Appendix D. PCATS Aim 2: CER Report on CTPs for Patients With pcJIA

PCATS Project Aim 2:

Comparative Effectiveness Research on Consensus Treatment Plans for pcJIA Patients

ABSTRACT

PUROSE: Juvenile idiopathic arthritis (JIA), a chronic inflammatory disease, is one of the most common childhood autoimmune diseases and a major cause of childhood disability. Treatment approaches are adaptive, where the 2nd line of treatment is determined adaptive to the patient responses to the 1st line treatment. It is unknown what 1st and 2nd line treatment strategy are most effective. The study is designed to compare effectiveness of two adaptive consensus treatment plans (CTP) recommended by childhood arthritis rheumatology research alliance (CARRA): early combination vs. the step-up plan, using advanced statistics causal inference methods applied to an electronic medical records dataset.

METHODS: We derived this inception cohort 1st DMARD user patient sample from a single institute electronic medical records Epic database dated from January 2009 to December 2017. Children 1-19 years of age, newly diagnosed with polyarticular course of JIA (pcJIA) and receiving their 1st prescription of disease-modifying anti rheumatic drugs (DMARDs) within 9 months after diagnosis are eligible. Patient with comorbid conditions of inflammatory bowel disease, celiac disease and trisomy 21 are excluded. The date of initiating 1st DMARDs prescription is the index visit. The primary study endpoint is clinical Juvenile Arthritis Disease Activity Score (cJADAS) at the 6 and 12 months follow up after the index visit. The secondary study endpoint is the patient reported health related quality of life as assessed by PedsQL™ at the 12 months follow up after index visits. Different statistics causal inference methods are considered for assessing the effectiveness of the 1st line treatment, including propensity score sub-classification, augmented inverse treatment probability weighting (AIPTW), regression adjustment, Bayesian additive regression tree (BART) and a Bayesian's structural model with GP prior as the matching tool (GPMatch). Bayesian GPMatch method, and an extended BART approach are applied to evaluate comparative effectiveness of the 1st and 2nd line time-varying adaptive treatment effect on the 12 months of cJADAS.

RESULTS: Out of 1,750 JIA patients enrolled in registry and captured in Epic during study period, 407 children were eligible: 283 (70%) treated with early combination plan and 124 (30%) step-up plan. Patients initiated on the early combination CTP had higher cJADAS scores compared to patients on the step-up CTP at the baseline, with the mean ± SD of 16.08 ± 7.10 vs. 12.43 ± 6.03 (Student P value <0.0001), and are more likely to be RF positive (13.7% vs. 4.9%, Chi-square P=0.002). Correcting for treatment-by-indication bias, the results of GPMatch suggest both CTP are effective in reducing disease activities, predicting 6.7±0.48 and 4.7±0.66 expected mean cJADAS score by 6 months if treated on the step-up and early combination respectively. Children treated on early combination plan on average gain 2.0 point more improvement in cJADAS by 6 month, with 95% confidence limit (CL) of (0.4, 3.6), compared to those treated on the step-up plan. Treatment effect does not vary by the JIA subtypes or the baseline cJADAS score. Other causal inference methods suggest similar effectiveness. The analyses of the 12 month outcome suggest early combination is more effective in reducing cJADAS regardless of the 2nd line treatment: the estimated improvement in cJADAS is 2.6 with 95% CL of (0.6, 4.6) if continued on the same treatment, and 2.3 (0.3, 4.14) if escalate from the

1st line treatment. Escalating the treatment at the 2nd line does not improve the cJADAS score at the 12 month, regardless of the initial treatment assignment. The predicted mean±SD potential outcomes under step-up CTP are 4.6±0.77 and 4.9±0.79 if the 2nd treatment is the same assignment and escalate respectively. The predicted mean±SD potential outcomes under early combination CTP are 2.1±1.11 (if stays the same) and 2.7±0.93 (if escalate). The same results are seen in the subset of patients who fail to achieve physician global <2 after six months of treatment.

CONCLUSIONS: Both early combination and step up CTP are effective in reducing disease activities, where early combination CTP is more effective leading to better cJADAS score by 6 and 12 months of treatment after diagnosis. Future studies are needed to investigate comparative effectiveness of CTP on the longer term of outcomes in cJADAS, including inactive disease and health related quality of life outcomes.

