Title:

Routine musculoskeletal ultrasound findings impact diagnostic decisions maximally in autoantibody-seronegative early arthritis patients

Authors:

Iqbal K, MBBS[1,2], Lendrem DW, PhD[1], Hargreaves B, BSc[1,2], Isaacs JD, BSc, PhD, MBBS[1,2], Thompson B, MBBS, PhD[2], Pratt AG, MBChB, BSc, PhD[1,2]

Affiliations:

1. Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK.
2. Musculoskeletal Services Directorate, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.

Correspondence to:

Dr. Arthur G Pratt,

Institute of Cellular Medicine (Musculoskeletal Research Group), Newcastle

University, Newcastle upon Tyne, UK. NE2 4HH

arthur.pratt@ncl.ac.uk

Key words:

Musculoskeletal ultrasound; early arthritis; diagnosis.

Word count:

1,995

Abstract.

Background: The diagnostic value added by musculoskeletal ultrasound (MSUS) over standard clinical and laboratory parameters has proved difficult to quantify. The additive contribution to diagnostic classification of a pragmatic, 15 minute MSUS protocol was appraised in a large, unselected cohort of early arthritis clinic attendees.

Methods: Detailed baseline characteristics were recorded. Semi-quantitative MSUS scoring of the most symptomatic wrist, $2^{nd}/3^{rd}$ MCPs and PIPs and $2^{nd}/5^{th}$ MTPs was recorded, along with the sonographer's scan impression ("definitely inflammatory," "possibly inflammatory" or "non-inflammatory"). MSUS findings were available to rheumatologist diagnosticians during subsequent consultations. Persistent inflammatory arthritis (PIA) was classified only where patients were started on $\geq$1 disease modifying anti-rheumatic drug (DMARD). Multivariate and ROC curve analyses were used to identify independent discriminators of PIA, and the added value of MSUS parameters.

Results: 831 patients were enrolled, of whom 31.3% acquired a PIA diagnosis. SJC, CRP, age and ACPA status were non-redundant clinical/laboratory predictors of a PIA diagnosis by consulting rheumatologists, with good discriminatory utility (AU ROC 0.88). While the additive contribution of summed parameters from the 7-joint MSUS protocol to this model was statistically significant ($p=0.004$) it was numerically small (ΔAU ROC 0.02). However, the additive contribution to diagnostic outcome of "sonographer's scan impression" over clinical parameters alone became substantial in the sub-cohort of ACPA-negative patients, increasing the AU ROC by 9% from 0.81 to 0.90 ($p<.0001$).

Conclusions: The clinical utility of a 15-minute MSUS screen for diagnosing PIA requiring DMARDs is most evident amongst ACPA negative patients attending an EA clinic.

The importance of prompt therapeutic intervention to improve outcomes for immune-mediated inflammatory arthritis [1] explains the recent growth in early arthritis clinics. Ever-earlier evaluation of patients with suspected inflammatory arthritis (SIA) poses new challenges to rheumatologists, not least when assessing synovitis which may remain subtle or sub-clinical in early disease. Musculoskeletal ultrasound (MSUS) is a potentially valuable tool in this context, having greater sensitivity for detecting abnormalities than clinical examination alone[2]. Expert consensus supports its use, but there is a paucity of real-world data on its value in busy clinical settings[3, 4].

Several outstanding issues warrant consideration. First, defining a sub-group of SIA patients for whom prioritised MSUS access maximally impacts clinical decision-making could significantly streamline service delivery. Here, the presence of MSUS abnormalities in the peripheral joints of autoantibody seropositive arthralgia patients who lack clinical synovitis has been shown to predict arthritis persistence[5]; limited data also indicate that power Doppler signal may increase the likelihood of persistent joint inflammation in an autoantibody *seronegative* group of patients with hand arthralgia and early morning stiffness[6]. Second, defining a "minimum MSUS dataset" that can rapidly be acquired for such patients is needed. Considering this in patients with early, clinically-detectable synovitis, Filer *et al* showed that summed power Doppler scores from 10 out of a total of 38 peripheral joints improved predictive value for the development of rheumatoid arthritis over clinical variables alone[7]. Whether such findings may be extrapolated for predicting persistent arthritis development in unselected SIA patients is unknown, although algorithms ignoring the wrists are unlikely to be informative[8]. Finally, the healthcare professional(s) best-placed to deliver MSUS assessments, and the format via which their findings are best reported to non-scanning diagnosticians, are often overlooked.

An evidence-based strategy for applying MSUS in the routine assessment of newly-referred SIA patients remains a pressing unmet need. In addition to studies of the kind referred to above (reviewed in Reference [3]), real-world data from large, unselected early SIA populations are likely to

inform their development. A published 7-joint ultrasound protocol[9] has been adapted for pragmatic application during 15-minute screening appointments that form part of all initial patient assessments in the Newcastle Early Arthritis Clinic (NEAC) since 2012. We determined that quantifying the additive contribution of MSUS to clinicians' diagnostic behaviour over and above clinical and laboratory parameters alone might (i) appraise the utility of our clinical strategy, and (ii) point towards a definable subgroup of SIA patients attending the clinic for whom such a strategy might be prioritised.

Methods.

For a full description of methods see on-line *Supplementary Information*.

Patients and clinical procedures. Consecutive treatment-naïve adults referred from primary care with SIA to the NEAC between January 2015 and August 2017 were enrolled. Such patients receive two initial assessment appointments one week apart. At the first, detailed baseline demographic and clinical parameters are undertaken, along with MSUS assessment. At the subsequent visit, patients are reviewed by a consultant rheumatologist with access to all results, and assigned an initial clinical diagnosis. We considered a diagnosis of 'persistent inflammatory arthritis' (PIA), warranting commencement of ≥1 disease modifying anti-rheumatic drug(s), as the most clinically relevant outcome. *Supplementary Figure S1A* summarises the clinical pathway within which data were collected.

