# ORACLE

# Examining factors contributing to age of late-onset autoimmune disease in U.S. adults using hierarchical clustering mechanisms

Authors: Jordyn Homoki, Vicky W. Li, Kyla Finlayson, Lulu K. Lee

Company: Oracle Life Sciences

## Background

The age of AD diagnosis is often associated with the degree of complications from disease, with an earlier age of diagnosis increasing risk of mortality and morbidity and later diagnoses increasing risk of simultaneous development of additional ADs[1].

Identifying potential demographic and health-related characteristics that naturally presented with an early age of disease onset could potentially aid earlier diagnosis, treatment, and ultimately, avoidance of aggressive progression of symptoms.

## Objective

This study examined potential distinct profiles of individuals with autoimmune disease(s) (AD) in a general U.S. adult population using unsupervised clustering.

## Methods

**Data Source**

This study used cross-sectional data from the 2022 and 2023 National Health and Wellness Survey (NHWS). It is an annual, self-administered, internet-based, nationwide survey comprised of ~75,000 US adults (age 18 or older) each year. Participants are recruited through an existing, general-purpose web-based consumer panel. The final sample matches the demographic composition of the US on age, gender, and racial/ethnic groups, based on data from the Current Population Survey of the US Census.

In 2022 and 2023, there were a total of 7,224 adults reported having been diagnosed with an AD condition.

A total of 23 variables were included in analysis, including AD-specific variables, socio-demographic characteristics, health-related characteristics, comorbidities diagnosed by a doctor, ever been diagnosed with COVID-19 or respiratory syncytial virus (RSV), family history of risk factors, as well as mental health and amount of social support received.

**AD-Specific variables**

The survey included year of diagnosis and 14 AD conditions ever been diagnosed by a doctor. AD conditions were examined individually, summed into total number of AD conditions, grouped into systemic and organ-specific conditions, and as early onset and late onset conditions.

Early onset conditions included: Type I Diabetes, Celiac Disease, Sickle Cell Disease, Hemophilia A, Hemophilia B. Late onset conditions included: Psoriasis, Rheumatoid Arthritis, Alopecia Areata, Crohn's Disease, Lupus, Multiple Sclerosis, Vitiligo, Sjogren's Syndrome, Latent Autoimmune Diabetes.

Systemic conditions included: Psoriasis, Rheumatoid Arthritis, Crohn's Disease, Lupus, Sjogren's Syndrome, Celiac Disease, Sickle Cell Disease, Hemophilia A, Hemophilia B. Organ-specific conditions included: Type I Diabetes, Alopecia Areata, Multiple Sclerosis, Vitiligo, Latent Autoimmune Diabetes.

Age of AD onset was defined as the earliest age that the participant was being diagnosed with an AD condition, if there were multiple conditions reported. Participants would be considered as having early onset of AD if they had an early onset condition or if the age at diagnosis was less than 18. Otherwise, they would be considered as having late onset of AD.

**Socio-demographic characteristics**

Analysis included age at the time of the survey, gender, race/ethnicity, education, employment status, total annual household income, marital status, number of adults living the same household, number of children under the age of 18 living in the same household.

**Health-related characteristics**

Analysis included smoking status, body mass index (kg/m²; BMI), and health insurance type.

**Comorbidities**

Analysis included other selective non-AD conditions that participants reported ever having been diagnosed by a doctor, including: hypertension, angina, atherosclerosis, migraines, history of alcoholism, COVID-19, and RSV.

**Mental health and social support**

Mental health was examined using the total scores from the Patient Health Questionnaire-9 (PHQ-9)[2] and the 7-item GAD questionnaire (GAD-7)[3]. The higher the scores, the more symptoms of depression and anxiety the participant presents.

Level of social support available was examined using the total score from the modified Medical Outcomes Study Social Support Survey (mMOS-SS)[4]. The higher the scores, the higher level of social support available to the participant.

**Family history of AD risk factors**

Analysis included a list of 17 conditions regarding the participants' family medical history, including asthma, cancer, cardiovascular and cerebrovascular conditions, mental health conditions, etc.

**Analysis Methods**

All study variables were entered into the clustering analysis. Gower's Distance was used to evaluate similarity between observations. Clusters were aggregated by within-cluster sum of squares. Average silhouette was examined using single, complete, average, and Ward's linkage. Principle component analysis (PCA) was used as a dimensionality reduction technique.

Bivariate comparisons were then used to compare the clusters generated.

