Results: The study included 30 patients with SLE. Internal consistency of the MFIS was excellent with Cronbach's $\alpha = 0.97$ for the complete scale. Excellent test-retest reliability was found with ICC = 0.95 (95% confidence interval: 0.88–0.98, $p < 0.05$). Construct validity was confirmed by Spearman's correlation (VT-SF36: $r_s = -0.73$, $p < 0.001$ (Fig. 1). MH-SF36: $r_s = -0.74$, $p < 0.001$ (Fig. 2)) and PCA with explained variance from the first two principal components (PC) (VT-SF36: PC1 = 60.2%, PC2 = 8.5%. MH-SF36: PC1 = 58.5%, PC2 = 7.4%). No significant correlation was found between the MFIS and SLEDAI ($r_s = 0.04$, $p = 0.84$) or SLICC Damage Index ($r_s = 0.32$, $p = 0.08$).

Figure 1. Scatter plot of the Modified Fatigue Impact Scale (MFIS) and the Short Form 36 vitality (VT-SF36) subscale.

Figure 2. Scatter plot of the Modified Fatigue Impact Scale (MFIS) and the Short Form 36 mental health (MH-SF36) subscale.

Conclusion: The present study found the multidimensional assessment of fatigue with MFIS to be a reliable and valid instrument in SLE. The MFIS might provide more detailed information about fatigue in future studies. In agreement with some previous studies we found no association between fatigue and SLEDAI or SLICC which raises

questions about the cause of this symptom. Further and larger studies are needed to investigate if any association between fatigue and disease components exist.

Disclosure of Interests: None declared

DOI: 10.1136/annrheumdis-2020-eular.1293

FRI0584 SCREENING AND IDENTIFICATION OF BIOMARKERS IN MYOSITIS

W. Liu¹, X. Zhang¹ on behalf of Department of Rheumatology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China. ¹Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Rheumatology, Guangzhou, China

Background: Myositis, including dermatomyositis and polymyositis, is autoimmune disorders that is characterized by muscle degeneration in the proximal extremities, with the complications of weakness of muscles, interstitial lung disease and vascular lesions, even leading to death in an acute progressive process^[1,2]. However, the molecular mechanisms of myositis are rarely understood.

Objectives: Identify the candidate genes in myositis.

Methods: Microarray datasets GSE128470, GSE48280 and GSE39454 were extracted from the Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) and function enrichment analyses were conducted. The protein-protein interaction network and the analyses of hub genes were performed with STRING and Cytoscape.

Results: There were 98 DEGs, of which the function and pathways enrichment analyses showed defense response, immune response, response to virus, inflammatory response, response to wounding, cell adhesion, cell proliferation, cell death and macromolecule metabolic process. 20 hub genes were identified, of which 7 including IRF9 TRIM22 MX2 IFITM1 IFI6 IFI44 IFI44L had not been reported in the literature, related to the response to virus, immune response, transcription from RNA polymerase II promoter, cell apoptosis, cell death. The verification analysis about the 7 genes in GSE128314 showed significant differences in myositis.

Conclusion: In conclusion, DEGs and hub genes identified in our study showed the potential molecular mechanisms in myositis, providing the helpful targets for diagnosis and clinical strategy of myositis.

References:

- [1] Wu H, Geng D, Xu J. An approach to the development of interstitial lung disease in dermatomyositis: a study of 230 cases in China[J]. Journal of International Medical Research. 2013;41(2):493–501.
- [2] Fathi M, Dastmalchi M, Rasmussen E, Lundberg IE, Tornling G. Interstitial lung disease, a common manifestation of newly diagnosed polymyositis and dermatomyositis[J]. Annals of the Rheumatic Diseases. 2004;63(3):297–301.

Figure 1. The protein-protein interaction network of 20 hub genes

Figure 2. 7 genes in GSE128314 showed significant differences in myositis

Acknowledgments: The authors acknowledge the efforts of the Gene Expression Omnibus (GEO) database. The interpretation and reporting of these data are the sole responsibility of the authors.

Disclosure of Interests: None declared

DOI: 10.1136/annrheumdis-2020-eular.1831

FRI0585 HIGH-THROUGHPUT METHODOLOGY FOR EMR-BASED IDENTIFICATION OF CLINICAL SUB-PHENOTYPES IN COMPLEX PATIENT POPULATIONS

M. Maurits¹, T. Huizinga¹, M. Reinders^{2,3}, S. Raychaudhuri^{4,5}, E. Karlson⁵, E. Van den Akker^{2,3}, R. Knevel^{1,4}. ¹Leiden University Medical Center (LUMC), Rheumatology, Leiden, Netherlands; ²Delft University of Technology (TU Delft), Bioinformatics, Delft, Netherlands; ³Leiden University Medical Center (LUMC), Molecular Epidemiology, Leiden, Netherlands; ⁴Brigham and Women's Hospital, Rheumatology, Boston, United States of America; ⁵Brigham and Women's Hospital, Data Sciences, Boston, United States of America

Background: Heterogeneity in disease populations complicates discovery of risk factors. To identify risk factors for subpopulations of diseases, we need analytical methods that can deal with unidentified disease subgroups.

Objectives: Inspired by successful approaches from the Big Data field, we developed a high-throughput approach to identify subpopulations within patients with heterogeneous, complex diseases using the wealth of information available in Electronic Medical Records (EMRs).

