

---

# Identification of Clinical Response Patterns Through Application of Unsupervised Machine Learning on Clinical Trial Time Series Data

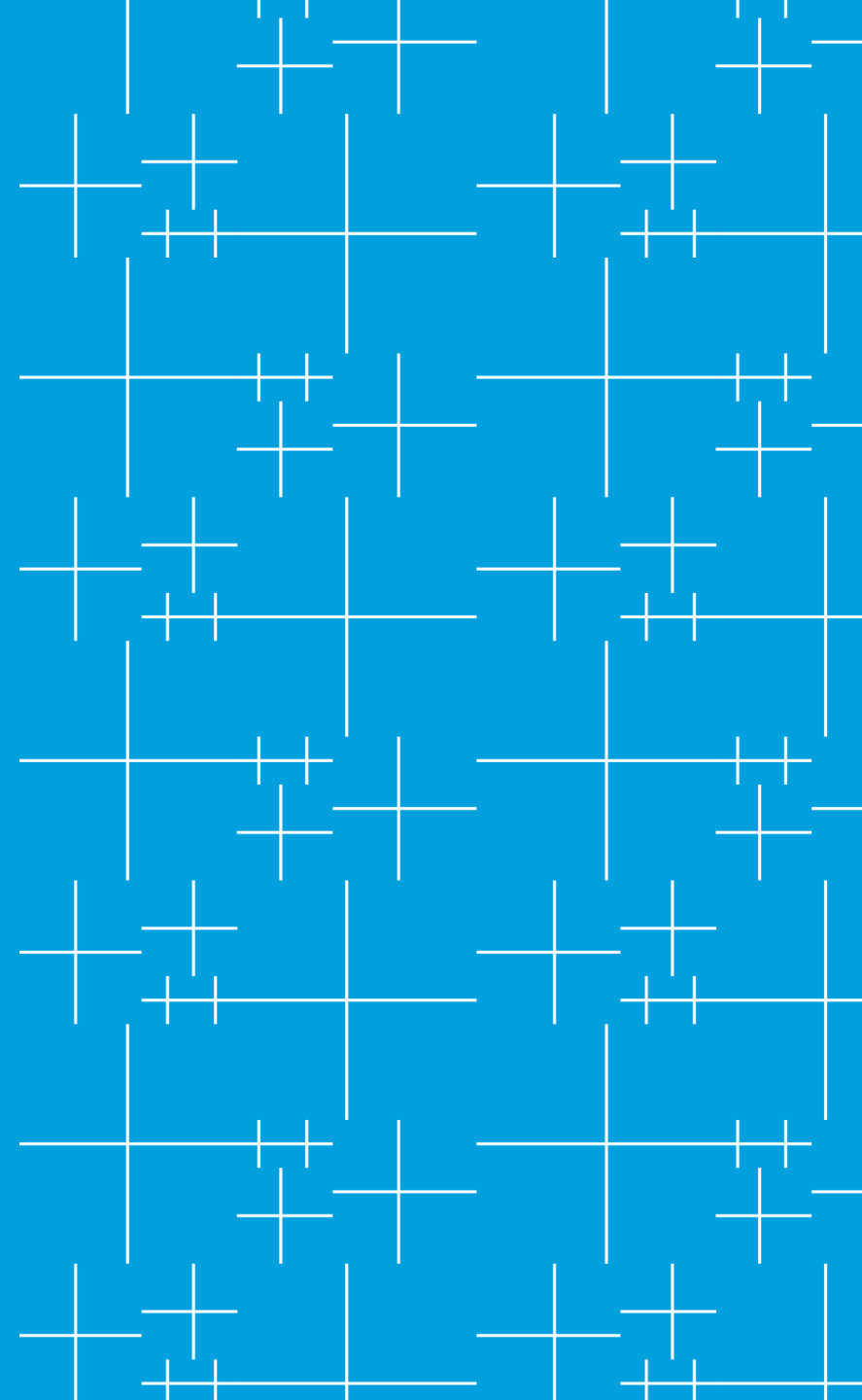
**Presenter: Bethany Hyde, MBS**

Senior Data Scientist, Janssen Commercial Data Sciences

Project Team: Maria Jazra, Sven Wegner, Bulent Ozturk, Jenna Eun, and Jeffrey Headd