## Results

- Age of onset of AD ranged from 17 years to 39 years on average across the clusters (**Table 1**)
- Most participants had one AD condition that is a systemic condition (**Table 1**)
- In general, the top 3 AD conditions were: psoriasis, rheumatoid arthritis, and type 1 diabetes (**Table 1**).
- Ward's linkage resulted in the most optimal clusters. Three clusters were identified (**Tables 1-3**):
  1. Most diagnosed with **systemic AD**, predominantly **female**, largest percentage of **African Americans**, least **educated** and lowest **income**, fewest number of adults and children **living in the same household**, highest **BMI**, and had the most **family history** risk factors
  2. Highest prevalence of **late-onset AD**, oldest average **age of late-onset** AD (41.5 years old), fewest **number of unique ADs**, predominantly **female**, largest percentage of **Asians**, majority **never smoked**, presented symptoms of **depression and anxiety**, had the highest level of **social support**, and generally had the fewest **comorbidities**
  3. Highest prevalence of **early-onset AD**, youngest average **age of late-onset** AD (33.2 years old), had the most **organ-specific ADs**, had the greatest number of **unique ADs**, predominantly **male**, contained the largest proportion of **Hispanics**, most educated and highest **income**, most **actively employed**, contained the most **current smokers**, lowest **BMI**, presented with the highest number of symptoms of **depression and anxiety**, and had the fewest number of **family history risk factors**

### Table 1. AD-specific variables across clusters of AD participants.

| | Cluster 1 | Cluster 2 | Cluster 3 | p-value |
|---|---|---|---|---|
| Total sample size (N=) | 1,886 | 3,885 | 1,453 | |
| **AD-specific variables** | | | | |
| First AD onset time | | | | <0.001 |
| Early onset | 54 (2.86%) | 85 (2.19%) | 955 (65.73%) | |
| Late onset | 1,832 (97.14%) | 3,800 (97.81%) | 498 (34.27%) | |
| Age of first AD onset | | | | <0.001 |
| Mean (SD) | 39.34 (15.10) | 40.83 (14.69) | 17.28 (12.95) | |
| Age of first late onset AD | | | | <0.001 |
| Mean (SD) | 40.18 (14.42) | 41.48 (14.15) | 33.15 (10.86) | |
| Missing | 44 | 61 | 802 | |
| First AD Type | | | | <0.001 |
| Systemic | 1,513 (80.22%) | 2,967 (76.37%) | 846 (58.22%) | |
| Organ-specific | 373 (19.78%) | 918 (23.63%) | 607 (41.78%) | |
| Number of ADs | | | | <0.001 |
| Mean (SD) | 1.26 (0.61) | 1.17 (0.46) | 1.37 (1.07) | |
| Number of ADs (categorical) | | | | <0.001 |
| 1 | 1,515 (80.33%) | 3,322 (85.51%) | 1,118 (76.94%) | |
| 2 | 289 (15.32%) | 480 (12.36%) | 247 (17.00%) | |
| 3+ | 82 (4.35%) | 83 (2.14%) | 88 (6.06%) | |
| **AD conditions** | | | | |
| Psoriasis | 593 (31.44%) | 1,445 (37.19%) | 439 (30.21%) | <0.001 |
| Rheumatoid Arthritis | 753 (39.93%) | 1,182 (30.42%) | 267 (18.38%) | <0.001 |
| Type 1 Diabetes | 151 (8.01%) | 351 (9.03%) | 447 (30.76%) | <0.001 |
| Alopecia Areata | 163 (8.64%) | 427 (10.99%) | 289 (19.89%) | <0.001 |
| Crohn's Disease | 160 (8.48%) | 292 (7.52%) | 152 (10.46%) | 0.002 |
| Lupus | 162 (8.59%) | 239 (6.15%) | 88 (6.06%) | 0.001 |
| Multiple Sclerosis | 137 (7.26%) | 215 (5.53%) | 63 (4.34%) | 0.001 |
| Vitiligo | 83 (4.40%) | 150 (3.86%) | 48 (3.30%) | 0.264 |
| Sjogren's Syndrome | 70 (3.71%) | 122 (3.14%) | 32 (2.20%) | 0.044 |
| Celiac Disease | 53 (2.81%) | 84 (2.16%) | 80 (5.51%) | <0.001 |
| Latent Autoimmune Diabetes | 21 (1.11%) | 32 (0.82%) | 32 (2.20%) | <0.001 |
| Sickle Cell Disease | 16 (0.85%) | 4 (0.10%) | 15 (1.03%) | <0.001 |
| Hemophilia A | 5 (0.27%) | 5 (0.13%) | 23 (1.58%) | <0.001 |
| Hemophilia B | 5 (0.27%) | 4 (0.10%) | 16 (1.10%) | <0.001 |

