**Predicting Primary and Secondary Adherence to RA Treatment: Revised Protocol**

A Joint Academic and Industry team of JHU, UAB, and Sun Pharmaceutical

University of Alabama at Birmingham: Jeffrey R. Curtis, MD, MS, MPH,

Sun Pharmaceutical team member: Charles Molta, MD

JHU team members: Hong J. Kan, PhD, Kirill Dyagilev, PhD, Peter Schulam, MS, Suchi Saria, PhD

**Introduction**

Rheumatoid arthritis is a chronic autoimmune disease that affects 1.3 million adult patients in the U.S. Available therapies for RA include NSAIDs, glucocorticoids, conventional Disease Modifying Anti-Rheumatic Drugs [DMARDs] (hydroxychloroquine, leflunomide, methotrexate, and sulfasalazine), biologic DMARDs (anti-TNFs: adalimumab, certolizumab, etanercept, golimumab, infliximab; others: abatacept, anakinra, rituximab, tocilizumab), and newer synthetic DMARDs (tofacitinib citrate).

Inflammatory conditions, such as rheumatoid arthritis, continue to rank as the costliest specialty therapy class.[[1]](#footnote-1) Adherence serves both as a goal for treatment improvement and as an overall measure of effective drug use in as much as it is a mediator of drug effectiveness in the short term, and it is a requirement for optimizing long term outcomes.

We propose to focus on predicting two adherence outcomes using integrated claims and EHR data:

* First, estimate predictors for “primary non-adherence” to new RA prescriptions, defined as prescriptions for RA medications that are written by a clinician, but not filled by the patient or administered (for infusion biologics);
* Second, conditional on patients starting prescribed treatments, predict persistence with those treatments, so that patients with low predicted probability to adhere can be the targets of multi-faceted systems, provider, and patient-level interventions to promote adherence and optimize treatment outcomes

We will also evaluate information gaps between claims only, EHR only, and EHR+claims for predicting adherence.

**Data Sources:**

Optum integrated claims and EHR data was used in this study. This integrated dataset was fully compliant with HIPPA requirements and was made available for this specific study through a contest organized by Health Datapolooza.

**Research Objectives and Study Design**

This is a retrospective cohort study using integrated claims and EHR data to understand factors associated with primary adherence and persistence in RA patients newly prescribed and treated with MTX and biologics/tofacitinib. Specific research questions are:

1. Evaluate the availability of various predictors of adherence available at baseline (time of drug prescribing/initiation) from EHR and claims.
2. Estimate predictors at baseline for primary non-adherence defined as prescriptions of MTX and biologics/tofacitinib recorded in EHR data but not filled or administered in RA patients who were newly prescribed with MTX and biologics/tofacinib.
3. Estimate predictors at baseline for persistence with MTX and biologics/tofacitinib in RA patients who initiated treatment with MTX and biologics/tofacitinib.

Study Sample:

The original sample provided contains about 120,000 patients from 2007 to 2015. Two cohorts of patients newly prescribed with methotrexate and biologics/tofacitinib will be extracted with the following criteria. A subcohort from each of the two cohorts that contains patients who initiated treatment will be extracted to study persistence.

1. Patients Newly Prescribed with Methotrexate (MTX)

This cohort consists of patients newly prescribed with an oral or subcutaneous (SC) methotrexate (index date is defined as the first prescription date) recorded in EHR. Below are the complete inclusion and exclusion criteria.

Inclusions: Patients must meet the following criteria: (1) with at least 12 months of pharmacy and medical insurance coverage before and at least 2 months of pharmacy and medical insurance coverage after a new MTX prescription recorded in prescription data[[2]](#footnote-2); (2) age >= 18 at index date; (3) at least 2 RA diagnoses made by a physician in medical claims that occurred 7 days apart between index-12 months and index+3 months. RA was defined with ICD9 diagnosis codes of 714.0, 714.2, or 714.81. An RA diagnosis must be made by a physician (defined by type of service in Medical claims indicating consultations, emergency room, inpatient visits, office visits, preventive medicine, or physical medicine/rehab). This cohort definition is supported by high quality validation studies that have shown that the positive predictive value of this algorithm exceeds 85% to identify RA patients.