**Janssen**  
**Commercial Data Sciences**



---

## Context and Motivation

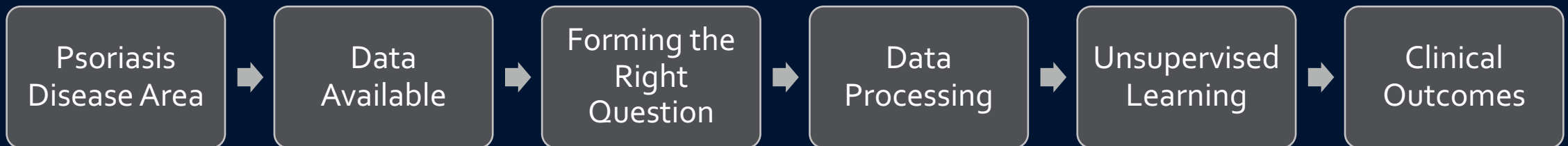
---

## Technical Deployment

---

## Results

---



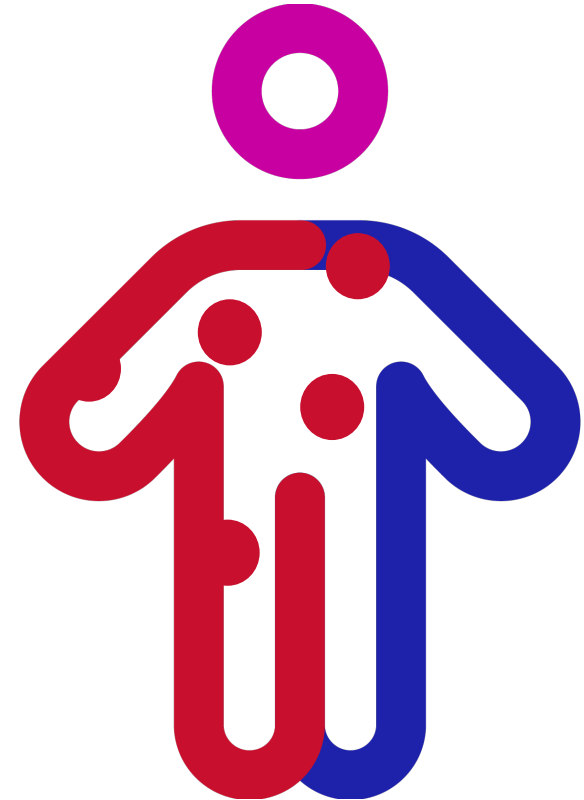
---

# Psoriasis Disease Area and Background

- 125 Million People Worldwide have Psoriasis
- Biologic treatment options for patients with moderate to severe plaque psoriasis
- Patient response to treatment measured by Psoriasis Area and Severity Index (PASI)

## What is PASI 90?

**PASI 90 response is defined** as **90%** improvement or more from baseline on **PASI** score.

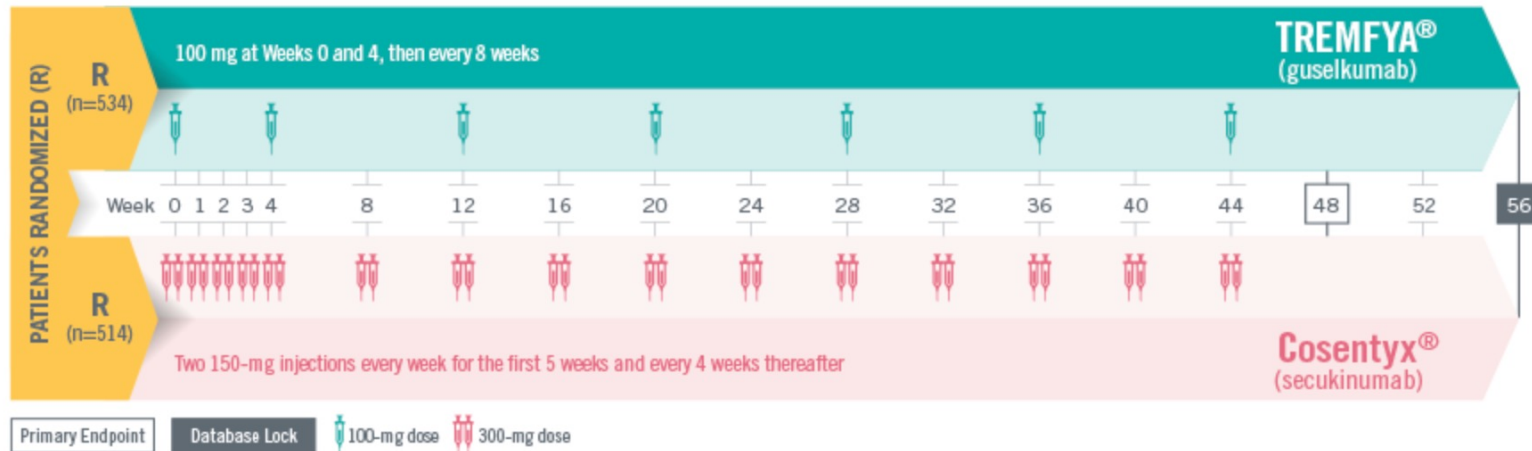


\*<https://www.healthline.com/health/psoriasis/facts-statistics-infographic#Prevalence>

# Available Data

Head-to-head Post Launch Trial - First comparator study of an IL-23 inhibitor, guselkumab (GUS) vs IL-17 inhibitor, secukinumab (SEC)

## ECLIPSE: PHASE 3 DOUBLE-BLIND TRIAL (N=1048)<sup>1,2</sup>



Cosentyx® is a registered trademark of Novartis AG.

## Data Captured

- Patient Demographics
- Baseline Measurements
- Patient Response over Time

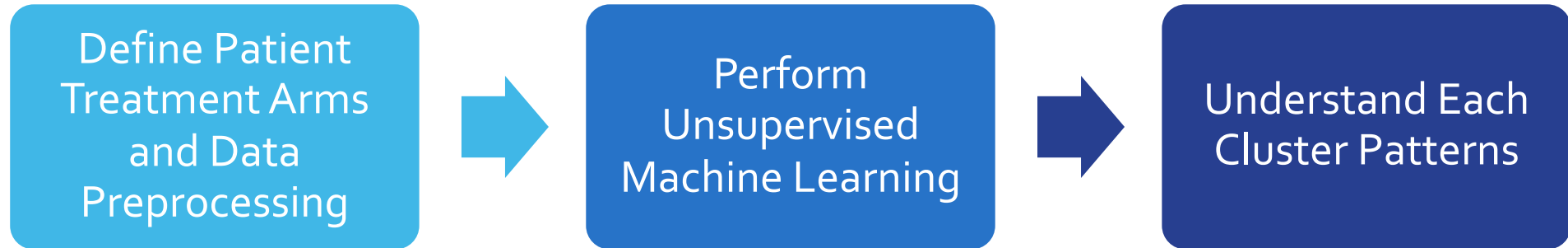
---

# Forming the Right Question

- 1. Identify different response patterns from patients who were treated with either guselkumab and secukinumab.**
- 2. Analyze patient profiles present in each PASI pattern cluster in order to better understand potential qualitative relationships between clinical characteristics.**

