy and muscle replacement by fibroadipose tissue. The visualization of these myocharacteristics by the US parameters of muscle echogenicity and muscle thickness [36] have led to this method also being used for other myopathies and as follow-up assessment for disease severity and residual muscle damage [40].

In detail, echogenicity is a hallmark feature of muscle replacement by fatty tissue and fibrosis [54] and is widely accepted as an ultrasound parameter in chronic muscle pathologies, such as myositis [55–57]. Mainly three methods are available to assess muscle echogenicity: (1) visual qualitative method to determine echogenicity in relation to other tissues; (2) semiquantitative evaluation by a scale [58]; and (3) grayscale-based quantitative measurement. Independent of the 40-year long availability, different evaluation methods for echogenicity are used in clinical trials, indicating a further need of harmonizing and combining them.

Standardization of muscle thickness is another issue to be solved, as various factors, such as body size of the patient, disease duration, sex and exact anatomical localization, might affect measurements [38,39].

Beside the patient-dependent factors leading to the heterogeneity in our meta-analysis, another major source is the dependency on the ultrasound operator primed with the anisotropic nature of the muscle, in which any trace change in the viewing angle alters the echo intensity. However, to overcome this, Scott et al. recommended taking extra care to reduce the probe tilt, especially while evaluating the echo-intensity values [57].

Other probable causes of heterogeneity include the variation in sampling, clinical characteristics of IBM patients and the differences in the control groups.

To our knowledge, this is the first systematic review and meta-analysis to convey the echogenicity and muscle thickness in IBM and to evaluate the sensitivity and specificity of ultrasonography to support the diagnosis of IBM by conducting diagnostic test accuracy meta-analysis. Although our literature searches were thorough and data extraction were cautious, limitations still exist in this review, such as the small number of published papers on IBM. However, as suggested by Callan et al., periodical evaluation of the current evidence is mandatory to increase awareness and improve research methodology [9]. Further meta-analysis is considered once additional homogenous studies become available.
In addition, upcoming studies on the diagnostic value of muscle ultrasonography should include larger samples of patients with IBM, as current evidence of sonography is not on a sufficient standardized level for the diagnosis of IBM. Nevertheless, we propose that the deployment of US, in combination with the standard clinical, histological and serological assessment, and with the long-term goal of its implementation, adds value to education and clinical practice.

Finally, the development of international standard operating procedures, e.g., as developed for measures of disease activity in myositis by iMACS [58], would equalize the imaging evaluation and provide the basis for longitudinal studies. Advances in artificial intelligence will further facilitate US as a useful diagnostic and follow-up/longitudinal technique [59], which will foster its use in clinical trials [60].

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cells11040600/s1: Figure S1—Forest plot of the muscle thickness of the FDP in mm of the included studies; Figure S2—SROC curve showing the AUC was 0.612 and the value with the highest sensitivity and specificity (Q* index) was 0.5848; Figure S3—Sensitivity analysis for FDP echogenicity meta-analysis; Figure S4—Sensitivity analysis for muscle thickness meta-analysis; Figure S5—Subgroup analysis for FDP echogenicity meta-analysis by disease duration; Figure S6—Subgroup analysis for GC echogenicity meta-analysis by disease duration; Table S1—Meta-regression for US device.

**Author Contributions:** Conceptualization, R.A.; methodology, R.A. and K.A.M.; validation, R.A., K.A.M. and A.E.; screening, A.E., Y.T.S., M.M.B., A.M.A., M.A.E., I.M., K.T.D., H.A.H. and H.M.G.; formal analysis, A.E., Y.T.S. and K.T.D.; data curation, R.A., K.A.M., A.E., Y.T.S. and K.T.D.; writing—original draft preparation, R.A., M.A.E., I.M., K.A.M., Y.T.S. and A.E.; writing—review and editing, R.A. and J.Z.; visualization, R.A. and J.Z.; supervision, R.A. and J.Z.; project administration, R.A. and K.A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge support by the Open Access Publication Funds of the Göttingen University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this manuscript are available from the corresponding databases: Medline (through PubMed) (https://pubmed.ncbi.nlm.nih.gov/ (accessed on 11 September 2021)), Scopus (http://scopus.com/home.uri (accessed on 11 September 2021)) and Web of Science (https://www.webofscience.com (accessed on 11 September 2021)).

**Acknowledgments:** We thank Borwin Bandelow for scientific advices. J.Z. is member of the European Reference Network for Rare Neuromuscular Diseases (ERN EURO-NMD).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Phillips, B.A.; Zilko, P.J.; Mastaglia, F.L. Prevalence of sporadic inclusion body myositis in Western Australia. *Muscle Nerve* 2000, 23, 970–972. [CrossRef]

2. Aoki, M.; Suzuki, N.; Kato, M.; Warita, H. Recent progress in diagnosis and pathomechanism of inclusion body myositis. *Rinsho Shinkeigaku* 2014, 54, 1115–1118. [CrossRef]