MSUS Assessment. Images recording longitudinal dorsal views of the wrist, 2nd and 3rd MCPs, 2nd and 3rd PIPs and 2nd and 5th MTPs of the most symptomatic hand/foot were mandated in all patients, based on the Backhaus protocol[9]. Imaging of additional joint areas was permitted at the sonographer's discretion according to individual patient symptoms, but with maximum scanning/recoding time of 15 minutes per patient. MSUS findings were recorded on a semi-quantitative (0-3) scale for greyscale (GS) and power Doppler (PD) domains according to consensus definitions[10, 11]. 7-joint GS and PD "synovitis load" was calculated for each patient as described in *Supplementary Information*. Relevant findings at non-mandated joint areas were noted as free-text. Finally, sonographers recorded their "scan impression", a global interpretation of mandated and non-mandated scan findings in relation to the presence of inflammatory arthritis (definite, possible, absent; scored 0, 1, 2, respectively). All findings were documented (*Supplementary Figure S1B*).

Data management and analysis. All analyses were carried out using JMP Pro 13.0 (SAS Institute Inc) using methods previously outlined[8]. Multivariate and ROC curve analyses were used to identify independent discriminators of PIA, and the added value of MSUS parameters.

Results.

Patients. 831 consenting patients, enrolled between January 2015 and August 2017, were followed up for a minimum of 6 months (median 15 months) during the study. The evolution of clinical diagnoses as assigned by consulting rheumatologists is summarised in *Supplementary Table S1* and *Supplementary Figure S1C*. Amongst PIA patients, DMARDs were commenced at the first consultant visit in the majority (75%) of cases, the remainder being commenced during follow-up. Baseline clinical, serological and MSUS characteristics of all patients are compared in *Table 1* according to whether they were or were not assigned a diagnosis of PIA.

Clinical and laboratory predictors of PIA. To identify clinical and laboratory parameters that had independent predictive value with respect to diagnostic outcome in the complete cohort, all 13 of the non-MSUS variables listed in *Table 1* were entered into a backward stepwise logistic regression analysis. Four non-redundant variables were independently associated with PIA: ACPA status, SJC, CRP and age. The final model containing these predictors was significantly associated with outcome (χ^2 [4 degrees of freedom] = 372.01; *P*<.0001). The strong discriminatory utility of ACPA autoantibody status for PIA was expected given its established positive predictive value for RA in early arthritis clinics[12, 13]. Indeed, we and others have highlighted the peculiar challenge of diagnosing seronegative SIA in this setting[14-16]. Mindful of the potential additive value of MSUS for this specific purpose [6], we stratified our cohort accordingly, repeating regression analysis in the ACPA-negative sub-cohort (n=703) amongst whom a significantly lower proportion of patients developed PIA during the follow-up period (20.8%, *versus* 89.1%; p=<.0001). Here, only three non-redundant variables (SJC, CRP and age) emerged as independent predictors of PIA, the resultant model also being associated with outcome (χ^2 [3 degrees of freedom] = 168.0; *P*<.0001). Results of these multivariable analyses are presented in *Supplementary Table S2*.

Additive contribution of MSUS parameters to diagnostic outcome. “Synovitis load” and “scan impression” (see *Methods*) were considered separately as alternate means by which MSUS

assessment could add independent diagnostic information. In the overall cohort and ACPA-seronegative sub-cohort, both synovitis load and scan impression demonstrated an additive predictive value for PIA outcome independently of clinical parameters (*Supplementary Tables S3* and *S4*). A range of cut-offs for a "risk metric" generated from each predictive model could then be used to plot paired ROC curves depicting the discriminatory utility of clinical parameters alone versus clinical + MSUS parameters with respect to the PIA outcome of interest[17]. Using this approach, clinical/laboratory parameters alone exhibited a good discriminatory ability for a diagnostic outcome of PIA in the complete cohort, with an area under (AU) ROC of 0.88 (95% CI =0.85-0.91). By incorporating "synovitis load" into the model there was a significant increase in AUROC, but the difference was nonetheless numerically small, with a difference between (delta, Δ) AUROC=0.02; p=004 (*Figure 1A*), and this translates to the need to scan 50 individuals before the information conferred in addition to routinely-obtained clinical parameters influences a single diagnostic decision (number needed to influence, NNI=50; see *Data Management and Analysis* section in *Supplementary Information*) . As anticipated, the discriminatory ability of clinical/laboratory parameters alone with respect to diagnostic outcome was diminished in the ACPA-seronegative sub-cohort (AUROC 0.81; 95% CI =0.77-0.85); the additive contribution of MSUS "synovitis load" was greater than that seen in the overall cohort, but remained numerically quite small (ΔAUROC = 0.04; p0.0138; *Figure 2B*), equating to a NNI of 25. Further ROC curve analysis (*Figures 2C-D*) revealed that the additive contribution of the alternative MSUS parameter "scan impression" was greater than that of "synovitis load" with respect to diagnostic outcome; this was most evident in the ACPA-seronegative sub-cohort, where the AUROC increased by 9% (Δ AUROC 0.07; p<.0001; *Figure 2D*), and the NNI reduced to 14.

To examine the possibility that the additive discriminatory value of "scan impression" might be susceptible to the experience of the US practitioner, a sensitivity analysis was undertaken in which the overall cohort was stratified according to whether MSUS assessment was undertaken by a trainer (BT, AGP; n=424) or a trainee (see *Supplementary Information*; n=407). No significant

difference was seen in ΔAUROC values derived when the two practitioner groups were compared. Following further analyses not presented here, neither were results impacted when considering semi-quantitative MSUS parameters as dichotomous (rather than continuous) variables – e.g. exploring the predictive utility of the number of joints (out of 7) that exhibited GS and/or PD synovitis above pre-defined thresholds, or EMS above and below 60 minutes.