Methods: We extracted longitudinal healthcare-interaction records coded by 1,853 PheCodes[1] of the 64,819 patients from the Boston's Partners-Biobank. Through dimensionality reduction using t-SNE[2] we created a 2D embedding of 32,424 of these patients (set A). We then identified distinct clusters post-t-SNE using DBscan[3] and visualized the relative importance of individual PheCodes within them using specialized spectrographs. We replicated this procedure in the remaining 32,395 records (set B).

Results: Summary statistics of both sets were comparable (Table 1). We found 284 clusters in set A and 295 in set B, of which 63.4% from set A could be mapped to a cluster in set B with a median (range) correlation of 0.24 (0.03 – 0.58).

Table 1. Summary statistics of the total Partners Biobank dataset and the 2 partitions.

	Set-A	set-B	Total
Entries	12,200,311	12,177,131	24,377,442
Patients	32,424	32,395	64,819
Patientyears	369,546.33	368,597.92	738,144.2
unique ICD codes	25,056	24,953	26,305
unique Phecodes	1,851	1,853	1,853

Figure 1. Two dimensional representation of Set A generated using dimensionality reduction (tSNE) and clustering (DBScan).

Clusters represented similar yet distinct clinical phenotypes; e.g. patients diagnosed with “other headache syndrome” were separated into four distinct clusters characterized by migraines, neurofibromatosis, epilepsy or brain cancer, all resulting in patients presenting with headaches (Fig. 1 & 2). Though EMR databases tend to be noisy, our method was also able to differentiate misclassification from true cases; SLE patients with RA codes clustered separately from true RA cases.

Figure 2. Phenotype Spectrographs (PheSpecs) of four clusters characterized by “Other headache syndromes”, driven by codes relating to migraine, epilepsy, neurofibromatosis or brain cancer.

Conclusion: We have shown that EMR data can be used to identify and visualize latent structure in patient categorizations, using an approach based on dimension reduction and clustering machine learning techniques. Our method can identify misclassified patients as well as separate patients with similar problems into subsets with different associated medical problems. Our approach adds a new and powerful tool to aid in the discovery of novel risk factors in complex, heterogeneous diseases.

References:

- [1] Denny, J.C. et al. Bioinformatics (2010)
- [2] van der Maaten et al. Journal of Machine Learning Research (2008)
- [3] Ester, M. et al. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. (1996)

Disclosure of Interests: Marc Maurits: None declared, Thomas Huizinga Grant/research support from: Ablynx, Bristol-Myers Squibb, Roche, Sanofi, Consultant of: Ablynx, Bristol-Myers Squibb, Roche, Sanofi, Marcel Reinders: None declared, Soumya Raychaudhuri: None declared, Elizabeth Karlson: None declared, Erik van den Akker: None declared, Rachel Knevel: None declared

DOI: 10.1136/annrheumdis-2020-eular.3489

FRI0586 HOW TO GET FROM THE MULTIDIMENSIONAL HEALTH ASSESSMENT QUESTIONNAIRE TO STANFORD HEALTH ASSESSMENT QUESTIONNAIRE DISABILITY INDEX SCORES IN PATIENTS WITH RHEUMATOID ARTHRITIS, PSORIATIC ARTHRITIS AND AXIAL SPONDYLOARTHRITIS: DEVELOPMENT AND VALIDATION OF A CONVERSION ALGORITHM

E. Svensson¹, K. Løngaard¹, L. Midtbøll Ørnberg², R. Meincke², J. K. Pedersen³, L. Dreyer⁴, N. Steen Krogh⁵, D. V. Jensen², M. L. Hetland². ¹RKKP - the Danish Clinical Quality Program, Aarhus, Denmark; ²DANBIO, Copenhagen Center for Arthritis Research, Rigshospitalet, Copenhagen, Denmark; ³Svendborg Hospital, Svendborg, Denmark; ⁴Aalborg University Hospital, Aalborg, Denmark; ⁵Zitelab Aps, Frederiksberg, Denmark

Background: In the DANBIO quality registry in Denmark, patients with rheumatoid arthritis (RA) psoriatic arthritis (PsA) and axial spondyloarthritis (axSpA) have reported Patient Reported Outcomes (PROs) including the Stanford Health Assessment Questionnaire Disability Index (HAQ-DI) for nearly twenty years as part of routine care. Patients' feedback have stressed a need for a shorter registration of disability (1). While the shorter Multidimensional Health Assessment Questionnaire (MDHAQ) is preferred by patients, the original HAQ-DI is the preferred tool in observational studies. Thus, a conversion algorithm between the MDHAQ and HAQ-DI scores is warranted.

Objectives: To develop and validate a simple conversion algorithm between MDHAQ and HAQ-DI scores in RA, PsA and axSpA patients.

Methods: Patients registered in DANBIO with a diagnosis of RA, PsA or axSpA who had completed both HAQ-DI and MDHAQ simultaneously at a visit +/- 30 days from start of conventional synthetic (cs)DMARD or biological (b)DMARD were eligible for the analysis, and randomly divided into development and validation cohorts stratified by diagnosis. The conversion algorithm was developed in the RA development cohort using linear regression with HAQ-DI as the dependent variable and MDHAQ as the independent variable. The predicted HAQ (pHAQ) scores were then