Contents

1. INTRODUCTION

Juvenile idiopathic arthritis (JIA) is one of the most common type of rheumatologic disease in children. The incident rate of JIA is approximately 10 per 100,000 for girls and 5.7 per 100,000 for boys[1].The cause of childhood arthritis is unknown, the current understanding of the disease etiology and pathogenesis are limited[2]. The advent of disease modifying anti rheumatic drugs (DMARDs), particularly biologic DMARDs, in the past two-decades have revolutionized the treatment approaches to JIA, making it possible to target for inactive disease as the treatment goal. Currently, the most prevalent practice is to start patient on a non-biological DMARD as the first line of treatment, then step-up by switching to or adding biologic DMARDs if the patients fail to make sufficient progress. The effectiveness of early combination of DMARDs vs. the mono therapy has been reported in studies in adult RA population[3]. In pediatric population, however, the evidence has been limited. In pediatric population, TREAT study randomized children with newly diagnosed non-systemic polyarticular course of JIA (pcJIA) to MTX and MTX plus etanercept groups. The study suggested that more patients achieve inactive disease status (40% vs. 23%, $p = 0.09$) after six month of treatment[4]. In systematic JIA, early admission of IL-1 treatment was shown to lead to more rapid achievement in clinical inactive disease[5,6]. Many suggest that there is a window of opportunity where early effective treatment could address underlying disease pathophysiology, prevent structural damage in joints, and thus promises for earlier and sustainable control of disease[4,7–9]. In the consensus conference led by Childhood Arthritis and Rheumatology Research Alliance (CARRA, https://carragroup.org/), a panel of clinical experts recommended early combination approach as one of three consensus treatment plans (CTP) for treating children with newly diagnosed pcJIA[10]. More recently, an international pediatric rheumatology task force recommended a newer set of guidelines for JIA treatment[7]. The report, while acknowledges the early aggressive treatment may better take advantage of the window of opportunities, caution there is still lack of sufficient evidence.

Utilizing electronic medical records collected during routine clinical care in a pediatric rheumatology patient cohort, this study is aimed to evaluate real world evidence on the effectiveness of early combination of biological and none biological DMARDs treatment, compared to the more conventional none biological DMARDs monotherapy, in treating children with newly diagnosed pcJIA. We design this observational study as an inception cohort new DMARD user study. Statistical causal inference methods are utilized to correct for treatment-by-indication bias[11], and to estimate the potential treatment outcomes if the patient could be treated on either early combination or step-up CTP adaptive the 2nd line treatment to the patient's responses to the 1st line assignment for 12 months. Because the performance of causal inference methods could be sensitive to model specifications, therefore, the study presented results of different causal inference methods applied to the data from electronic medical records.

2. METHODS

2.1. Study design

In order to evaluate the effectiveness of the early aggressive vs. step-up consensus treatment plan (CTP), we design this observational study as an inception cohort new DMARD user study. Patient 1-19 years of age, newly (<6 months) diagnosed with polyarticular course of JIA (pcJIA) following the CARRA operational definition based on the ILAR (International League of Associations for Rheumatology, http://www.ilar.org/) code, receiving prescription of either early combination DMARDs or none biologic DMARD monotherapy as the first line treatment within 9 months of diagnosis are eligible for the study. Patients with the following comorbid conditions of inflammatory bowel disease (IBD), celiac disease, and trisomy 2 are excluded, as DMARD may be used for treating comorbid conditions. No other exclusion criteria is imposed.

The study was approved by the IRB at Cincinnati Children's Hospital Medical Center (CCHMC), and was registered at the CT.gov (NCT02524340) and HSRProj (20153590).

2.2. Data Sources, Data Management and Quality Assurance

The electronic medical records (EMR) captured in the institutional Epic system serves as the primary resource for the study. A subset of the patients (N=215) captured in the Epic were also participants of a completed research study[12] . This NIH funded research study prospectively followed up a cohort of JIA patients cared for at the CCHMC pediatric rheumatology clinic (referred as QoL study hereafter). The data were collected from in-person interviews and manual chart review during patient's clinical visit. The clinical and demographics data elements collected from this completed research study overlap with the data information to be extracted from the Epic system, and the same participant in both study were linked by their medical record number. Therefore, the QoL study serves as the second data source for quality checking purpose. In addition, QoL study enriched quality of life measures not captured in the Epic system.

To check for data quality, the values of the same data fields were compared between EMR data extracted from Epic and the data from QoL study. The EMR data were flagged if any discrepancy were identified. A designate research assistant then conducted manual chart review on the flagged records, identifying: 1) if the data was collected in the Epic; 2) where it was recorded in the Epic; and 3) whether the same information were recorded at multiple locations within the Epic. Based on these findings and the follow-up discussions with the Epic system specialist and clinician Epic users, data extraction algorithms were revised, and re-evacuated. The process iterates until the results yielded minor or no discrepancies, either compared to the QA data, or from the previous iteration. Data elements to be extracted from the EMR and the corresponding clinical research forms are designed based on the existing literatures[10].