### Table 2. Socio-demographics and health-related characteristics across clusters of AD participants.

| | Cluster 1 | Cluster 2 | Cluster 3 | p-value |
|---|---|---|---|---|
| Total sample size (N=) | 1,886 | 3,885 | 1,453 | |
| **Sociodemographic characteristics** | | | | |
| Age at time of survey | | | | <0.001 |
| Mean (SD) | 55.10 (15.50) | 56.87 (14.89) | 40.03 (13.92) | |
| Gender | | | | <0.001 |
| Male | 619 (32.82%) | 1,615 (41.57%) | 806 (55.47%) | |
| Female | 1,267 (67.18%) | 2,270 (58.43%) | 647 (44.53%) | |
| Ethnicity | | | | <0.001 |
| African American | 342 (18.15%) | 460 (11.84%) | 125 (8.60%) | |
| Asian | 28 (1.48%) | 173 (4.45%) | 37 (2.55%) | |
| Hispanic | 200 (10.60%) | 420 (10.81%) | 237 (16.31%) | |
| White | 1,160 (61.51%) | 2,630 (67.70%) | 989 (68.07%) | |
| Mixed | 95 (5.04%) | 128 (3.29%) | 35 (2.41%) | |
| Other | 61 (3.23%) | 74 (1.90%) | 30 (2.06%) | |
| Education | | | | <0.001 |
| Less than college | 514 (27.25%) | 511 (13.15%) | 232 (15.97%) | |
| Some college | 911 (48.30%) | 1,216 (31.30%) | 352 (24.23%) | |
| Bachelor's degree and above | 457 (24.23%) | 2,156 (55.50%) | 862 (59.33%) | |
| Declined to answer | 4 (0.21%) | 2 (0.05%) | 7 (0.48%) | |
| Employment Status | | | | <0.001 |
| Employed full-time | 297 (15.75%) | 1,536 (39.54%) | 878 (60.43%) | |
| Self-employed | 152 (8.06%) | 245 (6.31%) | 75 (5.16%) | |
| Employed part-time | 139 (7.37%) | 256 (6.59%) | 140 (9.64%) | |
| Homemaker | 91 (4.83%) | 175 (4.50%) | 56 (3.85%) | |
| Retired | 641 (33.99%) | 1,277 (31.58%) | 87 (5.99%) | |
| Student | 40 (2.12%) | 32 (0.82%) | 55 (3.79%) | |
| Disability | 336 (17.82%) | 197 (5.07%) | 69 (4.75%) | |
| Declined to answer | 190 (10.07%) | 167 (4.30%) | 93 (6.40%) | |
| Number of People in Household | | | | <0.001 |
| Mean (SD) | 1.82 (0.99) | 2.16 (0.82) | 2.30 (0.99) | |
| Number of Children in Household | | | | <0.001 |
| Mean (SD) | 0.38 (0.86) | 0.44 (0.90) | 1.05 (1.14) | |
| Marital Status | | | | <0.001 |
| Married or living with partner | 384 (20.36%) | 2,899 (74.62%) | 937 (64.49%) | |
| Single, never married | 646 (34.25%) | 470 (12.10%) | 393 (27.05%) | |
| Divorced, separated, or widowed | 848 (44.96%) | 508 (13.08%) | 111 (7.64%) | |
| Declined to answer | 8 (0.42%) | 8 (0.21%) | 12 (0.83%) | |
| Household Income | | | | <0.001 |
| <$50000 | 1,432 (75.93%) | 874 (22.50%) | 394 (27.12%) | |
| $50000 to $100000 | 257 (13.63%) | 1,526 (39.28%) | 410 (28.22%) | |
| $100000 to $150000 | 90 (4.78%) | 741 (19.07%) | 374 (25.74%) | |
| >$150000 | 47 (2.39%) | 597 (15.37%) | 225 (15.49%) | |
| Declined to answer | 60 (3.18%) | 147 (3.78%) | 50 (3.44%) | |
| **Health-related characteristics** | | | | |
| Health insurance type | | | | <0.001 |
| Private Insurance | 258 (13.68%) | 2,187 (56.29%) | 902 (62.08%) | |
| Public Insurance | 1,477 (78.31%) | 1,533 (39.46%) | 373 (25.67%) | |
| Unknown | 151 (8.01%) | 165 (4.25%) | 178 (12.25%) | |
| Smoking status | | | | <0.001 |
| Never smoked | 772 (40.93%) | 2,250 (57.92%) | 622 (42.81%) | |
| Current smoker | 491 (26.03%) | 522 (13.44%) | 636 (43.77%) | |
| Past smoker | 623 (33.03%) | 1,113 (28.65%) | 195 (13.42%) | |
| BMI | | | | <0.001 |
| Mean (SD) | 30.96 (8.52) | 29.69 (7.37) | 26.80 (7.95) | |
| Missing | 43 | 116 | 162 | |

