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

FRIO585

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

M. Maurits1, T. Huizinga1, M. Reinders2,3, S. Raychaudhuri4,5, E. Karlsson6, E. Van den Akker2,3, R. Knevel4,1, Leiden 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 PheCodes1 of the 64,819 patients from the Boston's Partners-Biobank. Through dimensionality reduction using t-SNE2 we created a 2D embedding of 32,424 of these patients (set A). We then identified distinct clusters post-t-SNE using DBscan3 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).

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

<table>
<thead>
<tr>
<th></th>
<th>Set-A</th>
<th>Set-B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entries</td>
<td>12,200,311</td>
<td>12,177,131</td>
<td>24,377,442</td>
</tr>
<tr>
<td>Patients</td>
<td>32,424</td>
<td>32,395</td>
<td>64,819</td>
</tr>
<tr>
<td>Patientyears</td>
<td>369,546.33</td>
<td>366,597.92</td>
<td>736,144.2</td>
</tr>
<tr>
<td>Unique ICD codes</td>
<td>25,056</td>
<td>24,953</td>
<td>26,305</td>
</tr>
<tr>
<td>Unique Phecodes</td>
<td>1,851</td>
<td>1,853</td>
<td>1,853</td>
</tr>
</tbody>
</table>

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).

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:


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

FRIO586

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. Langgaard1, L. Midthull Overberg2, R. Meircke3, J. K. Pedersen4, L. Dreyer5, N. Steen Krogh5, D. V. Jensen5, M. L. Hetland4,1, RKKP - 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 the 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

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

Figure 2. Phenotype Spectrographs (PheSpecs) of four clusters characterized by “Other headache syndromes”; driven by codes relating to migraine, epilepsy, neurofibromatosis or brain cancer.
calculated by applying the conversion algorithm to the MDHAQ scores in the RA, PsA and axSpA validation cohorts. The pHAQ was validated against the HAQ-DI in the validation cohorts regarding criterion, correlational and construct validity.

**Results:** We included 8883/4410/1760 patients with RA/PsA/axSpA, respectively. The conversion algorithm pHAQ=0.15+MDHAQ*1.08 had the best fit (R²=0.83) in the RA validation cohort.

Criterion validity: The correlation coefficients between HAQ-DI/pHAQ and patient global score at baseline were 0.6/0.65. In groups of patients with high and low disability (defined as patient global score ≥50), standardized mean difference was -1.4 for HAQ-DI, and -1.4 for pHAQ.

Correlational validity: Correlation coefficients between HAQ-DI/pHAQ and ΔHAQ-DI/ΔpHAQ between baseline and first follow-up visit were r=0.91 and r=0.87, respectively. Correlation coefficients between HAQ-DI/pHAQ and pain score, DAS28CRP and physician global score were 0.63/0.64, 0.55/0.55 and 0.34/0.34, respectively. A Bland-Altman plot showed good agreement of HAQ-DI and pHAQ across all functional states.

Construct validity: HAQ-DI/pHAQ at the first follow-up visit after baseline was comparable between Patient Acceptable Symptom State groups (PASS=No: mean 1.17 vs 1.18/PASS=Yes: 0.55 vs 0.60). Similar results were seen for the external anchor (Figure 1).

In PsA and axSpA validation cohorts, similar results were found.

**Conclusion:** In PsA and axSpA validation cohorts, similar results were found.

**References:**

![Figure 1](http://ard.bmj.com/Ann Rheum Dis: first published as 10.1136/annrheumdis-2020-eular.1292 on 2 June 2020. Downloaded from http://ard.bmj.com)

**Construct validity:** Mean of HAQ-DI/pHAQ and ΔHAQ-DI/ΔpHAQ across external anchor responses.

**Acknowledgments:** The authors thank all Danish patients and Departments of Rheumatology, who conscientiously report to the DANBIO registry.