2.3. Patient Selection

All rheumatology clinical encounters for children diagnosed with pcJIA were extracted from Epic between January 1st 2009 and December 31, 2017. Patients who were diagnosed with pcJIA for at least two distinct visits by the pediatric rheumatologists were identified as pcJIA patients. To select the inception cohort, first, we identified for all pcJIA patients their first clinical encounter captured in Epic. If their first clinical encounter was within 6 months after the first date of diagnosis with pcJIA, then the patient was considered into inception cohort. Out of the inception cohort, new DMARDs user were identify next. For all inception cohort patient, the date of their first DMARDs prescription was set as the index visit. If the index visit was less than 9 months after the date of diagnosis, then the patient was identified as the new DMARD user. Finally, eligible patients were identified by applying the inclusion and exclusion criteria: 1-19 year of age, no comorbidity of IBD, celiac disease or trisomy 21, and been on their first DMARDs prescription for at least one month.

2.4. First Line and Second Line Treatment

All concurrent medication prescriptions recorded in the Epic for the eligible patients during the course of the study were extracted for each patient clinical encounters. All medication prescriptions were classified into biological DMARDs, none biological DMARDs, none steroid anti-inflammatory drug (NSAID), other none DMARD medication based on the prescriptions given. The start and end dates of the DMARDs at any given encounters were recorded. The type and duration of 1st line and 2nd line DMARD prescriptions were derived. If the patient received both biological and none biological DMARDs within 2 months of initiation of 1st DMARDs, then the patient is allocated into the early combination CTP arm. If a patient initiated on the non-biologic DMARDs and never taken biologic DMARDs for at least 3 months after, then the patient is allocated into the step-up CTP arm. The 2nd line of treatment is determined relative to their first line of treatment, depending on whether the treatment is adjusted from the previous treatment assignment.

2.5. Index and Follow-up Visits

For both groups of patients, their 1st DMARD assignment is considered as the index visit. The follow-up visits are determined relative to the index visit. The follow-up visits are identified from the clinical encounters that fall within the specified time. The timing of the 3 months follow up visits may vary by patients, determined according to the ending date of the first treatment course. If the first medication course ended within the 1-5 month after the index visit, then it is identified as the 3 months visit for the given patient. If the end date is longer than five months, the nearest clinical encounter next to 3 months following the index visit is identified as the 3-month visit. If no clinical encounter occurred during the 1-5 months window after index visit, then we consider the patient missing their 3 months follow up visit. Similarly, the 6 months follow-up visits were determined using the 1-5 months window after the 3 months visit date, or

using 5-8 months window after the index visit if the 3 month follow-up visit did not occur, by applying the same rule. The 9 and 12 months follow up visits were determined similarly. The time duration after the index visit are calculated for each patient. The asymmetric -1 and +2 month window were used to accommodate the possible delay in patient taking up medication after given prescription, as well as potential legged treatment effect. As an example, Figure 1 presents the clinical encounters (marked with circles, color coded according to the cJADAS score which is sub-classified into inactive, low, moderate, and high disease activities[13]), treatment courses (marked by colored lines – red denote bDMARD, blue nbDMARD, purple nb+bDMARD) and their identified index and follow up visits (marked by 3, 6, 9, 12 for the corresponding months of follow-up) in a subset of 20 patients initiated on nbDMARD.

