# Representativeness of a U.S. Biomarker Discovery Cohort Compared to Real-World Data from Enrolling U.S. Rheumatology Practices

## Application of Target Trial Emulation Approaches to Augment Longitudinal Data and Construct External Synthetic Comparator Cohorts

Jason Carlson[1], Shanette Daigle[2], Yujie Su[2], Stephanie S. Ford[3], Carleen Tabari[1], Jingyi Zhang[3], Fenglong Xie[2], Diana Abdueva[1], Amy Mudano[2], Angela Estrada Diaz[3], Jeffrey R. Curtis[2,3]

[1]Aqtual, Hayward, CA; [2]Foundation for Advancing Science, Technology, Education and Research, Hoover, AL; [3]University of Alabama at Birmingham, Birmingham, AL

## BACKGROUND

- Participants in prospective RA cohorts may differ from patients treated in routine practice, limiting their use as external synthetic comparators in real-world evidence studies.
- **PRIMA-102** is a U.S. prospective observational study collecting clinical and biomarker data to train and validate a cell-free DNA classifier predicting response to a new tumor necrosis factor inhibitor (TNFi), abatacept, and Janus kinase inhibitor (JAKi).
- PRIMA-102 enrolled RA patients at 52 U.S. sites at the time they were initiating a new biologic therapy or JAKi, collected biospecimens and clinical data, and repeated clinical assessment and biomarker collection at week 12 to gauge treatment response.

## OBJECTIVES

- Assess the representativeness of PRIMA-102 participants compared to RA patients treated by the same rheumatologists but not enrolled in PRIMA-102.
- Link participants to real-world data (RWD) using privacy-preserving tokens to extend follow-up and build external comparator cohorts, and compare characteristics of patients linked vs. not linked.

## METHODS

- Using the initial elements of a target trial emulation approach, we mirrored PRIMA-102 criteria: adults with RA initiating a new b/tsDMARD with prior cs/b/tsDMARD exposure; index = first new b/tsDMARD, June 2023–August 2025; baseline = the 12 months before index.
- **RWD sources (2):** commercial claims data and EHR data (a practice-based EHR network and a national EHR-based RA registry), restricted to the practices of PRIMA-102 enrolling physicians.
- **Linkage:** 1,131 RA patients at the 20 participating sites were compared to the two RWD sources using tokenized PHI (Figure 1), plus conditional matching on site, birth year, sex, RA diagnosis, prior b/tsDMARD, and an office visit within ±3 days of consent.
- **Validation:** claims matches required an RA diagnosis in claims; EHR matches had to be unique (1:1), with ties broken by race/ethnicity and diagnosis and medication history; 61 EHR network token matches were manually reviewed against source records; RWD index therapy was compared with study-reported therapy (Table 2).
- **Comparisons:** standardized mean differences (SMDs; ≥0.1 meaningful) for PRIMA-102 vs. each RWD cohort (Table 1) and for linked vs. non-linked RA patients at the same practices.

**Figure 1: Privacy-Preserving Tokenization**

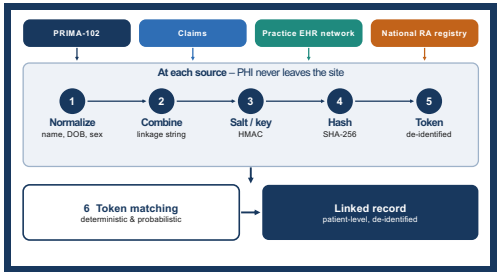



**Table 1: Baseline Characteristics, PRIMA-102 vs. RWD Cohorts at the Same Practices**

| Characteristic | PRIMA-102 (n=803) | Practice EHR network (n=2,221) | Claims (n=2,939) | SMD PRIMA vs. EHR | SMD PRIMA vs. claims |
|---|---|---|---|---|---|
| Age, years, mean (SD) | 58 (13) | 60 (13) | 57 (13) | **0.15** | 0.08 |
| Age ≥65 years | 31% | 38% | 28% | **0.15** | 0.07 |
| Female | 84% | 81% | 80% | 0.08 | **0.10** |
| BMI, kg/m², mean (SD) | 30.6 (7.3) | 31.2 (7.7) | *n/a* | 0.08 | — |
| CDAI (0–76), mean (SD)* | 32 (13) | 19 (13) | *n/a* | **1.00** | — |
| Seropositive (RF/anti-CCP)* | 52% | 55% | *n/a* | 0.06 | — |
| Prior csDMARDs, mean (SD)† | 1.9 (1.0) | 1.7 (0.9) | 1.6 (0.9) | **0.21** | **0.32** |
| Prior bDMARDs, mean (SD)† | 1.4 (1.4) | 1.2 (1.3) | 0.9 (1.0) | **0.15** | **0.41** |
| Prior JAKi, mean (SD)† | 0.2 (0.5) | 0.2 (0.5) | 0.2 (0.5) | 0.00 | 0.00 |
| Oral glucocorticoids‡ | 36% | 54% | 66% | **0.37** | **0.63** |