3. Paltiel, A.D.; Ingvarsson, E.; Lee, D.K.; Leff, R.L.; Nowak, R.J.; Petschke, K.D.; Richards-Shubik, S.; Zhou, A.; Shubik, M.; O’Connor, K.C. Demographic and clinical features of inclusion body myositis in North America. *Muscle Nerve* 2015, 52, 527–533. [CrossRef]

4. Benveniste, O.; Guiguet, M.; Freebody, J.; Dubourg, O.; Squier, W.; Maisinobe, T.; Stoijkovic, T.; Leite, M.I.; Allenbach, Y.; Herson, S.; et al. Long-term observational study of sporadic inclusion body myositis. *Brain* 2011, 134, 3176–3184. [CrossRef][PubMed]

5. Machado, P.; Brady, S.; Hanna, M.G. Update in inclusion body myositis. *Curr. Opin. Rheumatol.* 2013, 25, 763–771. [CrossRef]

6. Dimachkie, M.M.; Barohn, R.J. Inclusion body myositis. *Neurot. Clin.* 2014, 32, 629–646. [CrossRef]

7. Suzuki, N.; Mori-Yoshimura, M.; Yamashita, S.; Nakano, S.; Murata, K.Y.; Mori, M.; Inamori, Y.; Matsui, N.; Kimura, E.; Kusaka, H.; et al. The updated retrospective questionnaire study of sporadic inclusion body myositis in Japan. *Orphanet J. Rare Dis.* 2019, 14, 155. [CrossRef]
35. Zamora, J.; Abraira, V.; Muriel, A.; Khan, K.; Coomarasamy, A. Meta-DiSc: A software for meta-analysis of test accuracy data. **BMC Med. Res. Methodol.** 2006, 6, 31. [CrossRef]

36. Nodera, H.; Takamatsu, N.; Matsu, N.; Mori, A.; Terasawa, Y.; Shimatani, Y.; Osaki, Y.; Maruyama, K.; Iuzmi, Y.; Kaji, R. Intramuscular dissociation of echogenicity in the triceps surae characterizes sporadic inclusion body myositis. **Eur. J. Neurol.** 2016, 23, 588–596. [CrossRef] [PubMed]

37. Karvelas, K.R.; Xiao, T.; Langefeld, C.D.; Walker, F.O.; Pathak, S.; Caress, J.B.; Baute, V.; Cartwright, M.S. Assessing the accuracy of neuromuscular ultrasonography for inclusion body myositis. **Muscle Nerve** 2019, 59, 478–481. [CrossRef]

38. Leeuwenberg, K.E.; van Allen, N.; Christopher-Stine, L.; Paik, J.J.; Tiniakou, E.; Mecoli, C.; Doorduin, J.; Saris, C.G.J.; Albayda, J. Ultrasound can differentiate inclusion body myositis from disease mimics. **Muscle Nerve** 2020, 61, 783–788. [CrossRef] [PubMed]

39. Paramalingam, S.; Needham, M.; Raymond, W.; Mastaglia, F.; Lightowler, D.; Morin, N.; Counsel, P.; Keen, H.I. Muscle shear wave elastography, conventional B mode and power doppler ultrasonography in healthy adults and patients with autoimmune inflammatory myopathies: A pilot cross-sectional study. **BMC Musculoskeletal Disord.** 2021, 22, 537. [CrossRef]

40. Guimaraes, J.B.; Cavalcante, W.C.P.; Cruz, I.A.N.; Nico, M.A.; Filho, A.G.O.; da Silva, A.M.S.; Zanoteli, E. Musculoskeletal Ultrasound in Inclusion Body Myositis: A Comparative Study with Magnetic Resonance Imaging. **Ultrasound Med. Biol.** 2021, 47, 2186–2192. [CrossRef] [PubMed]

41. Noto, Y.; Shiga, K.; Tsuji, Y.; Kondo, M.; Tokuda, T.; Mizuno, T.; Nakagawa, M. Contrasting echogenicity in flexor digitorum profundus-flexor carpi ulnaris: A diagnostic ultrasound pattern in sporadic inclusion body myositis. **Muscle Nerve** 2014, 49, 745–748. [CrossRef]

42. Herbert, M.K.; Pruip, G.J. Novel serology testing for sporadic inclusion body myositis: Disease-specificity and diagnostic utility. **Curr. Opin. Rheumatol.** 2015, 27, 595–600. [CrossRef]

43. Arbor, P.-A.; Herberth, R.; Schütz, O.; Kollack-Walker, K.; Alonso-Ojeda, S.; Godeau, P.; et al. Magnetic resonance imaging pattern recognition in idiopathic inflammatory myopathies. **J. Neurol. Neurosurg Psychiatry** 2015, 86, 1152–1157. [CrossRef] [PubMed]

44. Lloyd, T.E.; Christopher-Stine, L.; Pinal-Fernandez, I.; Verbeek, M.; Cooper, R.G.; Roberts, S.A. Magnetic Resonance Imaging in Sporadic Inclusion Body Myositis: A Pilot Study. **Brain Pathol.** 2016, 26, 777–782. [CrossRef] [PubMed]