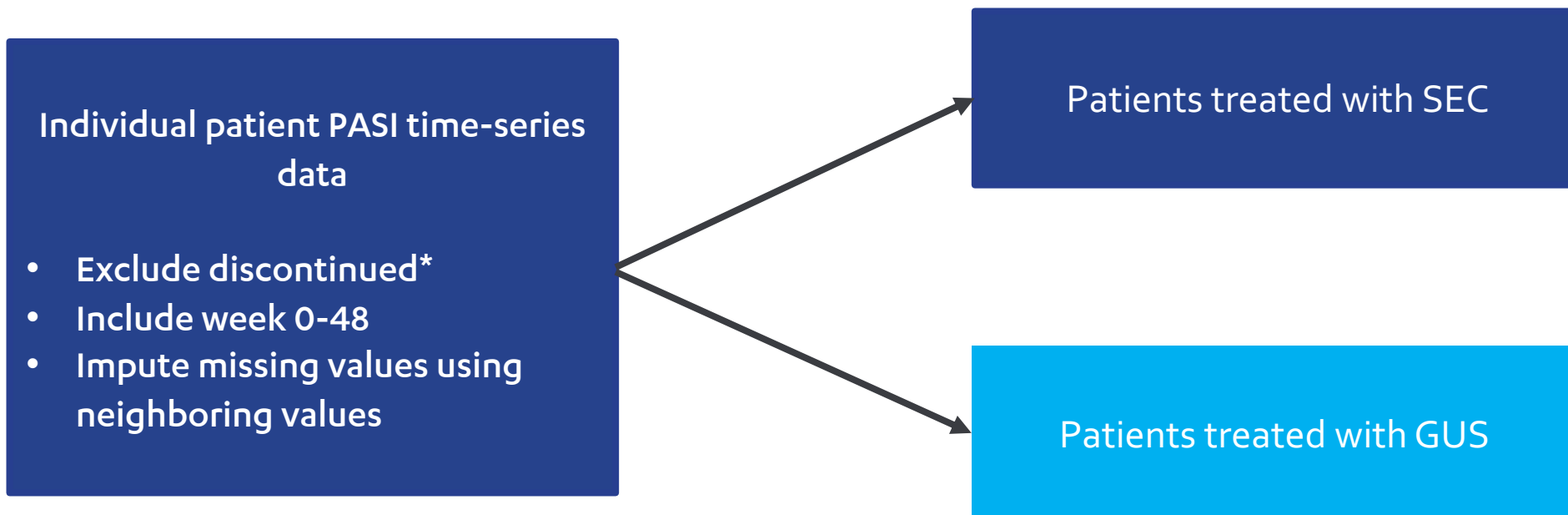
---

# Technical Methodology Overview



# Set-up of the Unsupervised ML analysis

## Define Patient Treatment Arms and Data Preprocessing



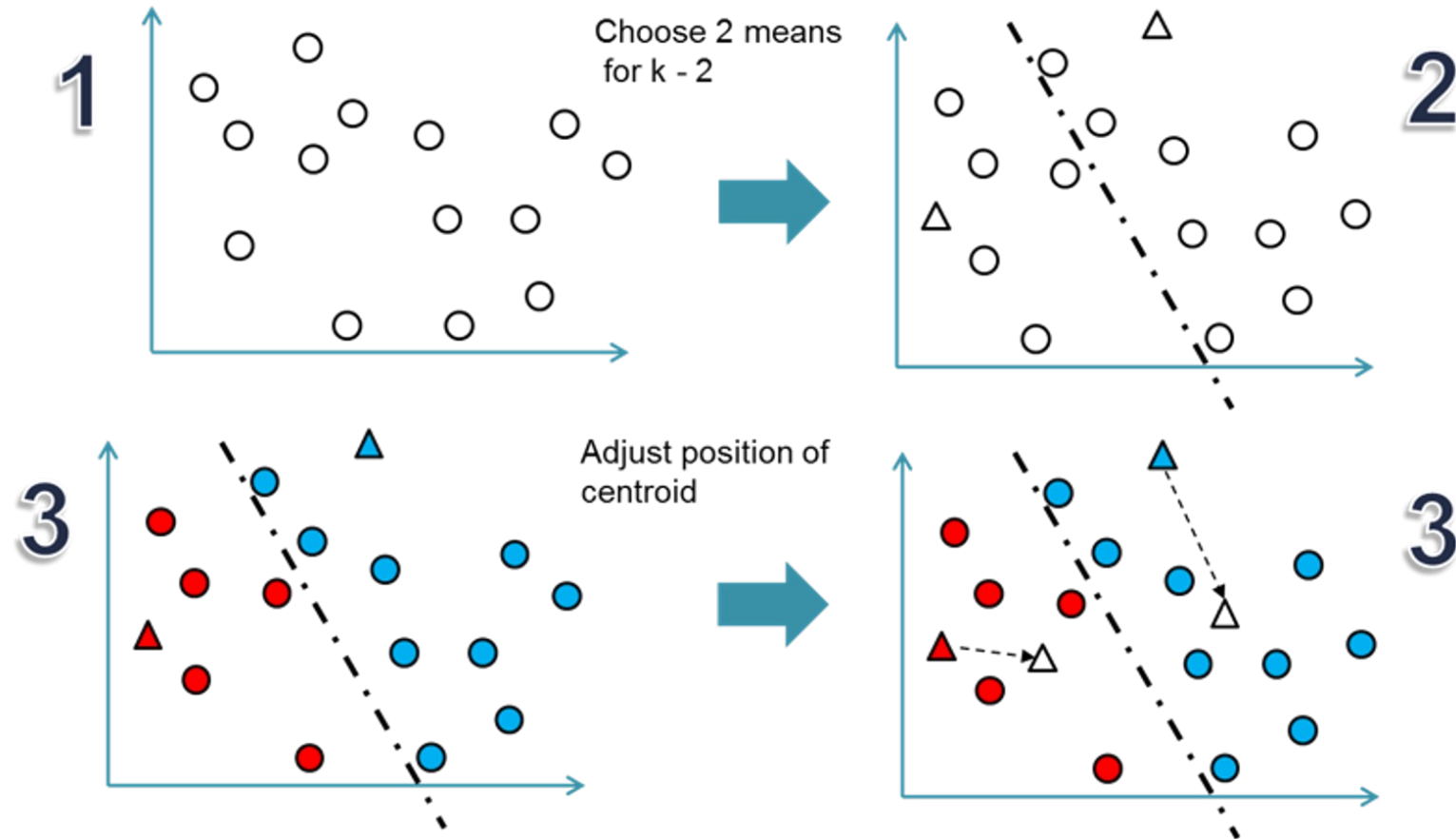
	#
Status = Discontinued	72
Total Patients	973

Treatment	#
Secukinumab	466
Guselkumab	507

\*Discontinued patients have too many missing data points, prohibiting reliable cluster analysis of PASI trends over time.

