for all patients were reviewed through June 2015. Patients with a diagnosis of PMR prior to 2000, or at another care facility, and those with an incorrectly registered PMR diagnosis code, were excluded. In a structured review of the case records, information required for classification according to the ACR/EULAR criteria, the Bird criteria, the Healey criteria, the Chuang&Hunder criteria and the Jones&Hazelman criteria was extracted. For the ACR/EULAR criteria, a modified version, in which patients who had never been tested for rheumatoid factor (RF) and anti-citrullinated peptide antibodies (ACPA) only required 2 points to be classified as having PMR, was used. Furthermore, as duration of morning stiffness (MS) was usually not recorded, criteria components for MS were considered to be fulfilled whenever MS was mentioned in the records. The reference method was an independent review, with assessment of the long term disease course and differential diagnoses, by an experienced rheumatologist with access to all electronic records.

Results: A total of 305 patients with a registered diagnosis of PMR were reviewed. Of these, 117 were excluded. Among 188 with an incident PMR diagnosis at the study sites during the study period, 49 (26 %) fulfilled the modified ACR/EULAR criteria, whereas 145 (77 %) fulfilled the Bird criteria and 93 (49 %) fulfilled the Healey criteria. Patients could not be classified according to the Chuang&Hunder or the Jones&Hazelman criteria due to missing data in most patients for several components. RF and ACPA were tested in only 42 cases (4 positive) and 29 cases (none positive), respectively. The PMR diagnosis was verified using the reference method in 113 cases (60 % of total; 68 % female, mean age at diagnosis 75 years). Among those fulfilling the modified ACR/EULAR criteria, the diagnosis was verified in 84 % of the patients. The corresponding proportion for the Bird criteria was 66 %, and for the Healey criteria 74 %.

Conclusions: In this study of patients with PMR diagnosed in PHC, the diagnosis could be verified in 60 % of the patients. This underlines the heterogeneity of PMR patients and related diagnostic procedures in PHC. A modified version of the ACR/EULAR criteria can be used to identify patients with a valid PMR diagnosis in retrospective surveys, but does not capture all PMR patients. The modified ACR/EULAR criteria appear to be more stringent than some of the older criteria sets.

REFERENCE:

1. Dasgupta B, et al. Ann Rheum Dis 2012;71:484–92.

Disclosure of Interest: None declared

DOI: 10.1136/annrheumdis-2018-eular.5790

FRI0654 IS THE PATIENT-ACCEPTABLE STATUS SIMILAR ACROSS 7 DOMAINS OF HEALTH IN PATIENTS WITH RHEUMATOID ARTHRITIS (RA)? POST-HOC ANALYSES FROM THE VALIDATION IN 549 PATIENTS OF THE RHEUMATOID ARTHRITIS IMPACT OF DISEASE (RAID) SCORE

C. Duarte^{1,2,*}, T. Kvien³, J. A. Pereira da Silva^{1,2}, L. Gossec⁴ on behalf of RAID Working Group EULAR. ¹Centro Hospitalar e Universitário de Coimbra, ²Faculdade de Medicina Universidade de Coimbra, Coimbra, Portugal, ³Diakonhjemmet Hospital, Oslo, Norway, ⁴Sorbonne Université, Paris, France

Background: Patient Acceptable Symptom State (PASS) is the highest acceptable level of symptoms which patients consider satisfactory. In the Rheumatoid Arthritis Impact Disease (RAID) questionnaire, seven domains of health of importance for RA patients are collected.¹ It is possible that levels judged acceptable by patients vary according to the domain of health.

Objectives: To explore the relationship between seven RA domains of health (collected in the RAID) and PASS, and to define their PASS cut-off values.

Methods: This is a *post-hoc* analysis of the cross-sectional study for RAID validation. Each of 7 domains (table 1) was evaluated through a Numeric Rating Scale from 0 (best) to 10 (worst). PASS was calculated using the anchored method based on patients' perspective. Disease activity was assessed based on the DAS28–3 values (joint counts and ESR). Comparison of patients in PASS or not was assessed through Mann-Whitney or Chi-square test, as adequate. Variables with $p < 0.05$ were included in multivariate logistic regression (Forward Conditional) analysis. The thresholds of PASS for each domain was calculated using the receiver-operating characteristic (ROC) curve and the optimal cut-off was determined through Youden Index. 75th percentile analyses were also performed (not shown).