Discussion.

To our knowledge this is the largest investigation to date of the added utility of MSUS over clinical and laboratory parameters in diagnosing SIA in an early arthritis setting. At one level, our "real-world" findings support previous work suggesting MSUS may be maximally informative for diagnostic purposes amongst ACPA-negative individuals[6], confirming this in an unselected SIA population. At another our findings inform debate, not only about the choice of scanning algorithm to employ when screening SIA patients in time-pressed early arthritis clinics, but also the means by which MSUS findings might best be communicated to diagnosticians who may not themselves be experts in the technology.

Several factors underpinned our decision to mandate, as a minimum, assessment of the Backhaus-7 joint areas[9] in our early arthritis clinic. Paramount amongst these were its coverage (albeit not entirely comprehensive) of areas shown to be maximally informative when predicting the persistence of very early arthritis[7]; but we felt it was equally important to *limit* the number of *mandated* joints to permit flexibility within the available time for discretionary assessment of additional symptomatic joints. This approach – balancing a defined consistency between *all* assessments with the ability to accommodate case-by-case clinical priorities for reporting purposes, is necessarily pragmatic and presents analytical caveats; we nonetheless suggest that it is vindicated by our findings. In particular, we observed that the apparent impact of a sonographer's "scan impression" – taking account of *all* findings (including non-mandated areas) – had an apparently more profound impact on diagnosticians' behaviour when forming an impression of PIA than did an arguably less subjective measure of "synovitis load" limited to pre-specified areas. Conceivably, scope within *any* MSUS algorithm, for *any* purpose, to permit discretionary flexibility from skilled sonographers when reporting findings to clinical decision-makers may be of pivotal importance: it may indeed be an ingredient whose absence underlies hitherto disappointing data to support MSUS use in routine clinical practice[18, 19] despite broad expert consensus of its merits[4, 20].

Caution must be exercised when interpreting our current analysis. Our data are extracted from routine records with no blinding procedures, and MSUS reports contributed to the diagnostic assessment. Our interpretation is therefore that the additive contribution of MSUS parameters quantified in our study relates to rheumatologists' diagnostic behaviour – and *not* to diagnosis *per se*. Hence, our data cannot exclude that MSUS findings – and in particular "scan impression" reports – might "mislead" diagnosticians into classifying early arthritis patients incorrectly with respect to a PIA outcome, engendering ascertainment bias. Although unlikely in our view, future blinded studies should address such concerns, whilst formally validating the reliability of "scan impression". Finally, the arguably modest absolute additive contribution of MSUS to diagnostic decision-making in our study may in part reflect the relatively low prior probability of PIA outcomes in population(s) studied), reinforcing the need to further refine the target population in which to apply this imaging modality for diagnostic purposes.

In conclusion, our data confirm that MSUS findings from a pragmatic 15-minute screen are influential upon clinical decision-making in a routine early arthritis setting, particularly for patients presenting with ACPA-seronegative SIA. Blinded studies and cost-benefit data amongst this subgroup are needed to further define the value of such an approach for broader implementation.

Key messages

- MSUS may be maximally informative for diagnosing inflammatory arthritis requiring DMARDs in ACPA-negative individuals.
- Sonographers' global "scan impression" has a greater impact on diagnostic decision-making than semi-quantitative scores alone.

Conflicts of Interest,

The authors declare no conflicts of interest.

Acknowledgments.

The authors wish to thank all patients who participated in the study, and the many clinical and administrative colleagues who made it possible.

Funding.

This work was supported by the National Institute of Health Research (Newcastle Biomedical Research Centre at Newcastle Hospitals NHS Foundation Trust and Newcastle University).

References.

1 van der Linden MP, le Cessie S, Raza K, et al. Long-term impact of delay in assessment of patients with early arthritis. Arthritis & Rheumatism 2010;62(12):3537-46.
2 Colebatch AN, Edwards CJ, Ostergaard M, et al. EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. Annals of the Rheumatic Diseases;72(6):804-14.
3 Lage-Hansen PR, Lindegaard H, Chrysidis S, Terslev L. The role of ultrasound in diagnosing rheumatoid arthritis, what do we know? An updated review. Rheumatology International 2017;37(2):179-87.
4 D'Agostino MA, Terslev L, Wakefield R, et al. Novel algorithms for the pragmatic use of ultrasound in the management of patients with rheumatoid arthritis: from diagnosis to remission. Annals of the Rheumatic Diseases 2016;75(11):1902-8.
5 van de Stadt LA, Bos WH, Meursinge Reynders M, et al. The value of ultrasonography in predicting arthritis in auto-antibody positive arthralgia patients: a prospective cohort study. Arthritis Research & Therapy 2010;12(3):R98.
6 Freeston JE, Wakefield RJ, Conaghan PG, Hensor EM, Stewart SP, Emery P. A diagnostic algorithm for persistence of very early inflammatory arthritis: the utility of power Doppler ultrasound when added to conventional assessment tools. Annals of the Rheumatic Diseases 2010;69(2):417-9.
7 Filer A, de Pablo P, Allen G, et al. Utility of ultrasound joint counts in the prediction of rheumatoid arthritis in patients with very early synovitis. Annals of the Rheumatic Diseases 2011;70(3):500-7.
8 Pratt AG, Lorenzi AR, Wilson G, Platt PN, Isaacs JD. Predicting persistent inflammatory arthritis amongst early arthritis clinic patients in the UK: is musculoskeletal ultrasound required? Arthritis Research & Therapy 2013;15(5):R118.
9 Backhaus M, Ohrndorf S, Kellner H, et al. Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: a pilot project. Arthritis & Rheumatism 2009;61(9):1194-201.
10 Szkudlarek M, Narvestad E, Klarlund M, Court-Payen M, Thomsen HS, Ostergaard M. Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance imaging, conventional radiography, and clinical examination. Arthritis & Rheumatism 2004;50(7):2103-12.
11 Wakefield RJ, Balint PV, Szkudlarek M, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology.[erratum appears in J Rheumatol. 2006 Feb;33(2):440 Note: Bruyn, George [corrected to Bruyn, George AW]]. Journal of Rheumatology 2005;32(12):2485-7.
12 van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, et al. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. Arthritis & Rheumatism 2004;50(3):709-15.
13 van Der Helm-van Mil AHM, Detert J, Cessie SL, et al. Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: Moving toward individualized treatment decision-making. Arthritis & Rheumatism 2008;58(8):2241-7.
14 Anderson AE, Pratt AG, Sedhom MA, et al. IL-6-driven STAT signalling in circulating CD4+ lymphocytes is a marker for early anticitrullinated peptide antibody-negative rheumatoid arthritis. Annals of the Rheumatic Diseases 2016;75(2):466-73.
15 Pratt AG, Lendrem D, Hargreaves B, Aslam O, Galloway JB, Isaacs JD. Components of treatment delay in rheumatoid arthritis differ according to autoantibody status: validation of a single-centre observation using national audit data. Rheumatology 2016;55(10):1843-8.
16 Boeters DM, Gaujoux-Viala C, Constantin A, Van der Helm-van Mil AH. The 2010 ACR/EULAR criteria are not sufficiently accurate in the early identification of autoantibody-negative rheumatoid