# Unsupervised Machine Learning

## How Clustering Algorithms Work in General



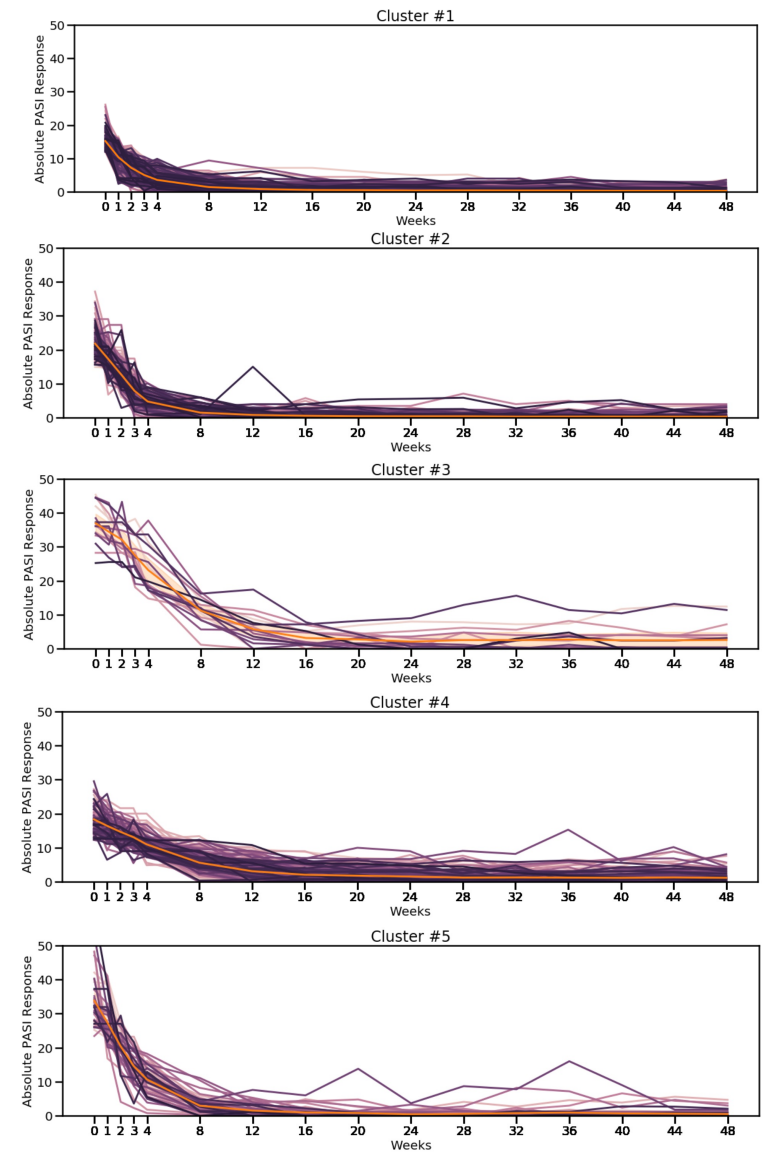
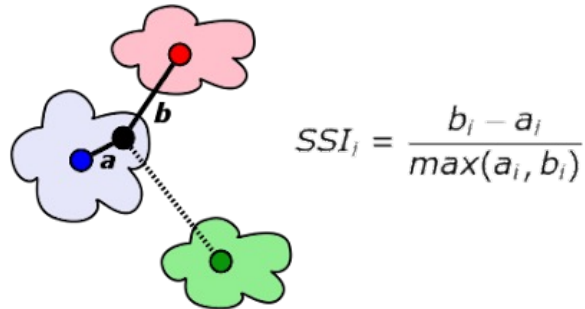


# Unsupervised Machine Learning Approach

## K-means time-series clustering algorithm

The number of clusters was determined by calculating the mean silhouette score

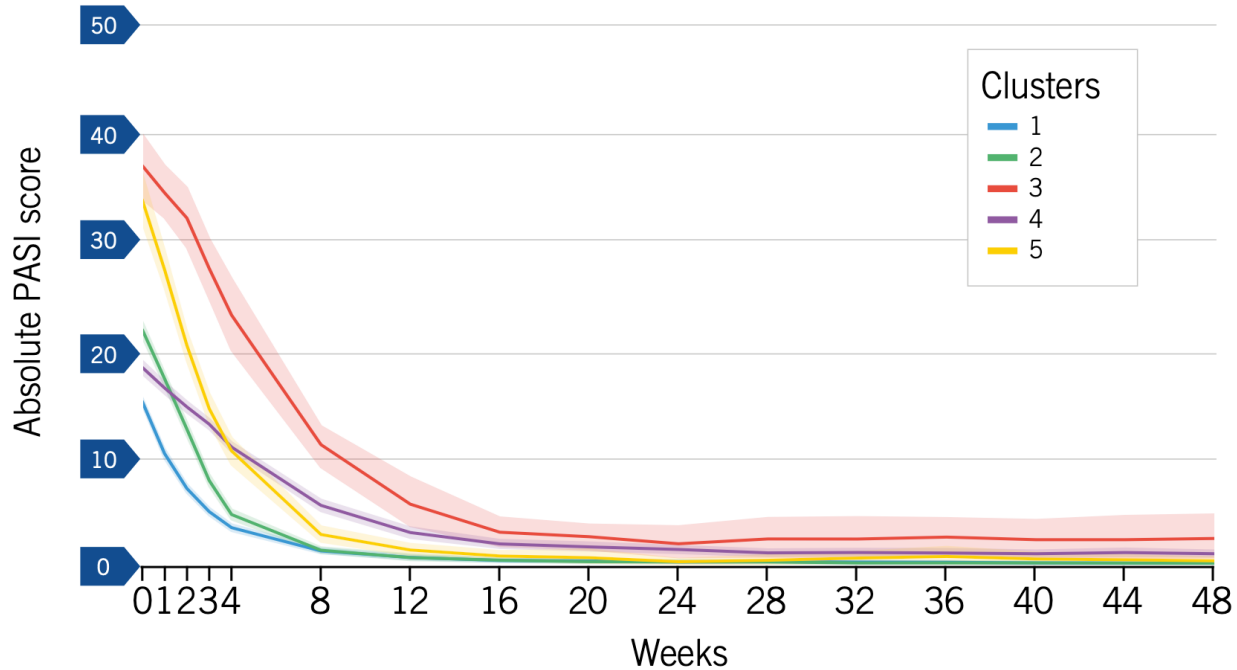
- Maximize Simplified Silhouette Index (SSI) while maintaining granularity of response patterns