Exclusions: (1) Using pharmacy and medical claims, any filling or administration of a biologic agent or tofacitinab during the period before index date + 30 days (? to be further tested) as long as a patient had full medical and pharmacy coverage (which may exceed 11 months in some cases depending on data availability); (2) Using pharmacy claims, any filling of MTX before index date; (3) Using medical claims, patients with diagnoses of psoriasis or psoriatic arthritis (696.x), inflammatory bowel disease (555.x, 556.x), ankylosing spondylitis (720.x), or cancer (ICD9 140.x – 208.x, other than 173.x [non-melanoma skin cancer is ok]) at least 12 months before index date as long as a patient had full medical and pharmacy coverage; (4) Patients with incomplete EHR data between index-12 months and index+2 months (This is to ensure availability of predictors from EHR during the 14 months period around index. But this criterion may need refinement if we end up not using any data from EHR 12 months prior).

The 12 months preceding MTX index date was defined as MTX baseline period. Note that in some cases, we included information during the 12-month baseline as well as the months before that as long as a patient had full pharmacy and medical coverage in order to avoid misclassification of certain key variables (e.g. prior biologic/MTX use) prior to that, if the data was available.

2. Patients Newly Prescribed with biologic/tofacitinib

The definition for the biologic/tofacitinib cohort is similar to that for MTX (note that these are 2 independent cohorts although one patient may be included in both cohorts). This cohort included patients who received a new prescription for a biologic (any of the biologics) or tofacitinib (b/t).

Inclusions: Patients must meet the following criteria: (1) with at least 12 months of pharmacy and medical insurance coverage before and at least 3 months of pharmacy and medical insurance coverage after a new b/t prescription was recorded in prescription data[[3]](#footnote-3); (2) age >= 18 at index date; (3) at least 2 RA diagnoses made by a physician in medical claims that occurred 7 days apart between index-12 months and index+3 months.

Exclusions: (1) Using pharmacy and medical claims, any filling or administration of a biologic agent or tofacitinab during the period before index date as long as a patient had full medical and pharmacy coverage (which may exceed 11 months in some cases depending on data availability); (2) Using medical claims, patients with diagnoses of psoriasis or psoriatic arthritis (696.x), inflammatory bowel disease (555.x, 556.x), ankylosing spondylitis (720.x), or cancer (ICD9 140.x – 208.x, other than 173.x [non-melanoma skin cancer is ok]) at least 12 months before index date as long as a patient had full medical and pharmacy coverage; (4) Patients with incomplete EHR data between index-12 months and index+3 months.

The 12 months preceding b/t index date was defined as b/t baseline period. Note that in some cases, we include information during the 12-month baseline as well as the months before that as long as a patient had full pharmacy and medical coverage.

In both cohorts defined above, we conduct the search for the first evidence of drug prescription starting from 2007 forward. If a patient had multiple episodes meeting the sample selection criteria, we keep the very first episode and discard the subsequent ones. This ensures that the sample only includes patients at the earliest stage of disease possible allowed by the data rather than late stage recalcitrant patients.

1. Subcohorts of patients who initiated MTX and biologics/tofacitinib

The subcohort of patients newly treated with MTX (b/t) includes patients from the cohort newly prescribed with MTX (b/t) who initiated the treatment within 2 months for MTX and 3 months for b/t. The rest of the sample selection criteria are the same except that we require 6 months of continuous enrollment with medical and pharmacy benefits after drug initiation in order to observe actual persistence (since death is allowed to happen anytime, in such cases, actual enrollment can be less than 6 months). Persistence of MTX, biologics and tofacitinib is evaluated based on pharmacy and medical claims.

Follow-up of drug use is right censored when patients lost full medical+pharmacy benefits after 12 months of drug initiation, at the time of death, or September 2015, whichever occurs first.

**Definition of Outcomes:**

Primary non-adherence:

For MTX, an oral or SC prescription recorded in EHR and not filled within 60 days in pharmacy claims is defined as primary non-adherence. Such cases were identified by searching for the first MTX prescription in the EHR and then verifying filling in claims. We ignored HCPCS codes for MTX injections in-office since this would not reflect the critical element of patients filling the drug at the pharmacy.

