Exploring Heterogeneity in Rheumatoid Arthritis: Patient Profiling Through Principal Component and Cluster Analysis of the BRASS Registry

Jeffrey R. Curtis,[1] Michael Weinblatt,[2] Kenneth Saag,[1] Vivian P. Bykerk,[3] Christina Charles-Schoeman,[4] Stefano Fiore,[5] Gregory St John,[6] Toshio Kimura,[6] Shen Zheng,[5] Clifton Bingham,[7] Grace Wright,[8] Martin Bergman,[9] Kamala Nola,[10] Daniel E. Furst,[4] Nancy Shadick[2]

[1]University of Alabama at Birmingham, Birmingham, AL, USA; [2]Brigham and Women's Hospital, Boston, MA, USA; [3]Hospital for Special Surgery, New York, NY, USA; [4]University of California, Los Angeles, CA, USA; [5]Sanofi, Bridgewater, NJ, USA; [6]Regeneron Pharmaceuticals, Inc., Tarrytown, NY, USA; [7]Johns Hopkins University, Baltimore, MD, USA; [8]Private Practice, New York, NY, USA; [9]Drexel University College of Medicine, Philadelphia, PA, USA; [10]Lipscomb University College of Pharmacy & Health Sciences, Nashville, TN, USA

Poster number 183

ABSTRACT

Background/Purpose: Data driven principal component (PC) and cluster analysis has the potential to identify previously unknown patient subgroups within a rheumatoid arthritis (RA) registry to establish prognosis, predict disease trajectory, and help inform treatment. The Brigham and Women's Rheumatoid Arthritis Sequential Study (BRASS), established in 2003, is a single-center, prospective observational registry cohort providing a comprehensive set of clinical disease activity measures in >1400 patients with RA. Our objective was to use PC and cluster analysis of baseline demographic, socio-economic, health and disease characteristics in BRASS to identify and characterize distinct patient clusters in RA. **Methods:** Patient variables recorded at entry into BRASS were refined and combined using PC analysis to reduce dimensionality and collinearity. The number of PCs was established by eigenvalue >1, cumulative variance, and interpretability. Patients were clustered using a k-means approach with non-hierarchical, exclusive, and complete clustering, with minimum cluster size 5% of population, and maximum 19 clusters. The final number of clusters was determined according to the cubic clustering criterion and pseudo F. **Results:** Analysis of baseline data from 1443 patients identified 41 PCs that capture the fundamental characteristics involved in RA. Cluster analysis distinguished 5 patient clusters. Each cluster reflected a different profile of PCs, and can be described based on overall health, RA disease activity and duration. Key differentiators between clusters include comorbidity PCs (metabolic comorbidities predominate in cluster 4, neurologic in cluster 3, and orthopedic in cluster 5) and patient characteristics/social PCs (greatest number of doctor visits and family history of MI in cluster 2, greatest BMI in cluster 4, highest income in cluster 1, lowest income in cluster 5, and least emotional support in cluster 4). **Conclusion:** Data-driven cluster analysis of RA patient characteristics at entry into the BRASS registry identified five distinct patient phenotypes, providing a convenient method to potentially derive novel insights into the multifactorial drivers, commonly co-occurring health conditions, and manifestations of RA. Investigation of longitudinal outcomes in these different clusters in the BRASS registry and validation in an independent dataset is ongoing.

BACKGROUND AND OBJECTIVES

- Data-driven principal component (PC) and cluster analysis has the potential to identify previously unrecognized patient subgroups within a rheumatoid arthritis (RA) registry to help establish prognosis, predict disease trajectory, and inform treatment
- Established in 2003, the Brigham and Women's Rheumatoid Arthritis Sequential Study (BRASS)[1] is a single-center, prospective, observational registry cohort providing a comprehensive set of clinical disease activity measures in >1400 patients with RA
- We used data-driven PC and cluster analysis of baseline demographic, socio-economic, health, and disease characteristics in BRASS to identify patient phenotypes that may have different trajectories for damage/disease progression, different treatment responses, and different risks for adverse events

METHODS

- Patient variables recorded at entry into BRASS (described in reference 1) were harmonized using PC analysis to reduce dimensionality and collinearity. The number of PCs was established by eigenvalue >1, cumulative variance, and interpretability
- Patients were clustered using a k-means approach with minimum cluster size 5% of the population
- The final number of clusters was determined according to the cubic clustering criterion and pseudo F statistics
- Cluster space was visualized using canonical variables, which are linear combinations of all PCs that capture the maximal multiple correlation, derived sequentially and orthogonal to the previous linear combination

RESULTS

- Analysis of 142 variables from 1443 patients identified 41 PCs that capture key characteristics involved in RA and accounted for 77% of the cumulative variance in the dataset
- Cluster analysis distinguished five patient clusters (Figures 1 and 2)
- Each cluster reflected a different profile of PCs, and could be described according to general health, RA control, and RA duration (Table 1)
- Clinical outcomes by cluster up to 4 years will be reported in oral presentation 2850 on Tuesday 4.30–6.00 pm (Session 5T114)