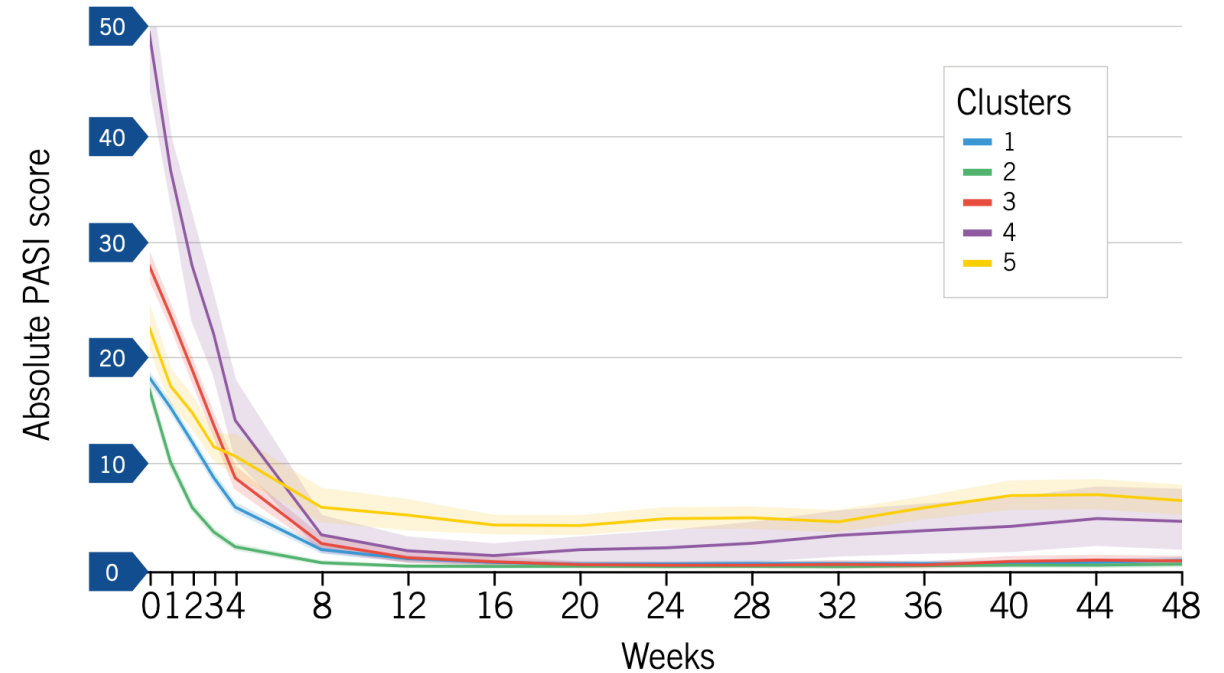
# Understand Profile of Patients in Each Cluster