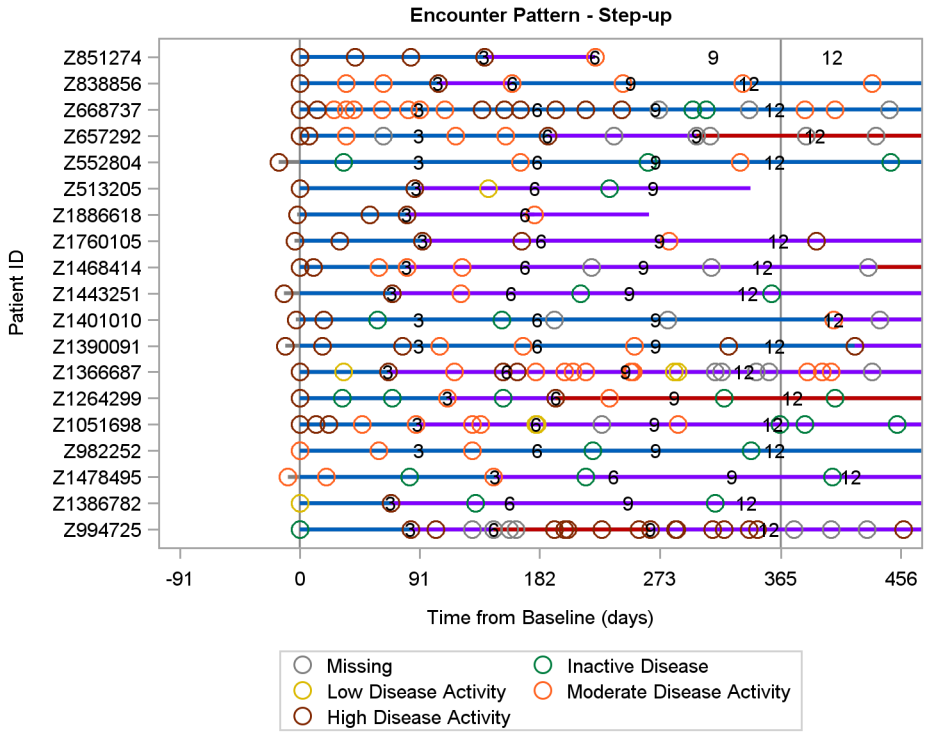


Figure 1 Clinical Encounter Pattern for Step-up CTP

2.6. Outcomes

The primary outcome is the clinical Juvenile Arthritis Disease Activity Score (cJADAS) at the 6 and 12 month of the follow-up visit. cJADAS is the disease activity measure validated, currently recommended and widely adopted as the clinical outcome measures in JIA research field[13]. The cJADAS is a summary score derived from physician global assessment of disease activity (ranges 0-10), patient global assessment of well-being (ranges 0-10), and active joint count truncated at 10[13]. The three core measures reflect different prospective of disease progression,

which are evaluated routinely during the clinical encounters. The secondary endpoint is the patient's health related quality of life as assessed by PedsQL general module total score[18]. Patients or their parent fill out the PedsQL generic module on an annual basis at CCHMC. Both of the cJADAS and PedsQL scores are bounded scores. cJADAS is bounded between 0 and 30, with higher score indicating more disease severity. PedsQL scores are bounded between 0 and 100, with higher score indicating better quality of life. The cJADAS are calculated for all visits using the three core measures extracted from Epic that is within +/- 1 month window around the identified follow up visit time point. If more than one clinical encounter occurred within the +/- 1 month window, then the averaged value of the specific core measure is used. The PedsQL generic scores are calculated for all visits following the user's manual. Since patients are only required to fill out PedsQL on an annual basis, the observed score is assigned to the nearest visit date for each patient within a 3-month window.

2.7. Covariates

Basic demographics of children include their age, race, gender, and insurance type. Disease characteristics include disease subtype, age of diagnosis, disease duration (i.e. difference between diagnosis and symptom onset), and age at the initiation of DMARDs. Biological variables include rheumatoid factor (RF; positive/negative), antinuclear antibodies (ANA; positive/negative), and erythrocyte sedimentation rate (ESR; normal range 0-10 mm/hour). Other than the three core measures used in the calculation of cJADAS score, patient reported global pain, duration of morning stiffness (none, <15 mins, >= 15 mins) and MD assessment of total number of joints with limited range of motion (LROM) are also collected. All these covariates are considered in the statistics causal inference procedure in order to correct for treatment-by-indication bias.

Over the course of the study, the clinical measures such as the biological, and cJADAS and other joint measures may change over time. These measures along with the duration of follow-up at each visit, are considered time-varying covariates in the analyses of adaptive CTP treatment.

2.8. Statistical Analyses

The baseline patient's demographic, insurance, and disease characteristics are compared between the two treatment arms. The propensity score is derived using the covariate balancing propensity score (CBPS)[14] method to ensure sufficient balance on the identified clinically important covariates between two treatment arms at the index visit on the age, gender, race, JIA subtype, insurance status, age at diagnose of disease, duration of disease at the time of diagnosis, MD global assessment, pain VAS, patient wellbeing, morning stiffness, ANA, RF, ESR, active number of joints, number of joints with lost range of motion, and baseline cJADAS score. The covariate balance plot presenting the standardized absolute mean difference between the two treatment groups before and after CBPS weighting. The balance is considered satisfactory if the

absolute standardized difference is <=0.2. The empirical distributions of the covariates are compared between the two treatment groups before and after weighted by the CBPS.