Absolute SMDs calculated from summary statistics; bold = |SMD| ≥0.1. Rows without PRIMA-102 data are omitted.
* Among patients with clinical data. † All available prior data. ‡ 12 months before index. n/a = not available in source (claims lack laboratory results). CDAI = Clinical Disease Activity Index; RF = rheumatoid factor; CCP = cyclic citrullinated peptide.

**Table 2: Linkage, Uniqueness, and Validation by RWD Source**

| Measure | Practice EHR network | Claims |
|---|---|---|
| Contributing sites | 9 | 20 |
| Participants eligible for linkage | 318 | 1,131 |
| Matching method | Token + conditional | Token |
| Matched | 308 | 543 |
| Validated / unique 1:1, n (%) | 305 (99.0%) | 519 (95.6%)[b] |
| Multiple candidates (excluded) | 3 | — |
| Linked to this source only | 53 | 277 |
| Index treatment discordance: index therapy differs from study record[a] | 3 / 305 (1.0%) | 9 / 111 (8.1%)[b] |

[a] RWD shows a different b/tsDMARD than the study record. [b] Claims matches confirmed by an RA diagnosis; index therapy compared among the 111 with ≥6 months of coverage before and after index. Manual review of 61 practice EHR network token matches against source records: 1 incorrect (different site, no visit within 30 days of consent), 2 with multiple candidates.

**ACKNOWLEDGEMENTS** This study was sponsored by Aqtual.

**DISCLOSURES** JC, CT, and DA are employees of Aqtual. JRC receives research support/consulting fees from AbbVie, Amgen, Aqtual, Bendcare, BMS, CorEvitas, GSK, Janssen, Lilly, Moderna, Novartis, Pfizer, Sanofi, Scipher, UCB, SetPoint Medical, FASTER, and Tnacity Blue Ocean. SD, YS, SF, JZ, FX, AM, and AED have nothing to disclose.

## RESULTS

**51.3%** of participants linked to ≥1 RWD source

**95.6%** of claims matches validated by RA diagnosis

**99.0%** of EHR matches unique (1:1)

- **Linkage:** of 1,131 PRIMA-102 participants at 20 sites, **580 (51.3%)** linked to ≥1 RWD source; 250 linked to >1 source.
- **Uniqueness:** 305 of 308 (99.0%) practice EHR network matches were unique 1:1; 3 patients with multiple candidate records were excluded (Table 2).
- **Validation:** 519 of 543 claims matches (95.6%) were confirmed by an RA diagnosis. Of 61 token matches manually reviewed against EHR source records, only 1 was incorrect.
- **Index treatment discordance:** RWD index therapy differed from the study-reported therapy for 3 of 305 EHR-linked (1.0%) and 9 of 111 claims-linked (8.1%) patients (Table 2).
- **Representativeness:** PRIMA-102 was similar to both RWD cohorts in age, sex, and BMI (SMD ≤0.15); CDAI was higher than in EHR (32 vs. 19; SMD 1.00), with more prior DMARD use and less oral glucocorticoid use than both RWD cohorts (SMD 0.15–0.63) (Table 1).
- **Linked vs. non-linked:** EHR network patients were balanced on age, sex, and BMI (SMD ≤0.12) but had higher CDAI (25.2 vs. 18.7; SMD 0.50); registry patients were younger, more diverse, with fewer comorbidities.

## CONCLUSIONS

- **PRIMA-102 participants were broadly representative** of RA patients at the same practices; disease activity was higher, although more refined matching/weighting can correct for this.
- Linkage based on a vector of multiple non-unique identifiers (without tokenization) was highly accurate (95–99+% with validated, unique matches); most study-reported index therapies were confirmed in the EHR and claims data with few mismatches (e.g., related to medication samples given in-office).
- Linked RWD will extend follow-up and may support an external synthetic comparator arm using EHR and/or claims data for comparisons with the clinical cohort.