arthritis: Results from the Leiden-EAC and ESPOIR cohorts. Seminars in Arthritis and Rheumatism 2017;47:170-4.
17 Agresti A. *Categorical Data Anlaysis*. 3 ed: Wiley; 2012.
18 Haavardsholm EA, Aga AB, Olsen IC, et al. Ultrasound in management of rheumatoid arthritis: ARCTIC randomised controlled strategy trial. British Bedical Journal 2016;354.
19 Dale J, Stirling A, Zhang R, et al. Targeting ultrasound remission in early rheumatoid arthritis: the results of the TaSER study, a randomised clinical trial. Annals of the Rheumatic Diseases 2016;75(6):1043-50.
20 Wakefield RJ, D'Agostino MA, Naredo E, et al. After treat-to-target: can a targeted ultrasound initiative improve RA outcomes? Postgraduate Medical Journal 2012;88(1042):482-6.

	Overall cohort (n=831)	PIA (n=260)	Non-PIA (n=571)	P value*
Clinical / laboratory parameters				
Age	53 (17-95)	58 (20-95)	50 (17-93)	<.0001
Sex (female)	586 (70.5)	162 (19.5)	424 (51.0)	0.0006
Symptom duration	20 (1-52)	16 (2-52)	24 (1-52)	<.0001
EMS (mins)	30 (0-480)	60 (0-480)	30 (0-120)	<.0001
TJC	5 (0-76)	7 (0-63)	4 (0-76)	0.0642
SJC	0 (0-33)	2 (0-33)	0 (0-33)	<.0001
GH VAS	50 (0-100)	50 (0-100)	50 (0-100)	0.1742
CRP	4 (4-278)	8 (4-278)	4 (4-179)	<.0001
ESR	10 (1-132)	21 (2-132)	8 (1-104)	<.0001
RF	171 (20.5)	112 (43.0)	59 (10.3)	<.0001
ACPA	128 (15.4)	114 (43.8)	14 (2.5)	<.0001
Ever smoked	209 (25.2)	80 (30.7)	129 (22.6)	0.0126
HAQ	0.88 (0-20)	1.13 (0-3)	0.75 (0-20)	0.0083
MSUS parameters				
GS synovitis load	2 (0-23)	5 (0-23)	1 (0-23)	<.0001
PD synovitis load	0 (0-18)	3 (0-18)	0 (0-15)	<.0001
Scan Impression:				
0	384 (47.0)	32 (12.5)	352 (62.8)	<.0001
1	134 (16.4)	22 (8.6)	112 (19.9)	<.0001
2	299 (36.6)	202 (78.9)	97 (17.3)	<.0001

Table 1*. Baseline clinical, laboratory and musculoskeletal ultrasound (MSUS) parameters of overall cohort, including univariate analysis amongst patients with persistent inflammatory arthritis (PIA)* versus *non-PIA outcomes. Data presented as median (range) and number (%) for continuous and categorical data, respectively. ACPA, anti-citrullinated peptide antibody (CCP2 assay); EMS, early morning stiffness; IQR, inter-quartile range; RF, rheumatoid factor; SD, standard deviation, SxDur, symptom duration; T/SJC (max 78), tender/swollen joint count; GH VAS, global health visual analogue scale. Baseline MSUS data and corresponding univariate analysis is also presented (see text including supplementary methods for definitions); note that despite being 100% predictive for a PIA outcome, erosions were scarce and afforded minimal discriminatory value overall, so were excluded from further analysis. *Mann-Whitney U/Student's t tests for skewed/normally-distributed continuous data respectively; Pearson's χ^2 with Fisher's exact test for dichotomous data. [1]Symptom duration was censored at 52 weeks if described as longer than this at presentation.*

Figure Legend

Figure 1. Comparative receiver operator characteristic (ROC) analyses depicting additive value of MSUS parameters (blue lines) over clinical parameters alone (red lines) for discriminating PIA *versus* non-PIA diagnoses. The independent additive value of total GS+PD synovitis load (**A**&**B**) or scan impression (**C**&**D**) is shown amongst the entire cohort (n=831; A & C) and the sub-cohort of ACPA-seronegative individuals (n=703; **B**&**D**). Area under (AU) ROC curves (95% confidence intervals) for corresponding curves are colour-coded in figures, and p-values denote significant differences between AUCs (see methods). See text for further details.

Figure 1.