For comparing effectiveness of the 1st line treatment, we considered some widely adopted causal inference methods, including propensity score (PS) sub-classification matching, linear regression with PS adjustment, linear regression with spline fit PS adjustment, augment inverse probability of treatment weighting (AIPTW), the Bayesian additive regression tree (BART) and GPMatch. GPMatch is a Bayesian nonparametric casual inference method using Gaussian process prior as the matching tool. The GP prior is formulated in such a way that, for each individual patient in the sample, the GP prior will allocate different weight to information obtained from other individual patients, based on the confounding variables specified in GP prior. As such patients who are similar (dissimilar) to the given patient are contributing more (or less) information in estimating the expected outcomes for the patient if treated under different treatment option. By implementing matching and flexible modeling in the same step, the GPMatch offers protection against potential model misspecifications and produce accurate treatment effect estimates for real world CER setting. Further, it offers a natural solution to evaluate comparative effectiveness of time-varying adaptive treatment. We also extended the BART for the two-staged time-varying adaptive treatment. Both GPMatch and the extended two-staged BART are used for evaluating the adaptive CTP, both 1st and 2nd line, on the 12 month endpoints.

For all methods, the same set of covariates, along with the duration of follow-up visit after the baseline are used in the outcome models. For the regression model with including spline fit PS, the B-spline of estimated CBPS is use. Since the cJADAS is a bounded summary score, Tobit regression is used in all regression type (except BART) of analyses[15]. The comparative effectiveness of early-combination CTP compared to the step-up CTP were reported for the 6 month and 12 month outcomes.

Missing data are expected in the study analyzing existing data. Analyzing existing electronic medical records data, the missing data could be due to two primary reasons. First, the EMR may fail to capture the data for some patient encounters. Second, patient may not interact with the health care system, and result in missing data for the given follow-up time points. For the first case, the best efforts were given to discover the existing data records from the EMR system. The 2nd case, however, are much harder, as there are many potential reasons underlie when and how frequent a patient may interact with their health provider. To handle missing data at baseline, we applied Bayesian multivariate missing data imputation technique, hierarchically coupled mixture model with local dependence (HCMM-LD)[16] structure method. HCMM-LD is a Bayesian nonparametric missing data imputation technique, specifically designed to model the jointly distribution of the multivariate data. By jointly modeling the multivariate data, this method avoids the issue of none congeniality of many widely used missing imputation methods, including the multivariate imputation chained equation (MICE). The diagnosis of missing data

imputation is presented by presenting kernel fit of the distributions of variables before and after imputation. For sensitive analyses, missing data are also imputed using MICE.

The outcome model included time duration since baseline in the model to adjust for the different follow-up time. At last, the causal treatment effect at 6 and 12 months are derived by estimating the averaged treatment effect over all five simulation sets. The final results from each of the five sets of multiple missing data imputation are combined using the widely used Rubin's rule[17].

It is perceivable that patient with different disease subtype, and disease activities at the baseline, and duration of treatment may have different treatment effect. The heterogeneous treatment effects (HTE) are evaluated using the GPMatch approach. The potential none linear treatment effect at different levels of baseline cJADAS are considered by including the corresponding its interactions with treatment. The model fitting is evaluated using Bayes Factor (BF). Only when the model with including heterogeneous treatment effect offers strong evidence (BF > 3) of better model fit, we consider the HTE.

Statistics causal inference methods require no presence of unmeasured confounders. To evaluate the sensitivity of estimates of causal treatment effect to potential unmeasured confounders, analyses were repeated by including additional baseline covariates. Specifically, patient reported health related quality of life at the baseline were not included in the primary analyses for the concern over the large percentage of missing (>50%) at the baseline. The sensitivity analyses included PedsQL measures available at the baseline into the multiple imputation of the missing baseline covariates, then applying the same analyses procedures. The estimated causal treatment effects are also reported for the 6 and 12 month outcomes.

3. RESULT

3.1. Descriptive Statistics of Study Sample

Out of 1,750 JIA patients enrolled in registry and captured in Epic during the period of January 1st 2008 to December 31st, 2017, date of diagnosis could not be found in 181 patients. Of the remaining 1,569, 1,048 (67%) patients were captured in the Epic as the new patients, of whom 219 (21%) patients did not take any DMARDs at any time. Of these 829 (79%) new patients took DMARDs at some point, 197 (24%) did not meet the operational definition of pcJIA; 7 (0.8%) patients fall outside the inclusion criteria for age (1<age<=19) at the time of receiving1st DMARD prescription; and 32 (4%) patients initiated DMARDs prior to diagnosis of pcJIA, 89 (11%) initiating DMARDs later than 9 months post diagnosis. Out of 530 eligible new patients, 47 of them were excluded due to comorbid conditions (IBD, celiac, and trisomy 21). At last, 76 patients initiated on biological DMARDs monotherapy, thus were not considered into this analyses. A total of 407 patients received either combination or nbDMARDs, with 283 (70%) initiated on the non-biology DMARD monotherapy; and 124 (30%) on early combination. The study Flowchart is presented in the Figure 2.