All biologics and tofacitinib were treated as one single “drug” in this study. Primary non-adherence is defined as a biologic/tofacitinib prescription recorded in HER that was not filled or administered according to claims within 90 days of prescription (more time is given as b/t may need more time to clear insurance hurdles and obtain prior authorization, or practical barriers such as travel arrangement for patients to return to start an infusion therapy). Such cases were identified by searching for the first biologics/tofacitinib prescription in the EHR and then verifying filling and administration status in claims. This outcome clearly harnesses the power of the integrated EHR and claims data together but juxtaposing what the clinician prescribed (in the EHR data) and what the patient actually filled/received (in the claims data).

Persistence:

Persistence with a medication was quantified by the number of days from initiation to discontinuation, as identified in pharmacy data. For MTX, a gap of 90 days in pharmacy claims data indicates discontinuation.

A gap of 90 days in pharmacy claims (for tofacitinib and self-injected biologics) and in combined pharmacy and medical claims (based on procedure codes for infused and self-injected drugs administered in the clinical setting) defined discontinuation, consistent with multiple prior studies[[4]](#footnote-4). The following tables showed how we calculated “days of supply” for IV and SC biologics and tofacitinib.

Table 1: Days of Supply for IV and SC biologics and tofacitinib

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | adalimumab | etanercept | Infliximab | certolizumab | golimumab | abatacept | rituximab | tocilizumab | tofacitinib |
| IV | - | - | 56 (usual starting dosing interval) | - | 56 (usual dosing interval) | 28 (usual dosing interval) | 183 (most typical interval between cycles) | 28 (usual dosing interval) | - |
| SC | Actual days of supply (indication: every 2 weeks, or once weekly if dose escalated) | Actual days of supply (indication: every 1 week) | - | Actual days of supply (indication: every 2 or 4 weeks) | Actual days of supply | Actual days of supply | - | - | Actual days of supply |

A binary indicator for persistence at 6 is created, which is the main outcome for persistence.

**Definition of Predicting Features:**

The following baseline factors are defined during the 12 months period before initiation or prescription of MTX and biologics/tofacitinib unless otherwise indicated. These factors were extracted from claims and EHR and tested for importance in predicting the outcomes using a formal variable selection procedure as well as clinical judgment. Not all of them were included in the final models.

* + Demographic data from the Patient table: race, region, gender, birth\_yr (or age), income (zip code based), education (zip code based), IDN vs. multispecialty practice, sourceid (provider group can be a hierarchy for modeling)
  + Insurance variables from the Member\_detail table: product (this variable changes over time), BUS (commericial or Medicare), ASO, CDHP (consumer driven health plan indicator for access to medical payment products such as HRAs or HSAs. this is potentially very important as high out of pocket will surely impact adherence including primary non-adherence)
  + Smoking status and BMI from OBSEVATIONS.
  + SDS (signs disease and symptoms) associated with all visits: RA-related recorded patient complaints such as sleep, depression, anxiety, daily activities, fatigue and pain (a full list of items is available upon request), and their positive and negative sentiment.
  + Hopkins ACG system outputs: EDCs and RxMG, high level indicators of disease groups based on diagnosis codes and pharmacy claims respectively (these variables are based on claims only for up to 12 months of data before drug initiation as the ACG system is based on 1 year inputs). They should include individual comorbidities such as depression, anxiety, sleep disorder, certain pain meds, COPD, etc.
  + Percentage of total visits with an RA diagnosis to rheumatologists based on Medical.provcat and provider.specialty during 12 months after RA incident diagnosis (ACR RA management guidelines recommend management of RA patients by rheumatologists. If there were too many PCP visits with RA diagnosis relative to rheumatologist visits, this may be an indicator for poor specialist access and other issues and may impact outcomes)
  + Number of office visits with a rheumatologist
  + Hospital use at baseline from the confinement table: number of hospitalizations and any use
  + ER visits at baseline from the confinement table: Number of visits and any use
  + Labs: Lab values extracted during the 12 months baseline before drug initiation/prescription using the test closest to drug initiation/prescription date as either a numeric value or, a binary indicator for abnormality from LABS tables. The following lab values were included: ESR (erythrocyte sedimentation rate), C reactive protein (CRP) [note that there are two kinds of inflammation tests, and two units for each of them – regular CRP, and high sensitivity CRP, and units can either be mg/L or mg/dL – the tests themselves do not always correspond to these units]. Other relevant labs include a rheumatoid factor (RF) and anti-CCP antibody, sometimes known as anti-citrullinated protein antibody (ACPA)
  + Disease activity: This information was based on RAPID3, CDAI, and SDAI at drug initiation/prescription +/- 1 months. We expected substantial missing data though. As a result, these multiple disease activity measures were grouped in the same categories of remission, low disease activity, moderate and high using their individual cutoff values as recommended by ACR[[5]](#footnote-5).
  + Pain scores recorded at drug initiation/prescription +/- 1 months in Observations, SDS and NLP\_measurement tables. Tender and swollen joint counts were extracted as well.
  + Physical functioning: HAQ, MHA, MDHAQ in NLP\_measurement during drug initiation/prescription +/-1 month.
  + “Allergies” to previous drugs: this information was searched in reasons for switching drugs in the drug rationale table.
  + Medication types: Anti-TNFs, other biologics, toficitinib, IV vs. SC.
  + Drug combination (presence of b/t for MTX and of MTX for b/t) around initiation/prescription.