45. Bohan, A.; Peter, J.B.; Bowman, R.L.; Pearson, C.M. Computer-assisted analysis of 153 patients with polymyositis and dermatomyositis. **Medicine** 1977, 56, 255–286. [CrossRef]

46. Blijham, P.J.; ter Laak, H.J.; Scheltens, P.M.M.; Rietveld, A.; Choma, M.; Lamb, J.A.; Cooper, R.G.; Roberts, M.; Badirising, U.A.; et al. Disease specificity of autoantibodies to cytosolic 5′-nucleotidase 1A in sporadic inclusion body myositis versus known autoimmune diseases. **Am. J. Med. 2016, 140, 2186–2192. [CrossRef] [PubMed]

47. Herbert, M.K.; Stammen-Vogelzangs, J.; Verbeek, M.M.; Rietveld, A.; Choma, M.; Lamb, J.A.; Cooper, R.G.; Roberts, M.; Badirising, U.A.; et al. Disease specificity of autoantibodies to cytosolic 5′-nucleotidase 1A in sporadic inclusion body myositis versus known autoimmune diseases. **Am. J. Med. 2016, 140, 2186–2192. [CrossRef] [PubMed]

48. Barkhaus, P.E.; Periuket, M.I.; Nandedkar, S.D. Quantitative electrophysiologic studies in sporadic inclusion body myositis. **Muscle Nerve** 1999, 22, 480–487. [CrossRef] [PubMed]

49. Bolko, L.; Leonard Louis, S.; Salmon, J.; Benveniste, O.; Allenbach, Y.; Maisonobe, T. Diagnostic Accuracy of Electromyogram for Myositis; Abstract. **Arthritis Rheumatol.** 2021, 73 (Suppl. S10), 0714. Available online: https://acrabstracts.org/abstract/diagnostic-accuracy-of-electromyogram-for-myositis (accessed on 20 December 2021).

50. 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. [CrossRef]

51. Malnar, T.; Bachasson, D.; Mercy, G.; Sarkis, E.; Anquetil, C.; Benveniste, O.; Allenbach, Y. MRI and muscle imaging for idiopathic inflammatory myopathies. **Brain Pathol.** 2021, 31, e12954. [CrossRef]

52. Dion, E.; Cherin, P.; Payan, C.; Fournet, I.C.; Papo, T.; Maisonobe, T.; Auberton, E.; Chosidow, O.; Godeau, P.; Piette, J.C.; et al. Magnetic resonance imaging criteria for distinguishing between inclusion body myositis and polymyositis. **J. Rheumatol. 2002, 29, 1897–1906. [CrossRef]

53. Tasc, G.; Monforde, M.; De Fino, C.; Kley, R.A.; Ricci, E.; Mirabella, M. Magnetic resonance imaging pattern recognition in sporadic inclusion-body myositis. **Muscle Nerve** 2015, 52, 956–962. [CrossRef]

54. Heckmatt, J.Z.; Dubowitz, V.; Leeman, S. Detection of pathological change in dystrophic muscle with B-scan ultrasound imaging. **Lancet** 1980, 1, 1389–1390. [CrossRef]

55. Pillen, S. Skeletal muscle ultrasound. **Eur. J. Transl. Myol.** 2010, 20, 145–156. [CrossRef]

56. Heckmatt, J.Z.; Leeman, S.; Dubowitz, V. Ultrasound imaging in the diagnosis of muscle disease. **J. Pediatr.** 1982, 101, 656–660. [CrossRef]

57. Dankel, S.J.; Abe, T.; Bell, Z.W.; Jessee, M.B.; Buckner, S.L.; Mattocks, K.T.; Mouser, J.G.; Loenneke, J.P. The Impact of Ultrasound Probe Tilt on Muscle Thickness and Echo-Intensity: A Cross-Sectional Study. **J. Clin. Densitom.** 2020, 23, 630–638. [CrossRef]

58. Rider, L.G.; Aggarwal, R.; Machado, P.M.; Hogrel, J.Y.; Reed, A.M.; Christopher-Stine, L.; Ruperto, N. Update on outcome assessment in myositis. **Nat. Rev. Rheumatol.** 2018, 14, 303–318. [CrossRef]
59. Regensburger, A.P.; Knieling, F.; Feldkamp, A.; Rascher, W.; Diesch, K.; Woelfle, J.; Prokosch, H.U.; Jüngert, J. Time Tracking of Standard Ultrasound Examinations in Pediatric Hospitals and Pediatric Medical Practices—A Multicenter Study by the Pediatric Section of the German Society of Ultrasound in Medicine (DEGUM). *Ultraschall Med.* 2021, 42, 379–387. [CrossRef]

60. Benveniste, O.; Hogrel, J.Y.; Belin, L.; Anoussamy, M.; Bachassin, D.; Rigolet, A.; Laforet, P.; Dzangué-Tchoupou, G.; Salem, J.-E.; Nguyen, L.S.; et al. Sirolimus for treatment of patients with inclusion body myositis: A randomised, double-blind, placebo-controlled, proof-of-concept, phase 2b trial. *Lancet Rheumatol.* 2021, 3, 40–48. [CrossRef]