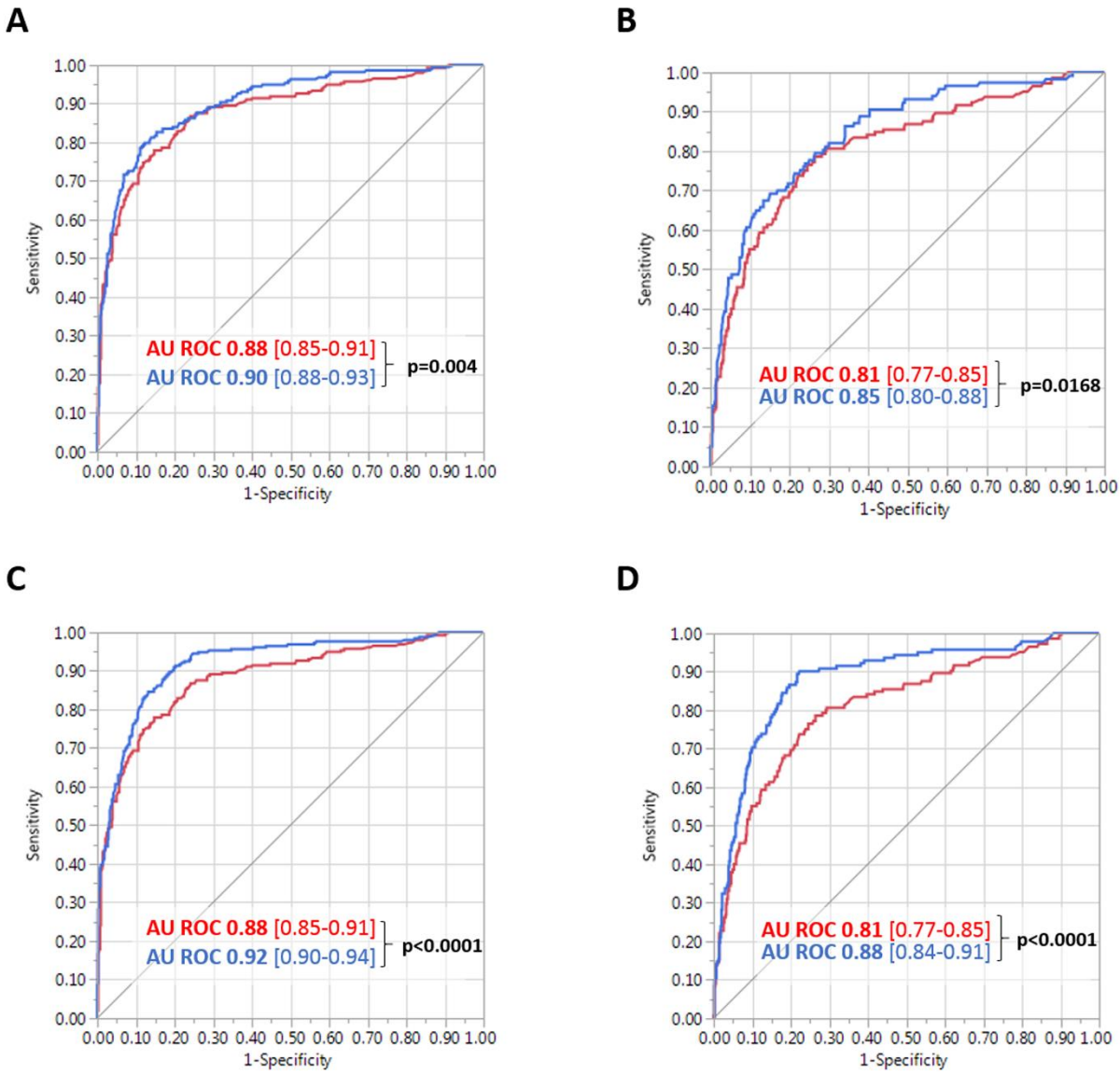

***Figure 1**. Comparative receiver operator characteristic (ROC) analyses depicting additive value of MSUS parameters (blue lines) over clinical parameters alone (red lines) for discriminating PIA versus non-PIA diagnoses. The independent additive value of total GS+PD synovitis load (**A**&**B**) or scan impression (**C**&**D**) is shown amongst the entire cohort (n=831; A & C) and the sub-cohort of ACPA-seronegative individuals (n=703; **B**&**D**). Area under (AU) ROC curves (95% confidence intervals) for corresponding curves are colour-coded in figures, and p-values denote significant differences between AUCs (see methods). See text for further details.*

Routine musculoskeletal ultrasound findings impact diagnostic decisions maximally in autoantibody-seronegative early arthritis patients.

Iqbal *et al*.

Supplementary Information.

Methods Detail.

Patients and clinical procedures. Consecutive patients ≥16 years of age referred with SIA to the NEAC between January 2015 and August 2017 who were naïve to immunomodulatory treatments were enrolled. Local guidelines for referral into the clinic include a symptom duration of <12 months and at least one of: morning stiffness >30 minutes; symmetrical distribution of arthralgia; positive MCP squeeze test[21]. All patients are offered two initial assessment appointments one week apart. At the first of these, detailed baseline demographic and clinical parameters are recorded by a nurse, MSUS assessment is undertaken and routine blood tests include acute phase markers and autoantibodies (rheumatoid factor, RF by nephelemetry and anti-citrullinated peptide antibodies by anti-cyclic citrullinated peptide, CCP2 test, Axis Shield). At the subsequent one-week visit patients are reviewed by a consultant rheumatologist with access to all of the data and investigation results, including up-to-date hand/foot radiographs; this results in an initial clinical diagnosis being assigned to each patient as previously described[22], though with reference to current disease classification criteria[23]. For practical purposes we considered a diagnosis of inflammatory arthritis which, in the opinion of the consulting rheumatologist, warranted commencement of one or more disease modifying anti-rheumatic drugs (DMARDs) during the study period, to be the most clinically relevant outcome category when considering the additive contribution of MSUS in clinical decision-making. *Figure 1A* summarises the clinical pathway within which data were collected in the study. All enrolled patients consented to participate in the study, which received a favorable review by the Newcastle and North Tyneside Local Research Ethics Committee (Reference 12/NE/0251).