Cluster 1

- **General health high, RA controlled, shorter RA duration**
  Had below-average score across comorbidity PCs and for the PC of doctor visits (reflected by mean 0.30 primary care physician [PCP] visits in the last 6 months)

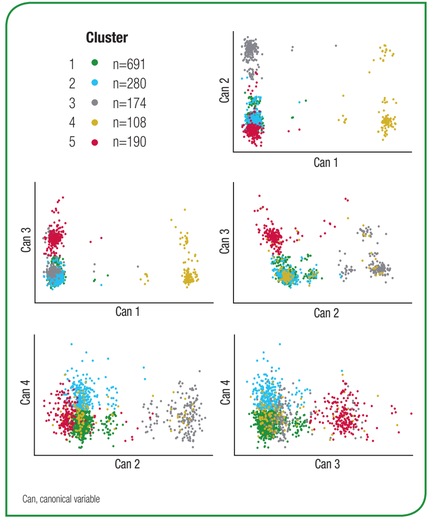
Cluster 2

- **General health high, RA controlled, longer RA duration**
  Exhibited above-average psychologic comorbidity (reflected by 51% prevalence of psychologic comorbidity ever), infection comorbidity (59% ever prevalence), and history of malignancy together with the highest score for doctor visits (reflected by mean 0.74 PCP visits in the last 6 months)

Cluster 3

- **General health low–moderate, moderate RA, moderate RA duration**
  Had the highest score for neurologic comorbidity, which predominantly consisted of migraine and dementia (ever prevalence 77% and 5%, respectively)

Figure 1. Cluster space visualization using canonical variables



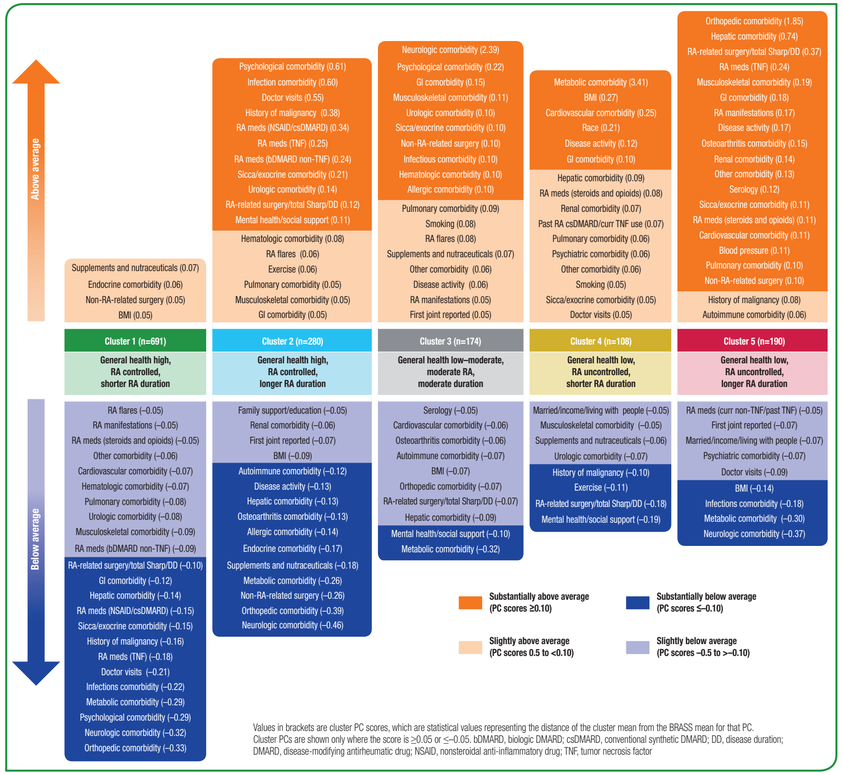
Cluster 4

- **General health low, RA uncontrolled, shorter RA duration**
  Was notable for a high prevalence of metabolic comorbidity, predominantly diabetes (ever prevalence 10% for type 1 and 80% for type 2). Cluster 4 also exhibited the highest scores for body mass index (BMI; median 31) and cardiovascular comorbidity

Cluster 5

- **General health low, RA uncontrolled, longer RA duration**
  Had the worst RA disease activity (reflected by 76% of patients having Clinical Disease Activity Index [CDAI] >10) and greatest prevalence of RA manifestations (35% and 22% for subcutaneous nodules and Sjögren syndrome, respectively). Cluster 5 also exhibited the highest score among the clusters for hepatic, gastrointestinal (GI), and autoimmune comorbidity

Figure 2. PCs across the five RA phenotype clusters, ranked by PC score indicating difference from overall BRASS average



Values in brackets are cluster PC scores, which are statistical values representing the distance of the cluster mean from the BRASS mean for that PC. Cluster PCs are shown only where the score is ≥0.05 or ≤–0.05. bDMARD, biologic DMARD; csDMARD, conventional synthetic DMARD; DD, disease duration; DMARD, disease-modifying antirheumatic drug; NSAID, nonsteroidal anti-inflammatory drug; TNF, tumor necrosis factor