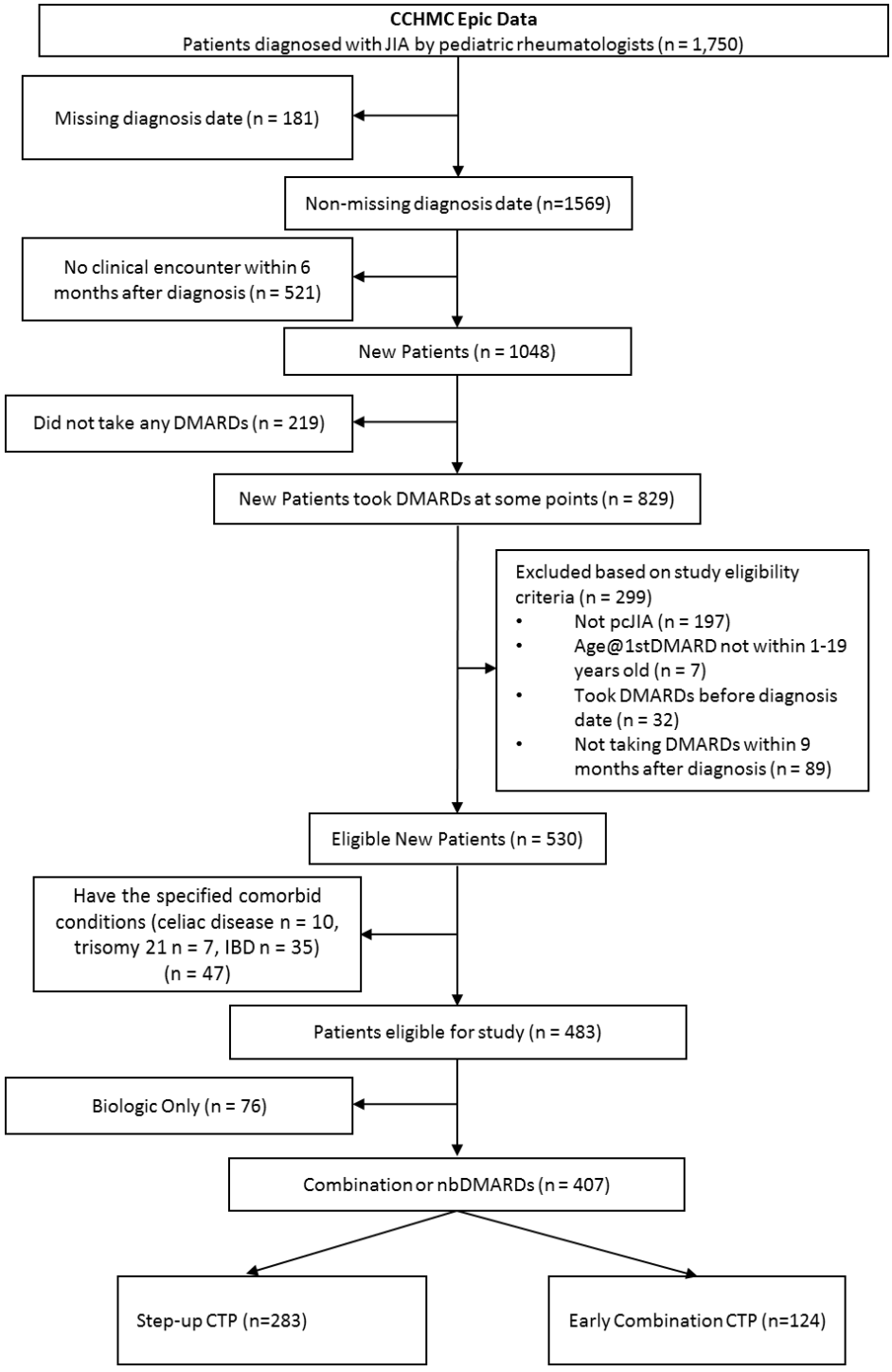


Figure 2 Study Flowchart

Table 1 provides the baseline characteristics of patients by treatment group. The index visit ranges from 0 to 5.9 month (median 0.3, Q1-Q3 0-1.5) after diagnosis, with mean±SD of 0.12±0.16 in the nbDMARD and 0.10±0.17 in the nb+bDMARD group (P=0.56). The treatment by indication bias is clearly evident, showing patients on early combination had significantly more active disease presentation at the index visit, with mean±SD of MD global 5.11 ± 2.70 vs.4.18 ± 2.48 (Student-T P=0.0031), patient wellbeing 4.48±2.70 vs. 3.45±2.49 (Student-T P=0.0005), active joint count 12.03±12.25 vs. 7.24±8.58 (Student-T P<0.0001), and cJADAS 16.08±7.10 vs. 12.43±6.03 (Student-T P<0.0001). Patients on early combination are more likely to be RF positive (13.7% vs. 4.9%, Chi-square P=0.002), reporting more than 15 minutes morning stiffness (72% vs. 56%, Chi-square P=0.03). Also presented in the *Table 1* are the follow up time for all patients at the 6 and 12 month of visit.

Table 1 Baseline Patient Characteristic by 1st Line Treatment

	nbDMARD (N = 283)		nb+bDMARD (N = 124)		
Baseline Variable	N	Mean ± SD	N	Mean ± SD	P-value
Age (year)	283	9.61 ± 5.14	124	10.19 ± 4.74	0.2824
Age of Diagnosis (year)	283	9.49 ± 5.16	124	10.09 ± 4.78	0.2726
Onset Age (year)	234	8.17 ± 4.93	101	8.69 ± 5.07	0.3782
Disease Duration at Diagnosis (year)	235	1.41 ± 2.30	100	1.76 ± 3.08	0.2629
Time Since Diagnosis (year)	283	0.12 ± 0.16	124	0.10 ± 0.17	0.5601
Six month Visit (in year after Baseline)	266	0.50 ± 0.06	117	0.49 ± 0.06	0.2524
Twelve month Visit (in year after Baseline)	245	1.00 ± 0.06	110	0.99 ± 0.07	0.1988
CJADAS10** (0 - 30)	194	12.43 ± 6.03	91	16.08 ± 7.10	<.0001
Active Joint Count** (0 - 71)	259	7.24 ± 8.58	116	12.03 ± 12.25	<.0001
Well Being** (0 - 10)	250	3.45 ± 2.49	111	4.48 ± 2.70	0.0005
MD Global** (0 - 10)	213	4.18 ± 2.48	98	5.11 ± 2.70	0.0031