MSUS Assessment. All patients underwent a MSUS assessment during their first NEAC visit. Our pragmatic scanning algorithm took account of local experience and available published reports at the study's inception[3]. We first defined a "minimum MSUS dataset" after the protocol described by Backhaus *et al*[9], mandating that longitudinal dorsal images of the wrist (midline and ulnar-carpal views), 2nd and 3rd MCPs, 2nd and 3rd PIPs and 2nd and 5th MTPs of the most symptomatic hand/foot were recorded for all patients (neutral position; dominant side if equally symptomatic). Imaging of additional joint areas was permitted at the discretion of the sonographer and as directed by individual patients according to symptoms, but with the total scanning/recoding time limited to 15 minutes per patient. MSUS findings were recorded on a paper proforma (*Figure 1B*), capturing semi-quantitative scores for greyscale (GS), power Doppler (PD) and erosion domains according to consensus definitions[10, 11] on a 0 to 3 scale from the system first suggested by Szudlarek *et al* [24], but adopting the modification of Scheel *et al* in respect of grey-scale synovitis[25]. GS and PD "synovitis load" was calculated for each patient, as the sum of semi-quantitative scores across 7 joints areas for each domain, yielding summary scores that incorporated only the highest-scoring of two probe positions (ulnar and midline) at the wrist to avoid "double-counting". MSUS parameters were also considered as dichotomous variables based on their presence or absence ($\geq$2 for GS and $\geq$1 for PD) at each joint; for simplicity these data are not shown since handling them in this way did not impact the outcome of any subsequent analyses. Relevant findings at non-mandated joint areas (including tenosynovitis and osteophytes, all recorded as free-text) were also noted. Finally, sonographers completed a "scan impression" section, recording a global interpretation of mandated and non-mandated scan findings in relation to the presence of inflammatory arthritis (definite, possible, absent). Completed proformas were available to consulting rheumatologists during subsequent one-week visits.

Two MSUS systems were in use in the NEAC during the study period; namely the *Aplio*™ Diagnostic Ultrasound System (Toshiba Medical Systems Corporation, Tochigi-Ken, Japan) and the Logiq GE E9™ Ultrasound System (General Electric Medical Systems Ltd, Chalfont St Giles, UK), respectively

employing 12MHz and 15MHz probes for all assessments. A total of 5 rheumatology trainees, as well as 2 allied healthcare trainees (one musculoskeletal physiotherapist and one radiographer) carried out MSUS assessments during the course of the current study, initially under direct supervision by appropriately experienced consultants (BT, AGP); where trainees were deemed competent to acquire images/complete proformas without direct supervision, operator identity was recorded to facilitate sensitivity analyses, with images regularly reviewed by trainers to ensure consistent semi-quantitative scoring. Good inter-observer (and excellent intra-observer) agreement of MSUS operators in the true-to-life environment described has previously been reported by our group[8, 26]. Sonographers did not have access to radiographs of hands and feet when making their assessments. ***Data management and analysis***. All baseline clinical and demographic data was contemporaneously recorded on a bespoke database. Because no added predictive value was conferred by considering any of these parameters as dichotomous (E.g. early morning stiffness > or $\leq$1 hour), all variables were treated as continuous where possible. Baseline semi-quantitative GS, PD and erosion scores of individual joints and "scan impression" MSUS data, as well as prospectively confirmed diagnostic and DMARD treatment data, were transcribed to the same database from clinic proformas after each visit. All analyses were carried out using JMP Pro 13.0 (SAS Institute Inc) using methods previously outlined [8]. Briefly, Student's *t*-tests or Mann-Whitney U tests and Pearson's chi-squared were used to describe data, and stepwise logistic regression and the construction of receiver operator characteristic (ROC) curves were carried out to determine non-redundant clinical parameters and their combined discriminatory utility with respect to diagnostic outcome. Paired and unpaired comparisons of AU ROCs and their 95% confidence intervals were calculated using DeLong's non-parametric method or approximated from 1000 bootstrap samples. Differences in AUROCs were tested using an approximate chi-square test for the difference in AUCs relative to their pooled standard errors respectively. In order to help conceptualise the potential clinical relevance of these results, the reciprocal of the difference between compared ROC curves was taken as an index of the number of individuals needed for MSUS data to influence one

diagnostic decision when considered alongside the most discriminatory clinical parameters (denoted *number needed to influence*, NNI). This NNI is averaged across all specificities for a pair of "risk metrics" that compare "clinical" with "clinical + MSUS" parameters, acknowledging that in reality it depends upon the prevalence of PIA in the population studied[27, 28].

Supplementary Figure.