### Table 3. Comorbidities, family history, mental health, and social support across clusters of AD participants.

| | Cluster 1 | Cluster 2 | Cluster 3 | p-value |
|---|---|---|---|---|
| Total sample size (N=) | 1,886 | 3,885 | 1,453 | |
| **Comorbidities** | | | | |
| High Blood Pressure/Hypertension | 1,063 (56.36%) | 1,936 (49.83%) | 357 (24.57%) | <0.001 |
| Angina/Angina Pectoris | 96 (5.09%) | 166 (4.27%) | 109 (7.50%) | <0.001 |
| Atherosclerosis | 70 (3.71%) | 148 (3.81%) | 118 (8.12%) | <0.001 |
| Migraines | 728 (38.60%) | 756 (19.46%) | 510 (35.10%) | <0.001 |
| Alcoholism | 244 (12.94%) | 171 (4.40%) | 247 (17.00%) | <0.001 |
| COVID-19 Ever | 469 (24.87%) | 1,489 (38.33%) | 599 (41.23%) | <0.001 |
| RSV Ever | 86 (4.56%) | 114 (2.93%) | 272 (18.72%) | <0.001 |
| **Mental health** | | | | |
| PHQ-9 Score | | | | <0.001 |
| Mean (SD) | 8.68 (7.43) | 5.40 (5.86) | 10.22 (7.67) | |
| GAD-7 Score | | | | <0.001 |
| Mean (SD) | 6.57 (6.32) | 4.05 (5.03) | 8.16 (6.25) | |
| **Social support** | | | | |
| mMOSS Score | | | | <0.001 |
| Mean (SD) | 53.50 (30.71) | 70.12 (27.23) | 62.61 (27.48) | |
| **Family history of AD risk factors** | | | | |
| Alcoholism | 704 (37.33%) | 1,075 (27.67%) | 438 (30.14%) | <0.001 |
| Asthma | 519 (27.52%) | 772 (19.87%) | 429 (29.53%) | <0.001 |
| Arthritis | 1,123 (59.54%) | 2,000 (51.48%) | 587 (40.40%) | <0.001 |
| Bipolar/Manic Disorder | 391 (20.73%) | 426 (10.97%) | 266 (18.31%) | <0.001 |
| High Blood Pressure/Hypertension | 1,296 (68.72%) | 2,432 (62.60%) | 601 (41.36%) | <0.001 |
| Cancer | 935 (49.58%) | 1,872 (48.19%) | 491 (33.79%) | <0.001 |
| High Cholesterol | 915 (48.52%) | 1,835 (47.23%) | 490 (33.72%) | <0.001 |
| Diabetes | 1,064 (56.42%) | 1,873 (48.21%) | 711 (48.93%) | <0.001 |
| Depression | 728 (38.60%) | 1,017 (26.18%) | 464 (31.93%) | <0.001 |
| Heart Attack | 791 (41.94%) | 1,398 (35.98%) | 381 (26.22%) | <0.001 |
| Heart Disease | 804 (42.63%) | 1,580 (40.67%) | 398 (27.39%) | <0.001 |
| Migraines | 492 (26.09%) | 649 (16.71%) | 333 (22.92%) | <0.001 |
| Multiple Sclerosis | 119 (6.31%) | 143 (3.68%) | 107 (7.36%) | <0.001 |
| Obesity | 687 (36.43%) | 1,146 (29.50%) | 332 (22.85%) | <0.001 |
| Osteopenia | 85 (4.51%) | 185 (4.76%) | 81 (5.57%) | 0.334 |
| Osteoporosis | 355 (18.82%) | 616 (15.86%) | 177 (12.18%) | <0.001 |
| Stroke/TIA | 564 (29.90%) | 1,001 (25.77%) | 253 (17.41%) | <0.001 |

## Conclusion

In a broadly representative survey of US adults, hierarchical clustering identified potential distinct profiles of individuals, specifically with regards to the characteristics that occurred with early- and late-onset of AD, to be considered in future clinical research for AD prevention and treatment development.

## References

1. Amador-Patarroyo, M. J., Rodriguez-Rodriguez, A., & Montoya-Ortiz, G. (2012). How does age at onset influence the outcome of autoimmune diseases?. Autoimmune diseases, 2012, 251730. https://doi.org/10.1155/2012/251730
2. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001 Sep;16(9):606-13. doi: 10.1046/j.1525-1497.2001.016009606.x. PMID: 11556941; PMCID: PMC1495268
3. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-1097.
4. Moser, A., Stuck, A. E., Silliman, R. A., Ganz, P. A., & Clough-Gorr, K. M. (2012). The eight-item modified Medical Outcomes Study Social Support Survey: psychometric evaluation showed excellent performance. Journal of clinical epidemiology, 65(10), 1107–1116. https://doi.org/10.1016/j.jclinepi.2012.04.007.

MSR156