Limited Range of Motion** (0 - 71)	259	5.38 ± 6.90	116	9.67 ± 11.69	<.0001
Erythrocyte Sedimentation Rate (mm/hr)**	151	20.58 ± 20.46	76	30.45 ± 28.30	0.0030
Global Pain VAS** (0 - 10)	254	4.16 ± 2.72	113	5.00 ± 2.72	0.0071
	N	Row %	N	Row %	
Female	204	72.1	91	73.4	0.7866
Race					0.8600
White or Caucasian	250	88.3	109	87.9	
Black or African American	17	6.0	8	6.5	
Other	11	3.9	6	4.8	
Unknown	5	1.8	1	0.8	
JIA Subtype+					0.0513
Poly RF-	99	35.0	56	45.2	

Poly RF+	21	7.4	14	11.3	
Oligo	90	31.8	26	21.0	
Other	73	25.8	28	22.6	
Insurance - Private	186	67.1	84	67.7	0.4550
Rheumatoid Factor - Positive**	14	4.9	17	13.7	0.0022
Antinuclear Antibodies - Positive	22	7.8	16	12.9	0.1016
Elevated C-reactive Protein**	35	23.8	34	43.0	0.0028
HLA-B27 - Present+	10	3.5	10	8.1	0.0516
Morning Stiffness*					0.0262
None	51	27.1	15	16.0	
15 Minutes	32	17.0	11	11.7	
> 15 Minutes	105	55.9	68	72.3	
Uveitis Ever	10	6.6	4	6.5	0.9727
Previous Treatment with NSAID**	221	78.1	75	60.5	0.0002
Previous Treatment with Prednisone	18	6.4	13	10.5	0.1489

Note: ** indicating P value <.01; * indication P value <.05. + indicating P value <.10.

3.2. Descriptive of Treatment Patterns

Figure 3 lists the top three DMARD prescription patterns in the step-up and early combination treatment group. Of the 283 patients initiated on the nbDMARDs, majority (N=271, 95%) were prescribed on methotrexate. Nine patients not yet had their 3 month follow up visit. Of the remaining 274 patients, 252 (92%) remained on the same prescription assignment, 21 (7.7%) got off from DMARDs, only 1 patient changed the DMARD prescription from MTX to Sulfasaline. At six month, additional 8 patient not yet had their 6 months follow up. Of the remaining 266 patients, 86 (32%) step-up in the sense that they changed their initial nbDMARD prescription or added bDMARDs; majority160 (60%) stayed on the same initial prescription; and 10 (3.7%) patients got off DMARDS.