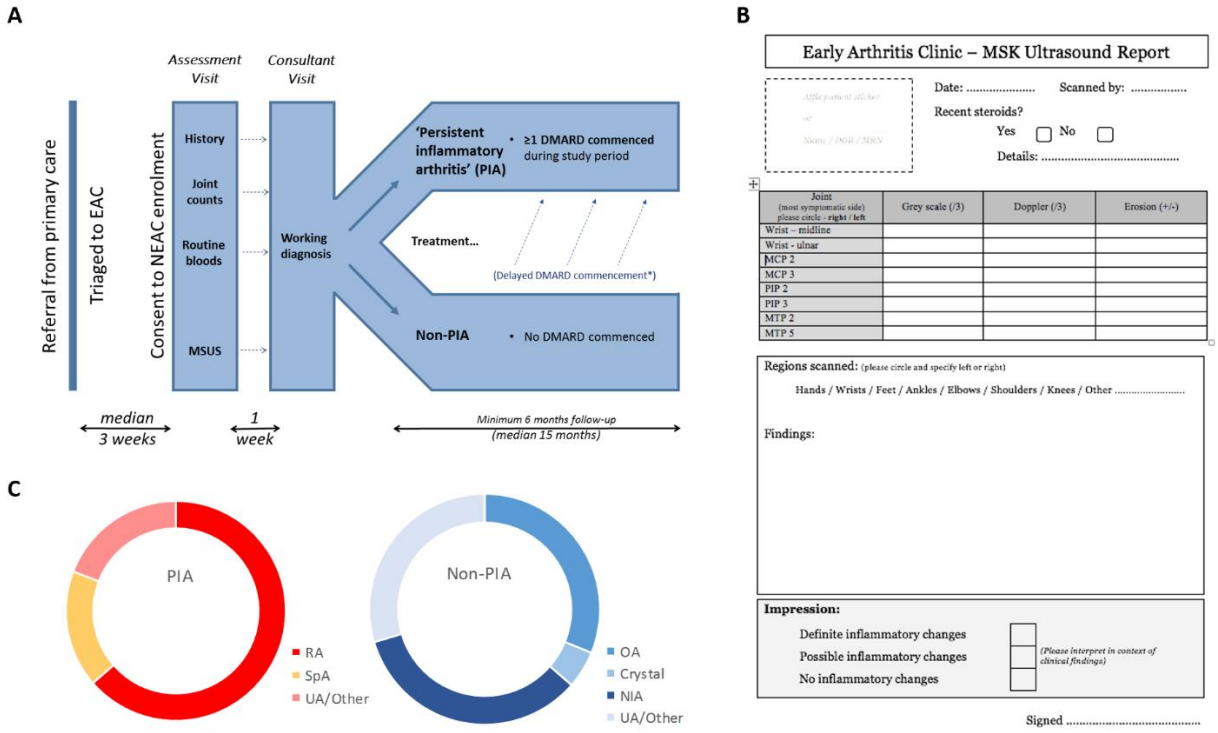

Figure S1. ***A***. *Scheme for patient referral, assessment, diagnosis and follow-up for 831 patients described in the observational inception cohort. See text; NEAC: Newcastle Early Arthritis Cohort.* ***B***. *Musculoskeletal (MSK) ultrasound report pro forma employed in routine clinical practice in the NEAC; semiquantiative GS and PD scores recorded in the upper table, together with the "Scan Impression" (lower portion) were incorporated into analyses as described.* ***C***. *Distribution of clinical diagnoses assigned within the PIA and non-PIA cohorts as defined at the end of the study's follow-up period. RA: rheumatoid arthritis; SpA: spondyloarthritis (includes psoriatic arthritis, ankylosing spondylitis and "other spondyloarthritis" categories from Table 1); OA: osteoarthritis; NIA: non-inflammatory arthralgia; Crystal: includes gout and pyrophosphate deposition; UA/Other: additional diagnostic assignments not otherwise accounted for amongst PIA/non-PIA categories.*

Supplementary Tables.

Working diagnosis	Number (%) assigned diagnosis (total = 831)	
	Inception	Close of follow-up (median 15 months)
Undifferentiated	75 (9.0)	65 (7.8)
Rheumatoid arthritis	158 (19.0)	171 (20.6)
Psoriatic arthritis	56 (6.7)	55 (6.6)
Ankylosing Spondylitis	6 (0.7)	5 (0.6)
Other Spondyloarthritis	36 (4.3)	35 (4.2)
Crystal arthritis	31 (3.7)	30 (3.6)
Connective tissue disease	12 (1.4)	12 (1.4)
Non-inflammatory arthralgia	385 (46.3)	382 (46.0)
Other	72 (8.7)	76 (9.1)

Table S1. *Clinical diagnoses assigned at inception and follow-up for all enrolled patients.*

Variable	**Coding if categorical**		**B**	**SE**	**Wald**	**P value**	**OR (95% CI)**
Complete cohort (n=831)							
ACPA	Pos	1			120.71	<.0001	30.77 (16.7-56.7)
	Neg	0	-1.71	0.16		<.0001	0.03 (0.02-0.06)
SJC	n/a		0.25	0.04	38.98	<.0001	1.29 (1.19-1.39)
CRP	n/a		0.02	0.01	19.88	<.0001	1.02 (1.01-1.03)
Age	n/a		0.02	0.01	8.97	0.0027	1.02 (1.01-1.03)
Constant	n/a		-1.41	0.37	14.38	-	-
ACPA-negative sub-cohort (n=703)							
SJC	n/a		0.24	0.04	34.70	<.0001	1.27 (1.18-1.38)
CRP	n/a		0.02	0.01	18.73	<.0001	1.02 (1.01-1.03)
Age	n/a		0.02	0.01	9.65	0.0018	1.02 (1.01-1.04)
Constant	n/a		-3.19	0.38	68.89	-	-

Table S2*. Results of backward stepwise logistic regression analysis to identify independent clinical (non-MSUS) predictors of PIA amongst EA clinic attendees. B values are regression coefficients. CI, confidence interval; OR, odds ratio; SE, standard error of B. Additional abbreviations as per Table 2.*

Variable	Coding if categorical	B	SE	Wald	P value	OR (95% CI)
Complete cohort (n=831)						
ACPA	*Positive*				<.0001	32.15 (16.7-61.9)
	Negative	-1.74	0.17	107.61	<.0001	0.03 (0.02-0.06)
SJC	*n/a*	0.22	0.05	21.76	<.0001	1.24 (1.14-1.36)
CRP	*n/a*	0.02	0.01	11.30	<.0001	1.02 (1.01-1.03)
Age	*n/a*	0.00	0.01	0.02	ns	0.99 (0.98-1.01)
Synovitis Load	***n/a***	***0.12***	***0.02***	***28.65***	***<.0001***	***1.13 (1.08-1.19)***
Constant	*n/a*	-0.91	0.41	4.80	-	-
ACPA-negative sub-cohort (n=703)						
SJC	*n/a*	0.21	0.05	19.63	<.0001	1.24 (1.12-1.36)
CRP	*n/a*	0.02	0.01	10.27	0.0014	1.02 (1.01-1.03)
Age	*n/a*	-0.00	0.01	0.01	ns	1.00 (0.98-1.02)
Synovitis Load	***n/a***	***0.13***	***0.02***	***26.28***	***<.0001***	***1.13 (1.08-1.19)***
Constant	*n/a*	-2.64	0.42	38.87	-	-