Results: 549 patients [78% female, mean age 56.7 years, mean disease duration 12.3 years, mean (SD) DAS 28–3 2.7(1.2)] were analysed. The majority of patients (60.7%) considered themselves to be in PASS. Disease activity (DAS 28–3 mean 2.3 vs 3.4, $p < 0.01$) and all seven domains of health were significantly lower in patients in PASS versus not in PASS ($p < 0.001$). In multivariate analyses, lower disease activity (OR 0.72; 95%IC 0.56–0.92), lower pain (OR 0.75; 95%

IC:0.65–0.86) and better physical well-being (OR 0.74 95%; IC:0.65–0.85) were associated with being in PASS.

The cut-off for PASS was ≤ 4.2 for the total RAID score but varied across the seven domains (table 1), with Pain (≤ 5) and Fatigue (≤ 5) having the highest acceptability cut-offs. Sleep Disturbance and Coping were the domains with lowest thresholds compatible with PASS (≤ 3).

Table 1 Thresholds for Acceptable Status for each RAID domain, AUC: Area Under the Curve, and sensitivity and specificity versus PASS

Domain	Cut-off	AUC	Sensitivity (%)	Specificity (%)
Pain	≤ 5	0.84	80.5	72.9
Function	≤ 4	0.82	69.7	80.8
Fatigue	≤ 5	0.75	74.9	62.2
Sleep	≤ 3	0.70	64.2	67.4
Emotional Well-Being	≤ 4	0.76	75.0	64.5
Physical Well Being	≤ 4	0.81	69.3	77.1
Coping	≤ 3	0.79	65.5	77.1

Conclusions: Pain and physical well-being appeared as major drivers of PASS. The cut-offs defining PASS were not the same for all RAID domains, indicating that being in PASS doesn't mean the same acceptable severity for all domains of health. This observation suggests that individualised management, for each domain, should be considered.

REFERENCE:

- 1) Gossec L, et al. Ann Rheum Dis 2011;70:935–942.

Acknowledgements: RAID Study Group: Balanescu A, Boumpas DT, Carmona L, de Wit M, Dijkman BA, Dougados M, Englbrecht M, Gogus F, Heiberg T, Kirwan JR, Mola EM, Cerinic MM, Otsa K, Sokka T

Disclosure of Interest: None declared

DOI: 10.1136/annrheumdis-2018-eular.2997

FRI0655 TITLE: THE CONSISTENCY OF OUTCOMES REPORTED IN TRIALS OF SYSTEMIC SCLEROSIS. IMPROVING OVER TIME?

D. J. Sumpton^{1,2,3,*}, A. Bigot⁴, B. Sautenet⁵, J. C. Craig^{2,3}, G. Hassett⁶, V. Thakkar^{7,8}, P. Tugwell⁹, A. Tong^{2,3}. ¹Department of Rheumatology, Concord Repatriation General Hospital, ²Sydney School of Public Health, The University of Sydney, ³Centre for Kidney Research, The Children's Hospital Westmead, Sydney, Australia, ⁴Service de Médecine Interne, ⁵Service de Néphrologie, Hôpital Bretonneau, CHRU Tours, Tours, France, ⁶Department of Rheumatology, Liverpool Hospital, Sydney, ⁷Department of Rheumatology, Liverpool Hospital, ⁸School of Medicine, Western Sydney University, Sydney, Australia, ⁹Department of Medicine, University of Ottawa, Ottawa, Canada

Background: Standardisation of outcome domains and measures in trials in rheumatology has generally improved the consistency and relevance of outcomes reported in rheumatic conditions over the last two decades. The Outcome Measures in Rheumatology (OMERACT) 'core response set' for trials in Systemic Sclerosis (SSc) was developed in 2008 and comprises 11 domains and 31 measures¹.

Objectives: We aimed to assess the scope, and consistency of outcomes reported in trials of SSc, and the uptake of this core response set.

Methods: MEDLINE, Cochrane CENTRAL, Embase and clinicaltrials.gov were searched to identify randomised clinical trials published between 2000 and June 2016 in adults with SSc. Outcomes and measures were recorded for each trial and classified into domains. The scope and consistency of domains were compared between those trials published from 2000 to 2010 and 2011 to mid-2016 to determine whether has been uptake of the core domain and measurement set reported. A two-year lag between publication of the core set in 2008 and start of the second-time period was set to allow for core response set uptake.