**Disclosure of Interests:** Elisabeth Svensson: None declared. Katja Løngaard: None declared. Merete L. Hetland: None declared. Jens Kristian Pedersen: None declared. Lene Jensen: None declared. Merete L. Hetland Grant/research support from: BMS, Dr. Pincus holds a copyright and trademark on MDHAQ and RAPID3 for which he receives royalties and license fees from profit-making organizations, Theodore Pincus Shareholder of: Rush University, Rheumatology, Chicago, United States of America; ZtieLab, Copenhagen, Denmark.

**Background:** A multidimensional health assessment questionnaire (MDHAQ) includes RAPID3, which distinguishes active from control treatments in rheumatoid arthritis clinical trials, and documents change comparatively to disease-specific indices in all diseases studied. The MDHAQ also includes a standard, structured 60-symptom checklist, to recognize comorbidities, provide review of systems, and serve on a fibromyalgia assessment screening tool (FAST3) as a clue to identify patients with fibromyalgia. A new MDHAQ is to recognize adverse events to high-risk medications on a standard, structured, protocol-driven MDHAQ 60-symptom checklist. A structured list, rather than a ‘subjective’ narrative medical history, is needed as many adverse events are common symptoms, e.g., headache, fatigue; prior negative data facilitates recognition of a new symptom as a possible adverse event. Similar strategies have been reported in oncology, pulmonology and other specialties, but not in rheumatology.

**Objectives:** To use a remote electronic MDHAQ, completed weekly at home, to recognize RAPID3 clinical status changes and adverse events on the 60-symptom checklist, for early detection of medication adverse events.

**Methods:** All patients with all diagnoses complete an MDHAQ at all visits in routine care at one rheumatology site. An electronic flowsheet (Table) is used to monitor 0-30 RAPID3, its components, and report of specific symptoms on the 60-symptom checklist, which appears required to document earlier absence of a common symptom and signal that a common symptom may be an adverse event. Results are depicted for an individual patient with pulmonary fibrosis, seen because of a positive rheumatoid factor.

**Results:** A flowsheet of a pulmonary fibrosis patient over 2018 indicates initial RAPID3 of 14/30 and 10 symptoms at first visit of 19 Jan (Flowsheet). Treatment with low-dose methotrexate (MTX) and prednisone (PRED) led to substantial improvement over 6 months - RAPID3 3.5 and 6 symptoms on 2 Aug. On 15 Aug, MTX and PRED were discontinued by another physician, who prescribed pirfenidone. The patient telephoned on 24 Sep indicating distress. A home-completed remote MDHAQ indicated RAPID3 of 19.5 and 15 symptoms - 7 not reported on 2 Aug were among 16 listed pirfenidone adverse events. Discontinuation of pirfenidone and resumption of PRED and MTX with weekly remote electronic MDHAQ monitoring documented improvement of RAPID3 to 4.2 and 6 symptoms, including resolution of pirfenidone-specific symptoms, on 24 Dec (Flowsheet).

**Conclusion:** Weekly remote electronic MDHAQ monitoring after initiation of a high-risk medication to monitor treatment responses and adverse events may provide a cost-effective approach to reduce morbidity and mortality of adverse events, involving about 10 minutes weekly (2 hours over 12 weeks) of patient time. 78-year-old man monitored over 2018—all data from self-report on MDHAQ – pirfenidone highlighted (many entries deleted for space considerations).

**Disclosure of Interests:** Theodore Pincus Shareholder of: Dr. Pincus holds a copyright and trademark on MDHAQ and RAPID3 for which he receives royalties and license fees from profit-making organizations, all of which are used to support further development of quantitative clinical measures for patients and health professionals., None declared. Lykke Midtbøll Ørnbjerg Grant/research support from: Novartis, None declared. Niels Steen Krogh: None declared. Dorte Vendelbo Rikke Meincke: None declared. Jens Kristian Pedersen: None declared. Lene Jensen: None declared. Merete L. Hetland Grant/research support from: BMS, None declared.