Does It Matter if I Stay or Go?  
Predicting Patient-Level Attrition to Evaluate Study Generalizability

Urmila Chandran, PhD1, Jenna Reps, PhD1, Patrick B Ryan, PhD1, Paul E Stang, PhD1  
1Janssen Research & Development, LLC, Titusville, NJ, USA

BACKGROUND
- Commercial claims databases are often criticized for unstable patient populations and limited follow-up time on patients due to varying insurance plans and employment status, affecting coverage.
- Disease registries and randomized trials have been touted as less biased data sources due to better subject retention and longer follow-up than claims databases.
- However, claims database studies often do not assess if loss of subjects actually induces selection bias and loss of generalizability within a given population.

OBJECTIVES
1. To evaluate if subjects with rheumatoid arthritis (RA) starting a biologic therapy that remain in a claims database for ≥5 years differ from those lost to follow-up.
2. To observe if loss to follow-up in RA patients in a claims database is comparable to loss to follow-up proportions in RA registries and randomized trials.

METHODS
Data sources: Truven MarketScan Commercial Claims and Encounters (CCAE) and Optum Clininformatics® Extended DataMart (OPTUM)
Study population: 1) Adults (≥18 years) with at least 2 claims for a diagnosis of RA on separate visits or 2) Adults with 1 prescription claim for a disease modifying anti-rheumatic drug (DMARD) within 90 days following a diagnosis claim for RA.

ANALYSIS – OBJECTIVE 1
- Variables: Large number of measured variables (age, gender, conditions, drugs, procedures, measurements, observations, Charlson index score) in 365 days prior to index were included in the prediction model.
- Models: Regularized logistic regression, random forest, and gradient boosting machine were evaluated.
- Training: Used 10-fold cross validation on training data (75% of data) to select optimal hyper-parameters.
- Validation: Internal validation of model done on remaining 25% of the data.

ANALYSIS – OBJECTIVE 2
- Average length of follow-up for RA patients from start of biologic therapy to end of 2016 was computed.
- Follow-up time computed from claims databases were contextualized against follow-up time observed in RA registries and randomized trials.

RESULTS – OBJECTIVE 1
- Based on measured variables in OPTUM, there is weak discriminative ability with test set AUCs ranging from 0.59-0.63 comparing RA patients that stayed for ≥5 years vs. those available for less than 5 years.
- In CCAE, discrimination was higher, but primarily due to age>65 being a discriminatory factor – most patients transition to Medicare around age 65 years.
- When excluding patients older than 60 years, discrimination dropped for all models in CCAE.
- Similar results of moderate to poor discrimination was also observed for patients with Crohn’s Disease, another autoimmune condition (data not shown).

CONCLUSIONS
- Patient populations in claims databases may have less biased samples than previously thought.
- Loss to follow-up in claims databases for patients using biologics for their RA appear to be similar to that observed in RA registries and randomized trials.
- The impact of attrition comparing initial users to those with several years of follow-up can be assessed using patient-level prediction in claims data studies, and should be encouraged to further characterize new user cohorts.