CONCLUSIONS

- Data-driven cluster analysis of the BRASS registry identified 41 PCs that capture key patient characteristics of RA
- The PCs directed the identification of five distinct patient phenotypes that could be described based on general health, RA control, and RA duration
- Results suggest the clusters represent clinically meaningful phenotypes of RA that may have different trajectories of disease progression and treatment response
- Validation in an independent dataset is ongoing

Table 1. Summary of clusters

Cluster	1 (N=691)	2 (N=280)	3 (N=174)	4 (N=108)	5 (N=190)
Summary descriptors					
General health	High	High	Low–moderate	Low	Low
RA control	Controlled	Controlled	Moderate	Uncontrolled	Uncontrolled
RA duration	Shorter	Longer	Moderate	Shorter	Longer
Illustrative variables					
>1 comorbidity, n (%)[a]	285 (41)	147 (53)	112 (64)	100 (93)	133 (70)
Mean number of comorbidities (SD)[a]	2.74 (1.92)	3.84 (2.17)	5.34 (2.47)	5.45 (2.53)	4.70 (2.24)
CDAI ≤10, n (%)	298 (43)	130 (46)	63 (36)	30 (28)	45 (24)
Mean number of RA-related surgeries (SD)	0.47 (1.10)	0.94 (1.94)	0.62 (1.31)	0.28 (0.76)	1.38 (2.37)
Median total Sharp score (range)	1 (0–230)	1 (0–241)	1 (0–220)	0.5 (0–157)	10.5 (0–270)
Median RA duration, years	5	11	9	8	22

[a]Charlson Comorbidity Index. SD, standard deviation

Reference

1. Iannaccone CK, et al. *Rheumatology* 2011;50:40–6

Acknowledgments

The BRASS registry is supported by Sanofi Genzyme, Regeneron Pharmaceuticals, Inc., Bristol-Myers Squibb, and Crescendo Bioscience. Analyses presented here and medical writing support (Sophie Albon of Adelphi Communications Ltd, Macclesfield, UK) were funded by Sanofi Genzyme (Cambridge, MA, USA) and Regeneron Pharmaceuticals, Inc. (Tarrytown, NY, USA USA) in accordance with Good Publication Practice (GPP3) guidelines. Data in this poster were previously presented as an oral presentation at the European Congress of Rheumatology (EULAR) 2019

Disclosures

- J.R. Curtis has received grants and/or consulting fees from AbbVie, Amgen, Bristol-Myers Squibb, Corrona, Janssen, Lilly, Myriad, Pfizer, Regeneron, Roche, and UCB
- M. Weinblatt has received grants and/or consulting fees and/or has ownership interests in AbbVie, Amgen, Bristol-Myers Squibb, Canfite, Corrona, Crescendo Bioscience, Gilead, GlaxoSmithKline, Horizon, Lilly, Lycera, Novartis, Pfizer, Roche, Samsung, Sanofi Genzyme/Regeneron, Scipher, and SetPoint
- K. Saag has received grants and/or consulting fees from AbbVie, Amgen, Bayer, Gilead, Horizon, Ironwood/AstraZeneca, Kowa, Radius, Roche/Genentech, Sobi, Takeda, and Teijin
- V.P. Bykerk has received grants and/or consulting fees and/or other financial/material support from Amgen, The Cedar Hill Foundation, Gilead, NIH, Pfizer, Sanofi Genzyme/Regeneron, Scipher, and UCB
- C. Charles-Schoeman has received grants and/or consulting fees from AbbVie, Amgen, Bristol-Myers Squibb, Gilead, Octopharma, Pfizer, and Sanofi Genzyme/Regeneron
- S. Fiore and S. Zheng are employees of Sanofi and may hold stock and/or stock options in the company
- G. St John and T. Kimura are employees of Regeneron Pharmaceuticals, Inc. and may hold stock and/or stock options in the company
- C. Bingham has received grants and/or consulting fees from AbbVie, Bristol-Myers Squibb, Genentech/Roche, Lilly, and Sanofi Genzyme/Regeneron
- G. Wright has received consulting and/or speakers' bureau fees from AbbVie, Amgen, Bristol-Myers Squibb, Exagen, Lilly, Myriad, Novartis, Pfizer, Sanofi Genzyme/Regeneron, and UCB
- M. Bergman has received consulting and/or speakers' bureau fees and/or is a shareholder in AbbVie, Bristol-Myers Squibb, Celgene, Genentech, Gilead, GlaxoSmithKline, Horizon, Janssen, Johnson & Johnson (parent of Janssen), Merck, Novartis, Pfizer, Sandoz, and Sanofi Genzyme/Regeneron
- K. Nola has received consulting fees and/or speakers' bureau fees and/or is a shareholder in Coherus, Gilead, Johnson & Johnson, and Sanofi Genzyme/Regeneron
- D.E. Furst has received grants and/or consulting fees from Actelion, Amgen, Bristol-Myers Squibb, Corbus, Galapagos, GlaxoSmithKline, NIH, Novartis, Pfizer, Roche/Genentech, and Sanofi Genzyme/Regeneron
- N. Shadick has received grants from Bristol-Myers Squibb, Crescendo Bioscience, Mallinckrodt, and Sanofi Genzyme/Regeneron