More details of the extracted features are available upon request.

**Methods:**

For each of our four outcomes, we identified an initial set of candidate features. The candidates included demographic features (e.g. gender and age), information about the individual’s insurance plan (obtained from the MEMBER\_DETAIL table), clinical observation features, signs diseases and symptoms (SDS) features extracted from clinical notes, and features extracted using the Johns Hopkins University ACG system (these include high-level features extracted from an individual’s claims history).

This resulted in hundreds of possible features, which would cause our model estimates to over-fit to our relatively small sample. To address this issue, we used L1-penalized logistic regression (the “lasso”) to perform model selection. Specifically, we evaluated the cross-validated misclassification error across a wide range of L1-penalty parameters and chose the largest such penalty (i.e. the simplest model) that was within one standard deviation of the parameter that achieved the minimum. This criterion generally selects more conservative model complexities, which was especially important in this study because we primarily focused on model interpretability. We note, however, that in some cases the one-standard-deviation rule returned the null model (i.e. that including the intercept only), and so we chose the penalty that achieved the minimum cross-validated misclassification error.

After selecting an L1-penalty, we refit the penalized model on the entire dataset to select the features that we would include in our final model (i.e. those with non-zero coefficient in the regression). Finally, to obtain our effect estimates we refit the logistic regression without the L1-penalty using only the features with non-zero weights from the variable selection step.

To assess the generalization of our final feature set, we used four-fold cross-validation to estimate the logistic regression and make predictions on unseen patients. We computed the area under the curve (AUC) using the model’s predicted outcome probability for each fold. We note that while the weights for these models were learned independently for each fold, the way in which we chose the final features out of our candidate pool used all of the data. Because of this, our estimates of generalization performance are likely optimistic (i.e. biased upward). In future work, our goal is to obtain a larger study cohort so that we can assess generalization using a fresh sample.

1. <http://lab.express-scripts.com/lab/insights/industry-updates/us-rx-spending-increased-13-percent-in-2014> [↑](#footnote-ref-1)
2. We allowed death to occur anytime after MTX index date. Therefore, a patient may die within 3 months of index date, in which case the full medical and pharmacy coverage period may end at the time of death. [↑](#footnote-ref-2)
3. We allowed death to occur anytime after MTX index date. Therefore, a patient may die within 3 months of index date, in which case the full medical and pharmacy coverage period may end at the time of death. [↑](#footnote-ref-3)
4. [Use of oral and subcutaneous methotrexate in rheumatoid arthritis patients in the United States.](https://www.ncbi.nlm.nih.gov/pubmed/24942466) Curtis JR, Zhang J, Xie F, Beukelman T, Chen L, Fernandes J, Ginsberg S, Spettell C, Yun H, Saag KG, Schiff M. Arthritis Care Res (Hoboken). 2014 Nov;66(11):1604-11. [↑](#footnote-ref-4)
5. Anderson J1, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, Saag KG, O'Dell JR, Kazi S. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. Arthritis Care Res (Hoboken). 2012 May;64(5):640-7. doi: 10.1002/acr.21649. [↑](#footnote-ref-5)