Results: Overall 114 trials (4860 patients, median sample size of 33) were identified. The majority of trials were interventions for immunotherapeutic agents or vasodilators and included a total of 2736 measures (of 78 domains), with a mean of 24 measures per trial. The proportion of trials reporting any outcome from the each domain is listed: (% of 2000–2010 trials, % of 2011–2016 trials, change in %): health-related quality of life and function (42.6, 56.5, +13.9); skin (39.7, 47.8, +8.1); pulmonary (33.8, 43.5, +9.7); global health (14.7, 21.7, +7.0); gastrointestinal (4.4, 10.9, +6.5); ; cardiac (13.2, 15.2, +2); biomarkers of ESR/CRP (7.4, 8.7, +1.3%) musculoskeletal (5.9, 6.5, +0.6); Raynaud's phenomenon (20.6, 19.6, -1.0) renal (14.7, 10.9, -3.8); digital ulcers (23.5, 19.6, -3.9%). Six measures had a greater than 5% increase in reporting frequency across trial periods: HAQ-DI (from 38.2% to 47.8%), SF-36 (from 17.6% to 28.3%), Modified Rodnan Skin

Score (from 38.2 % to 47.8%), pulmonary function tests (from 30.9% to 43.5%), measures of dyspnoea (4.4% to 10.9%) and patient global disease (from 10.3% to 21.7%).

Conclusions: There was a wide range of domains reported in trials in systemic sclerosis. The uptake of domains and measure as per the core response set is low in SSc trials compared to other rheumatic diseases with only modest improvement in reporting of 6 out of 31 measures. Improvements in reporting of specific measures align with the recent development of a composite response index in systemic sclerosis (CRISS)2.

REFERENCES:

1. Khanna D, et al. Development of a provisional core set of response measures for clinical trials of systemic sclerosis, *Ann Rheum Dis* 2008;67:703–9.
2. Khanna D, et al. The American College of Rheumatology Provisional Composite Response Index for Clinical Trials in Early Diffuse Cutaneous Systemic Sclerosis. *Arthritis Res Ther* 2016;68:299–311.

Disclosure of Interest: None declared

DOI: 10.1136/annrheumdis-2018-eular.6001

FRI0656 PREDICTIVE VALIDITY OF PRESENTEEISM MEASURES WITH DUAL ANSWER KEYS IN INFLAMMATORY ARTHRITIS

M. Donaldson^{1,2}, A. Kobza^{2,3}, D. E. Beaton⁴, H. Xie^{2,5}, E. C. Sayre², M. A. M. Gignac^{4,6}, D. Lacaille^{1,2}. ¹University of British Columbia, Vancouver, ²Arthritis Research Canada, Richmond, ³University of Ottawa, Ottawa, ⁴Institute of Work and Health, Toronto, ⁵Simon Fraser University, Burnaby, ⁶University of Toronto, Toronto, Canada

Background: Employment studies in arthritis have emphasized the importance of decreased productivity at work, or presenteeism. However, consensus is still lacking on how to best measure this construct. Instruments used include the Work Limitations Questionnaire (WLQ), which measures the amount of time people are limited, and the Workplace Activity Limitations Scale (WALS), which measures the degree of difficulty. We modified the response keys of the WLQ and WALS, creating a dual answer key assessing both degree of difficulty and amount of time with difficulty. Previous work by our group has shown that scores obtained from combining responses to the dual answer keys exhibit good internal consistency, criterion and construct validity.

Objectives: The objective of this study was to evaluate the ability of the combined scores to predict future work cessation and number of work days missed due to arthritis.

Methods: A longitudinal study used baseline and 6-month data from the RCT of an employment intervention, the "Making It Work" Program. Inclusion criteria were: inflammatory arthritis, age 19–59, working at baseline, and concerns about arthritis affecting work. 327 participants were included (RA:173, PsA:48, SLE:42, AS:64). Combined scores were obtained by i) multiplying, and ii) adding, the scores of difficulty and time answer keys at the item level. We assessed the predictive ability of the combined scores, and the original scores, of the WALS and WLQ to predict work cessation due to arthritis within 6 months using binary logistic regression models, or number of work days missed due to arthritis using zero-inflated Poisson models.

Results: WLQ and WALS original and combined scores were significant predictors of future work cessation and number of work days missed. The WLQ combined scores predicted work cessation better than the original score (AIC:117.6 & 118.2 for multiplicative and additive combinations, respectively vs. 125.4 for time) and work days missed (AIC:2179.8 & 2177.8 vs. 2188.7). WALS original score (ie. degree of difficulty) predicted work cessation better than combined scores (AIC:95.8 vs. 109.3 & 109.4). However, WALS combined scores predicted number of work days missed better (AIC:1696.9 & 1673.4 vs. 1795.2).

Conclusions: Combined scores from dual answer keys applied to the WLQ and WALS significantly predicted future work cessation and number of work days missed. Using combined scores, rather than original answer keys, improved the ability of the WLQ to predict both work cessation and work days missed, but only improved the ability of the WALS to predict work days missed.