## Cluster Average PASI Response Comparison

### Guselkumab



### Secukinumab

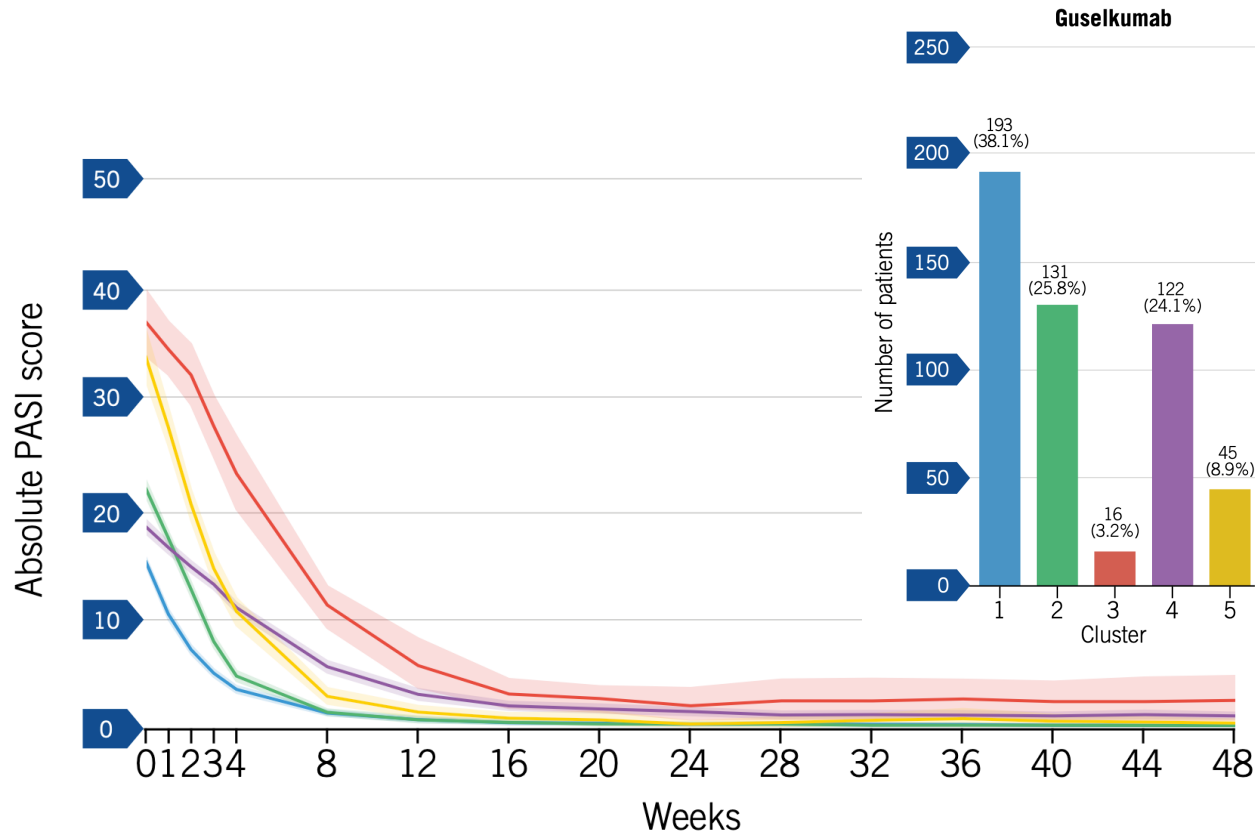


Shaded areas represent the 95% confidence interval of the mean PASI response.

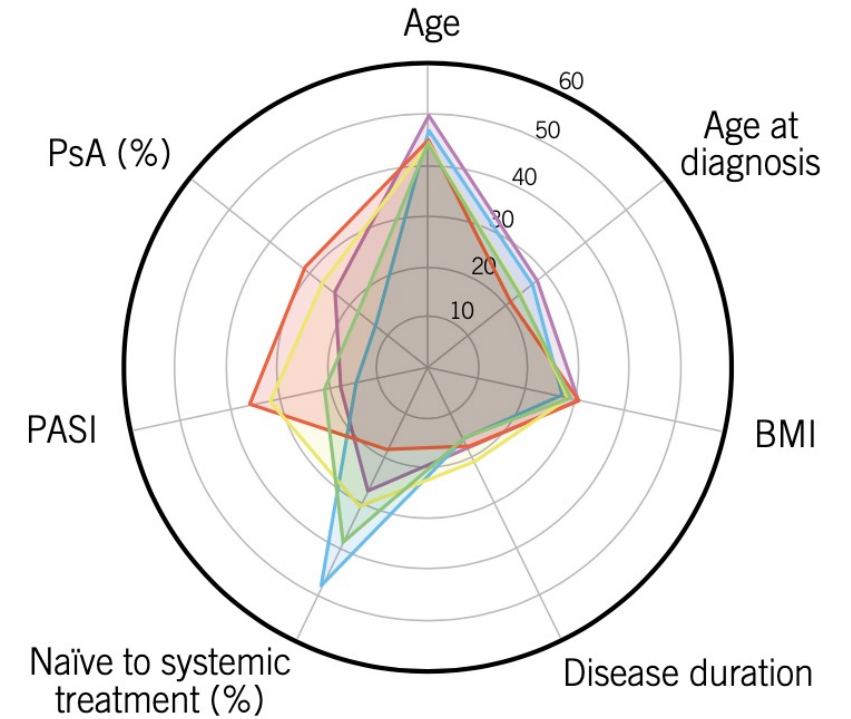
# Guselkumab

## Understanding Cluster Patterns

Mean PASI Response for Each Cluster



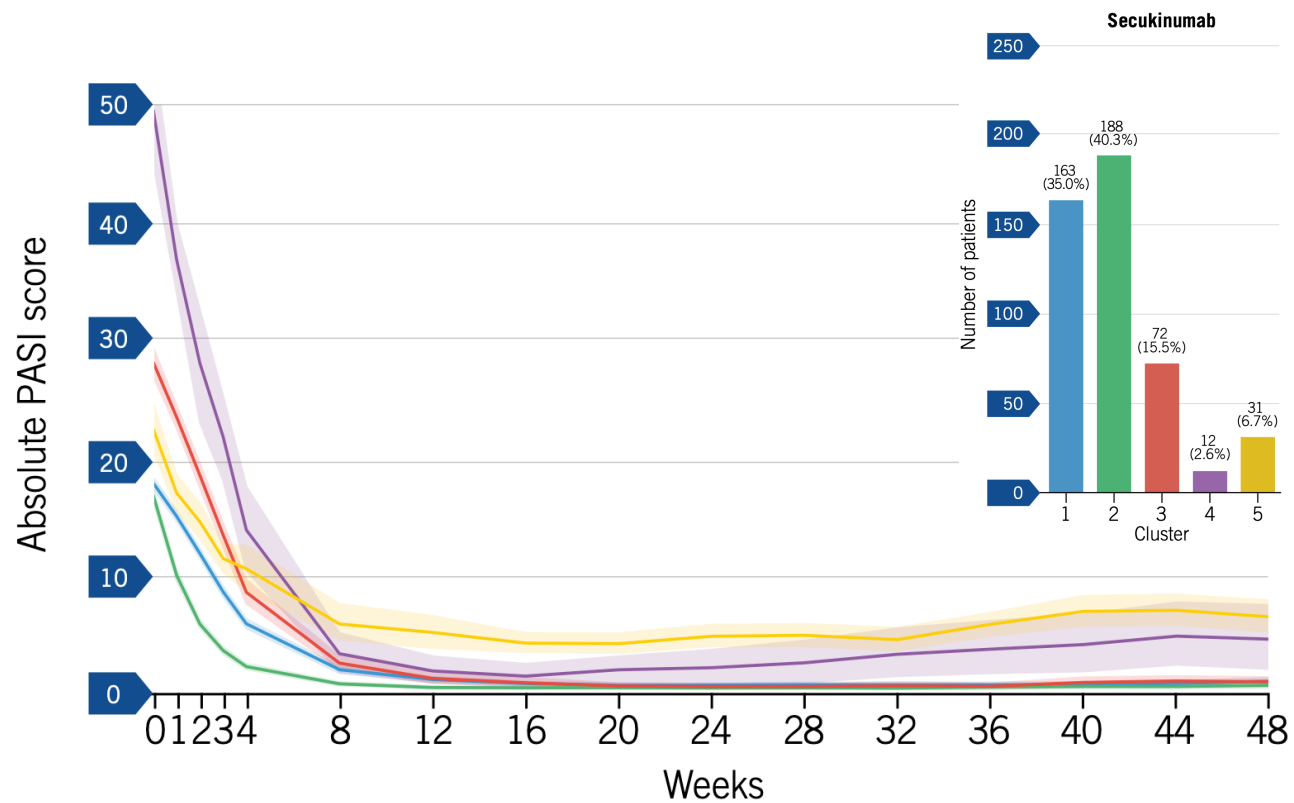
Key Baseline and Clinical Characteristics in Each Cluster



