Does It Matter if I Stay or Go?

Predicting Patient-Level Attrition to Evaluate Study Generalizability

Urmila Chandran, PhD[1], Jenna Reps, PhD[1], Patrick B Ryan, PhD[1], Paul E Stang, PhD[1]

[1]Janssen 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 Clinformatics® 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.

Index date: First ever prescription claim (new user) for an approved biologic between 01-01-2008 and 12-31-2008.

Outcome of interest: Continuous observation for 1825 days (5 years) from index

Biologics approved for RA in US as of Dec 2008:
infliximab, adalimumab, etanercept, rituximab, abatacept, anakinra

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.

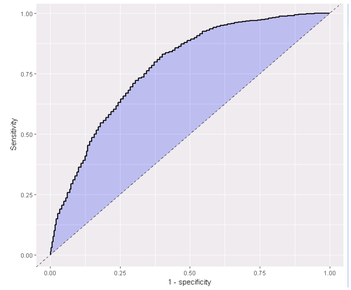
RESULTS – *OBJECTIVE 1*

- 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*).

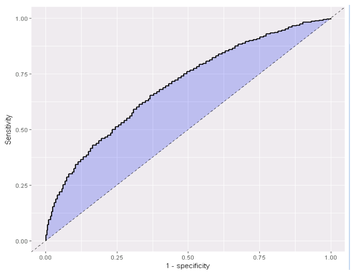
Discriminative performance of machine learning classifiers

Database	Classifier	Test set AUC
CCAE	Lasso Logistic Regression	0.69
CCAE	Random Forest	0.68
CCAE	Gradient Boosting Machine	0.68
OPTUM	Lasso Logistic Regression	0.59
OPTUM	Random Forest	0.63
OPTUM	Gradient Boosting Machine	0.59
CCAE	Lasso Logistic Regression (Age <=60)	0.62
OPTUM	Lasso Logistic Regression (Age <=60)	0.61

ROC plots for regularized logistic regression model (test set)



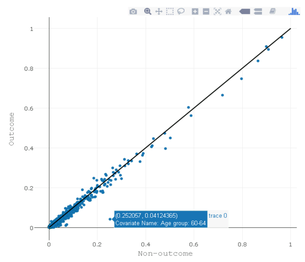
Trained on CCAE data



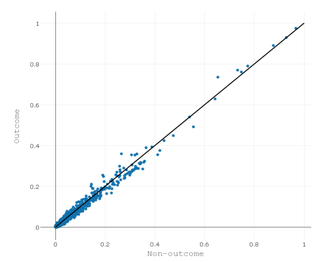
Trained on OPTUM data

Variable plots of outcome mean vs. non-outcome mean

CCAE data



OPTUM data



RESULTS – *OBJECTIVE 2*

- For RA patients that were new users of biologics in CCAE and OPTUM databases, and who had the chance to be in the database for at least 5 years, the average length of follow-up generally ranged from 3 years to over 6 years across biologics.
- Observed median follow-up time of patients in RA registries in North America and Europe range from 2.5-7.2 years[a].
- Drop out rates >20% in 1 year were observed in most RCTs on RA biologic therapies[b,c].

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.

Conflict of Interest Statement

The authors of this poster are full time employees of Janssen Research and Development, a unit of Johnson and Johnson. The work on this study was part of their employment. They also hold pension rights from the company, and own stock and stock options.

References: [a]Nyberg F, Askling J, Berglind N, Franzen S, Ho M, Holmqvist M, et al. Using epidemiological registry data to provide background rates as context for adverse events in a rheumatoid arthritis drug development program: a coordinated approach. Pharmacoepidemiology and drug safety. 2015 Nov;24(11):1121-32. [b]Singh JA, Hossain A, Tanjong Ghogomu E, Mudano AS, Maxwell LJ, Buchbinder R, et al. Biologics or tofacitinib for people with rheumatoid arthritis unsuccessfully treated with biologics: a systematic review and network meta-analysis. The Cochrane database of systematic reviews. 2017 Mar 10;3:CD012591. [c]Singh JA, Hossain A, Tanjong Ghogomu E, Mudano AS, Tugwell P, Wells GA. Biologic or tofacitinib monotherapy for rheumatoid arthritis in people with traditional disease-modifying anti-rheumatic drug (DMARD) failure: a Cochrane Systematic Review and network meta-analysis (NMA). The Cochrane database of systematic reviews. 2016 Nov 17;11:CD012437.