Disclosure of Interest: None declared

DOI: 10.1136/annrheumdis-2018-eular.5970

FRI0657 CONSTRUCT VALIDITY AND RESPONSIVENESS OF THREE STIFFNESS ITEMS IN RHEUMATOID ARTHRITIS: AN APPLICATION OF THE OMERACT FILTER 2.1

E. T. Craig^{1,*}, C. O. Bingham III¹, S. J. Bartlett², U. Haque¹, A. Leong³, A.-M. Orbai¹. ¹Department of Rheumatology, Johns Hopkins University, Baltimore, United States, ²Center for Mind Body Research, McGill University, Montreal, Canada, ³Patient Research Partner, Santa Barbara, United States

Background: Stiffness is frequent and important to people with rheumatoid arthritis (RA), but measuring it is an unmet need. Qualitative research with people with RA has shown that stiffness duration, severity and interference have face and content validity. These items are being evaluated to develop a patient reported outcome measure (PROM) for stiffness.

Objectives: To evaluate individual stiffness item measurement properties using the OMERACT Filter 2.1.

Methods: Consecutive patients in a longitudinal RA cohort were enrolled. Participants completed 3 items assessing stiffness severity, interference, and duration. PROMs collected included patient global assessment (PGA), pain VAS, self-reported flare (Y/N), PROMIS measures, and patient global impression of change (PGIC). Physician assessments included 28 tender/swollen joint counts, and global (EGA). Disease activity indices were CDAI and DAS28-CRP. Stiffness construct validity was assessed using Spearman correlation coefficients. To assess discrimination in groups expected to change vs stable groups, we calculated standardized response means (SRMs) using multiple anchors.

Results: 196 patients with RA were included. **Construct validity:** all stiffness items correlated highly with pain; moderately with DAS-28 CRP and CDAI; and low with joint counts and EGA. Stiffness severity and interference correlated highly with PGA, pain interference and fatigue. Stiffness interference correlated highly with physical function. **Discrimination:** all stiffness items had moderate effect size for resolution of flare, and for 'much improved' by PGIC. Interference had moderate effect for improved CDAI category and for 'little improved' by PGIC. Stiffness duration had large effect for new RA flare, and 'much worse' by PGIC; severity and interference had moderate effect for RA flare; severity had moderate effect for worsened CDAI and for 'little worse' and 'much worse' by PGIC (TbI).

Table 1 Discrimination of stiffness items by anchor in RA longitudinal sample (n=181). SRM=mean change/pooled SD of change. SRM >0.8 large, 0.5–0.79 moderate, 0.2–0.5 small effect.

Stiffness Anchors	Duration SRM	Severity SRM	Interference SRM
Flare status (n=181)	0.67	0.68	0.54
Flare to No flare (improved) (n=26)	0.01	0.01	0.14
No flare to No flare (unchanged) (n=111)	0	0.19	0.27
Flare to Flare (unchanged) (n=14)	0.81	0.60	0.54
No flare to Flare (worse) (n=30)			
PGIC vs baseline (n=181)	0.52	0.61	0.79
Much improved (n=26)	0.43	0.46	0.60
Little improved (n=25)	0.08	0.08	0.11
No change (n=68)	0.41	0.67	0.42
Little worse (n=49)	1.05	0.59	0.39
Much worse (n=13)			
CDAI vs baseline (n=181)	0.25	0.21	0.61
Better CDAI category (n=45)	0.07	0.12	0.04
Unchanged CDAI category (n=87)	0.48	0.52	0.29
Worse CDAI category (n=49)			

Conclusions: In this cohort of RA patients there is emerging evidence for good construct validity for stiffness duration, severity and interference; moderate discrimination for improvement in RA disease state, and moderate to high discrimination for worsening RA. Additional items are being tested to assemble an instrument for measuring stiffness in RA.

Disclosure of Interest: None declared

DOI: 10.1136/annrheumdis-2018-eular.2142

FRI0658 PROGNOSTIC MARKERS FOR RESPONSE ACCORDING TO THE NEW ACR/EULAR 2016 RESPONSE CRITERIA FOR IDIOPATHIC INFLAMMATORY MYOSITIS

E. Espinosa-Ortega¹, M. Holmqvist², M. Dastmalchi¹, I. Lundberg¹, H. Alexanderson¹. ¹Rheumatology Unit, Department of Medicine, Karolinska Institutet, ²Institutionen för medicin, Solna, Stockholm, Sweden

Background: Idiopathic inflammatory myopathies (IIM) is a heterogeneous group of chronic, inflammatory diseases with varying response to treatment. To