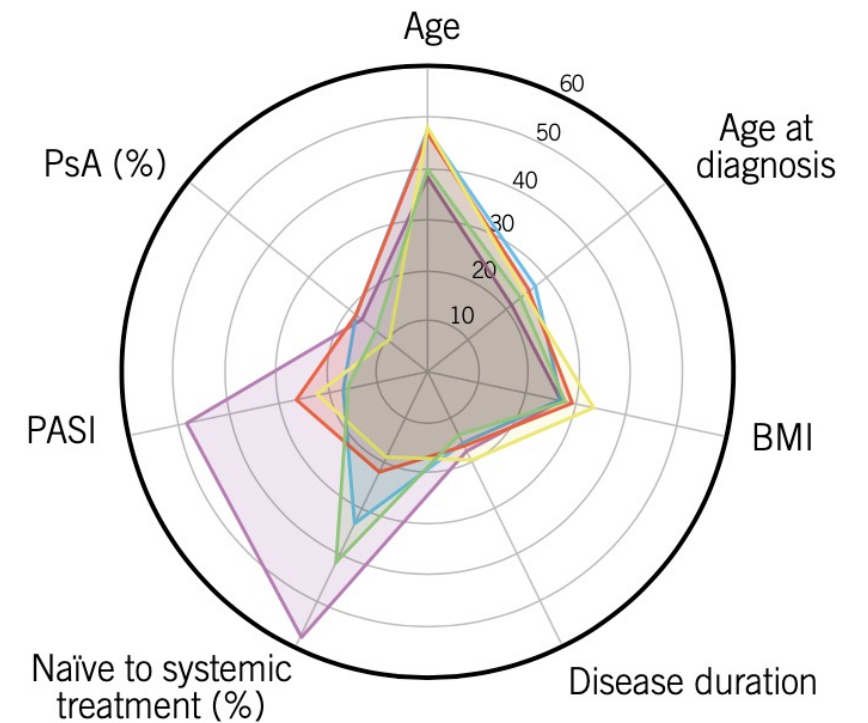
# Secukinumab

## Understanding Cluster Patterns

Mean PASI Response for Each Cluster



Key Baseline and Clinical Characteristics in Each Cluster



Shaded areas represent the 95% confidence interval of the mean PASI response.

---

# Clinical Outcomes

What can the patient profiles by symptom response pattern tell us?

## Guselkimab

↓ % Naïve to Systemic Treatment

↑ % Psoriatic Arthritis

✓ Prior Biologic Treatment

Responding later with higher  
variability over time

## Secukinumab

↑ % Naïve to Systemic Treatment

↓ % Psoriatic Arthritis

✗ Prior Biologic Treatment

Higher rate of losing response after  
initially a good response

---

## Conclusion

Clinical trial data used for patient profiles of symptom response

### **Clinical Trial Time-Series Data**

- Patient Symptom response over time

### **Unsupervised Machine Learning**

- Cluster Optimization using Silhouette Score
- K-Means Time Series Clustering Algorithm

### **Five distinct PASI response patterns were identified**

- Independent of and unbiased by prespecified assumptions

**Patient profiles provide insight into patient symptom response patterns**

---

# References

1. Reich et al. Lancet 2019;394:831–839.
2. Romain, T. et al. (2020). Tslern, A Machine Learning Toolkit for Time Series Data. *Journal of Machine Learning Research*. <http://jmlr.org/papers/v21/20-091.html>
3. Brownlee, J. (2020). 10 Clustering Algorithms With Python. *Machine Learning Mastery*. <https://machinelearningmastery.com/clustering-algorithms-with-python/>
4. Bhardwaj, A. (2020). Silhouette Coefficient: Validating clustering techniques. *Medium*. <https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c>
5. Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. *Pediatrics*, 133(1), e54–e63. <https://doi.org/10.1542/peds.2013-0819>
6. Newcomer SR, Steiner JF, Bayliss EA. (2011) Identifying subgroups of complex patients with cluster analysis. *The American Journal of Managed Care*.
7. Lefèvre T, Rondet C, Parizot I, Chauvin P. (2014). Applying Multivariate Clustering Techniques to Health Data: The 4 Types of Healthcare Utilization in the Paris Metropolitan Area. *PLoS ONE* 9(12): e115064. <https://doi.org/10.1371/journal.pone.0115064>