Table S3*. Results of logistic regression to demonstrate independent predictive value of synovitis load (defined as sum of GS and PD scores from 7 mandated joint areas as described in text) with respect to PIA diagnosis when considered alongside clinical parameters previously observed to have independent discriminatory value. Synovitis Load adds value over and above Clinical Parameters especially for ACPA-negative patients.*

Variable	Coding if categorical	B	SE	Wald	P value	OR (95% CI)
Complete cohort (n=831)						
ACPA	*Positive*				<.0001	38.01 (18.7-77.5)
	Negative	-1.82	0.18	100.35	<.0001	0.03 (0.01-0.05)
SJC	*n/a*	0.14	0.04	11.77	0.0006	1.15 (1.06-1.24)
CRP	*n/a*	0.04	0.01	7.02	0.0081	1.01 (1.00-1.03)
Age	*n/a*	0.00	0.01	0.18	ns	1.00 (0.99-1.02)
Scan Impression	*None-Possible Synovitis[0,1]*	***0.84***	***0.37***	***5.33***	***<.0209***	***0.15 (0.08-0.29)***
	Possible-Definite synovitis[1-2]	***1.87***	***0.32***	***33.81***	***<.0001***	***6.49 (3.5-12.20)***
Constant	*n/a*	-1.60	0.41	15.26	-	-
ACPA-negative sub-cohort (n=703)						
SJC	*n/a*	0.14	0.04	11.23	0.0008	1.14 (1.06-1.25)
CRP	*n/a*	0.01	0.01	5.78	0.0162	1.01 (1.00-1.02)
Age	*n/a*	0.00	0.01	0.08	ns	1.00 (0.99-1.02)
Scan Impression	*None-Possible synovitis[0,1]*	***1.27***	***0.41***	***9.60***	***0.0020***	***0.15 (0.08-0.29)***
	Possible-Definite synovitis[1-2]	***1.73***	***0.32***	***28.88***	***<.0001***	***5.68 (3.0-10.63)***
Constant	*n/a*	-3.60	0.45	62.73	-	-

Table S4*. Results of logistic regression to demonstrate independent predictive value of global scan impression (as defined in text) with respect to PIA diagnosis when considered alongside clinical parameters previously observed to have independent discriminatory value. Estimates for Scan Impression are for the odds of PIA diagnosis comparing a score of 0 (None) to 1 (Possible), and for 1 (Possible) to 2 (Definite). Scan Impression adds value over and above Clinical Parameter especially for ACPA-negative patients.*

References for Supplementary Methods.

1 Emery, P. *et al.* Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. *Annals of the Rheumatic Diseases* **61**, 290-297.

2 Pratt, A. G. *et al.* A CD4 T cell gene signature for early rheumatoid arthritis implicates interleukin 6-mediated STAT3 signalling, particularly in anti-citrullinated peptide antibody-negative disease. *Annals of the Rheumatic Diseases* **71**, 1374-1381 (2012).

3 Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* **62**, 2569-2581, doi:10.1002/art.27584 (2010).

4 Lage-Hansen, P. R., Lindegaard, H., Chrysidis, S. & Terslev, L. The role of ultrasound in diagnosing rheumatoid arthritis, what do we know? An updated review. *Rheumatology International* **37**, 179-187 (2017).

5 Backhaus, M. *et al.* Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: a pilot project. *Arthritis & Rheumatism* **61**, 1194-1201 (2009).

6 Szkudlarek, M. *et al.* Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance imaging, conventional radiography, and clinical examination. *Arthritis & Rheumatism* **50**, 2103-2112 (2004).

7 Wakefield, R. J. *et al.* Musculoskeletal ultrasound including definitions for ultrasonographic pathology.[erratum appears in J Rheumatol. 2006 Feb;33(2):440 Note: Bruyn, George [corrected to Bruyn, George AW]]. *Journal of Rheumatology* **32**, 2485-2487 (2005).

8 Szkudlarek, M. *et al.* Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis.[see comment]. *Arthritis & Rheumatism* **48**, 955-962 (2003).

9 Scheel, A. K. *et al.* A novel ultrasonographic synovitis scoring system suitable for analyzing finger joint inflammation in rheumatoid arthritis.[see comment]. *Arthritis & Rheumatism* **52**, 733-743 (2005).

10 Pratt, A. G., Lorenzi, A. R., Wilson, G., Platt, P. N. & Isaacs, J. D. Predicting persistent inflammatory arthritis amongst early arthritis clinic patients in the UK: is musculoskeletal ultrasound required? *Arthritis Research & Therapy* **15**, R118 (2013).

11 Baker, K. F., Thompson, B., Lendrem, D. W., Pratt, A. G. & Isaacs, J. D. Ultrasound Measures of Synovitis Are Independent of Clinical Parameters in The Setting of Rheumatoid Arthritis Remission: A Cross-Sectional Analysis. *Annals of the Rheumatic Diseases* **75 (Suppl. 2)**, 1228, doi:10.1136/annrheumdis-2016-eular.4131 (2016).

12 Bandolier. How good is that test? II. *Bandolier* **27**, 2 (1996).

13 Mitchell, A. J. Sensitivity × PPV is a recognized test called the clinical utility index (CUI+). *European Journal of Epidemiology* **26**, 251-252 (2011).