Name:	Urmila Chandran, PhD
Affiliation:	Janssen R&D LLC
Email:	uchandr1@ITS.JNJ.com
Presentation type (s):	**Poster**

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 R&D LLC, Titusville, NJ, USA

Abstract

Commercial claims data are a rich and critical resource for observational research. Due to the linkage to employment, researchers have lamented the perceived short follow-up time for patients in health insurance claims databases. We examined the length of follow-up in 2 US claims databases for patients with rheumatoid arthritis (RA), to compare them to dropout rates in RA disease registries and randomized trials. We also evaluated whether those patients with longer follow-up (≥5 years) represent a biased sample of the inception cohort. We used 3 predictive modeling approaches to predict who will be observed for the full follow-up period of 5 years using a large number of predictor variables. All 3 models showed low discrimination, suggesting that those with longer follow-up time in the claims data are not essentially different than those in the inception cohort. The average follow-up time among RA patients initiating biologics in the claims data was also comparable to that observed in RA disease registries and randomized trials.

Introduction

'Turnover' in the US healthcare system, where commercially insured employees (and their dependents) are often offered choice of insurance plans annually, have historically been a concern of observational database research in these populations[1]. In addition, the shift in culture of frequent changes in employment status or organization have added to the concern that these databases have limited follow-up time on patients. It has been speculated that disease registries and randomized trials may have better subject retention (and therefore less bias) than observational database studies. Although researchers typically address limitations of study designs and data sources, they fail to evaluate if the loss of subjects actually biases their sample within a given population. With a focus on RA, we evaluated if subjects that are lost to follow-up in a claims database differ from those that remain, and if loss to follow-up in RA patients starting a biologic therapy in claims data are comparable to observations from registries and randomized trials.

Methods:

The analyses were performed in two US claims databases, Truven Commerical Claims & Encounters (CCAE) and Optum Extended DataMart (Optum). Prediction models were developed

to assess differences between RA patients that were new users of biologics in 2008 and were followed up for ≥5 years versus those that dropped out before 5 years. This was done using a large number of measured variables in the 365 days prior to new use of a biologic. We used 10-fold cross validation on the training data (75% of the data) to select the optimal hyper-parameters and then internally validated the model on the remaining 25% of the data. Regularized logistic regression, random forest, and gradient boosting machine classifiers were trained and evaluated. The average length of follow-up for RA patients identified from start of biologic therapy to end of 2016 was also computed and contextualized against follow-up time observed in RA disease registries and randomized trials.

Results:
The results show that based on the recorded variables in the database, there is weak discriminative ability in Optum across the machine learning algorithms with test set AUCs ranging between 0.59-0.63 comparing RA patients that remained at least 5 years and those that dropped out before 5 years. In CCAE, the discrimination was higher with the regularized logistic regression obtaining an AUC of 0.68, but this was primarily due to age>65 being a discriminatory factor, as most patients transition from private insurance to Medicare at that time. When excluding those over 60 years of age in CCAE, the discrimination dropped to 0.62-0.64. The same effect was observed in Optum with AUCs of 0.57-0.61 among RA patients ≤60 years.

The average length of follow-up for patients starting different biologics in the claims databases generally ranged 3 years to up to 6 years depending on how long a drug had been on the market. Results were mostly similar in CCAE and Optum databases. Median time of follow-up ranged from 2-5 years depending on the biologic. In the RA literature, majority of randomized trials have observed more than 20% patient drop out within the first year[2]. Disease registries experience a wide range (3%-19%) of subject loss in the initial year, with median follow-up time across RA registries in North America and Europe observed to range from 2.5-7.2 years[3].

Conclusion

Populations in claims databases may have longer follow-up and less biased samples than previously believed, and loss to follow-up proportions in claims databases appear to be comparable to that observed in RA disease registries and randomized trials. Patient-level prediction of attrition can be used as a tool to assess study generalizability.

References

1. Marquis MS, Kapur K. Employment transitions and continuity of health insurance: implications for premium assistance programs. Health affairs. 2003 Sep-Oct;22(5):198-209. PubMed PMID: 14515896.
2. 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. PubMed PMID: 28282491.
3. 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.