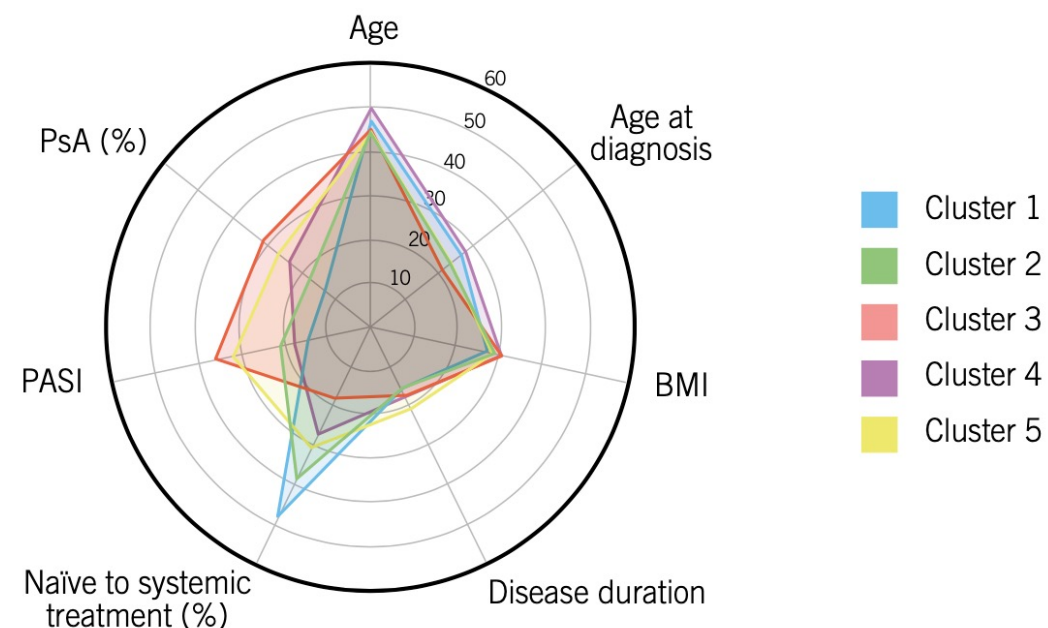
Janssen  
Commercial Data Sciences



# Guselkumab

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Total
Patients, n (%)	193 (38.1)	131 (25.8)	16 (3.2)	122 (24.1)	45 (8.9)	507
Age, years	47.0	45.0	45.0	49.5	44.0	47.0
Age at diagnosis, years	26.0	23.0	20.5	27.0	21.0	25.0
BMI, kg/m <sup>2</sup>	27.0	29.0	30.5	30.0	28.0	28.4
Disease duration, years	16.0	16.0	17.5	18.0	21.0	17.0
Naïve to systemic treatment, %	48.7	38.9	18.7	27.8	31.1	38.7
PASI at baseline	14.8	21.0	36.2	18.0	32.2	18.0
Self-reported PsA prevalence, %	13.4	17.5	31.2	23.7	26.6	18.7
BSA affected, %	15.0	24.0	43.0	20.0	45.0	20.0
Prior biologics	0	0	1	0	0	0

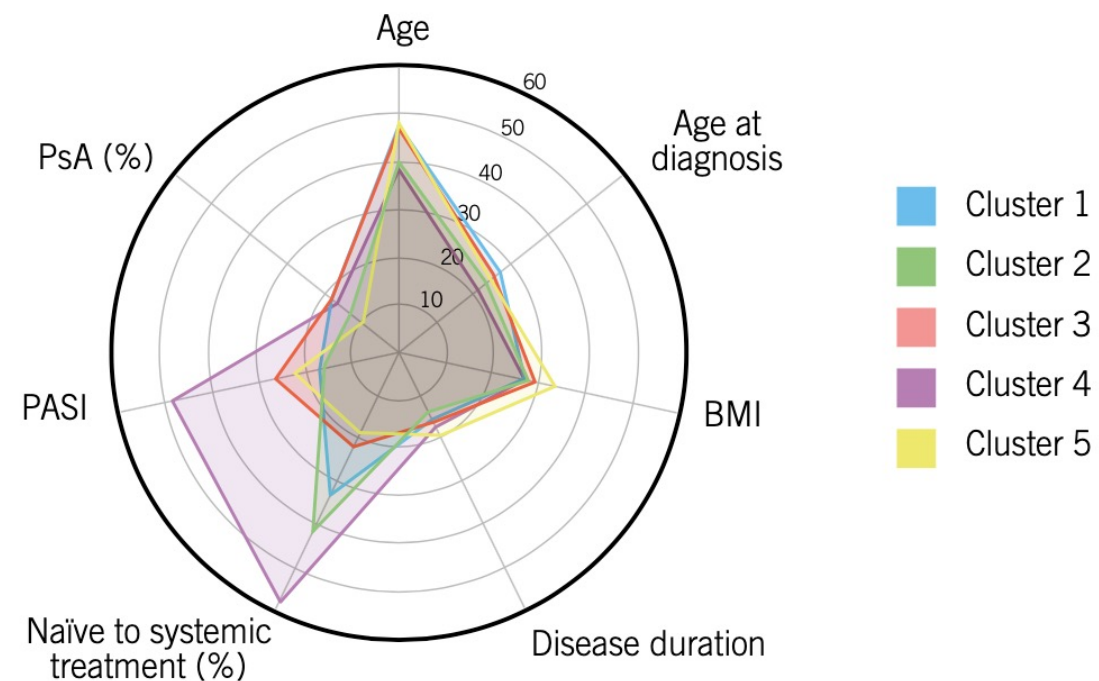
Data are medians unless otherwise indicated.



# Secukinumab

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Total
Patients, n (%)	163 (35.0)	188 (40.3)	72 (15.5)	12 (2.6)	31 (6.7)	466
Age, years	46.0	40.0	46.5	39.0	49.0	43.0
Age at diagnosis, years	27.0	24.0	25.5	21.0	24.0	25.0
BMI, kg/m <sup>2</sup>	29.0	28.0	29.0	27.5	34.0	29.5
Disease duration, years	15.0	14.2	16.9	17.5	21.0	15.0
Naïve to systemic treatment, %	34.9	42.5	22.2	58.3	19.3	35.6
PASI at baseline	17.4	16.2	27.1	48.4	22.4	18.0
Self-reported PsA prevalence, %	17.2	12.7	18.1	16.6	9.6	15.0
BSA affected, %	19.0	17.0	36.0	73.0	29.0	20.0
Prior biologics	0	0	0	0	1	0

Data are medians unless otherwise indicated.





**Janssen**

Commercial Data Sciences