Of the 124 patients initiated on b+nbDMARDs, 75 (61%) are prescribed on the MTX and Etenercept combination, and 32 (26%) were prescribed on the MTX and adalimumab combination. Only 1 patient not yet been followed up for 3 months. Of the remaining 123 patients, at the 3 month follow up visit, nearly all patients (103, 84%) stayed on the same prescription from the baseline. At the six month, additional 6 patient not yet had their 6 month follow. Of the remaining 117 patients, majority (91, 78%) continued on the same initial prescription, 11(9.4%) and 15(12.8%) patients either adjusted up (change or add new medication) or adjust down (drop previous mediation).

Given the treatment patterns, the baseline medication is allocated as the 1st line treatment, and the medication at the 6 months follow up is allocated as the 2nd line treatment.

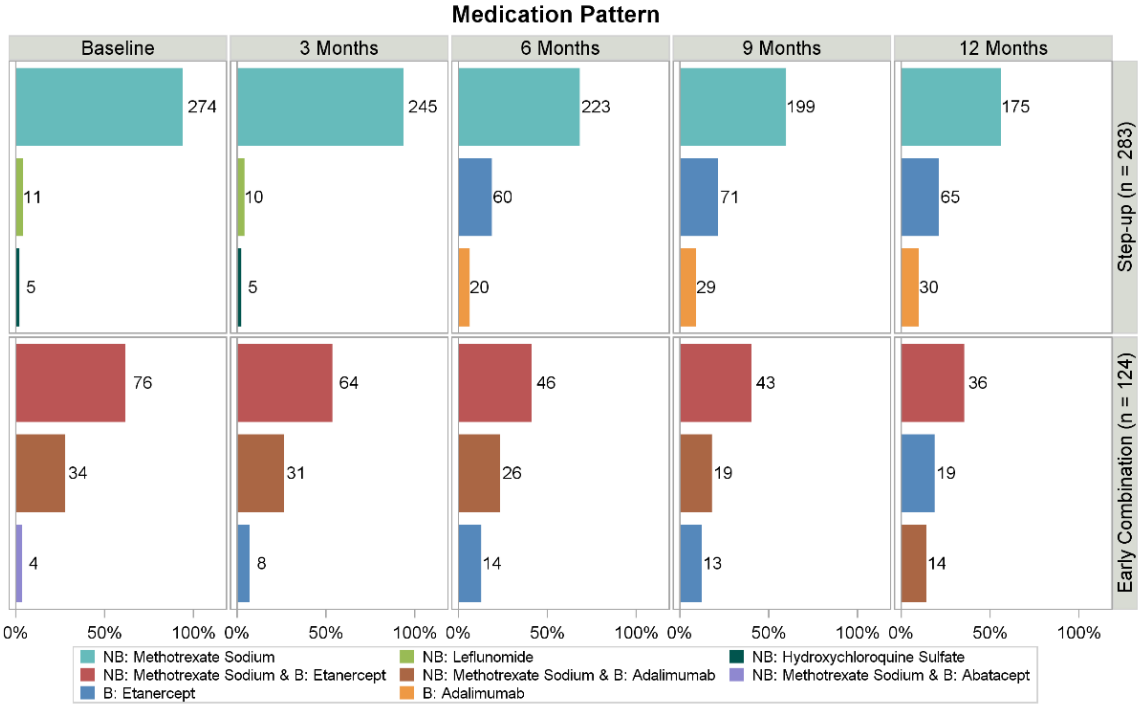


Figure 3 Top Three DMARD Prescription Patters over Study Visits

3.3. Descriptive of Outcome Measures

cJADAS is a summary score of three core measures (physician global, patient wellbeing and the active joint count), missing of any of the three core measure will lead to missing cJADAS. Because of retrospective study, not all patients are captured during the study visit window for the given visit. In addition, some patients are not yet had their 6 or 12 month follow-up visits, The Table 2 reported the number of non-missing cJADAS and the number of patients had the corresponding follow up visit, and the % of missing. We can see from Table, the cJADAS were observed in 194 (68.55%) and 91 (73.39%) patients in step-up and Early combination group at the baseline respectively (Table 2)., The missing cJADAS outcome were slightly higher in the month 6 and 2 follow-up visit. . The box-whisker plot of the cJADAS score by study arm shows disease activity improves in both arms. Although patients prescribed on the early combination presented much sever disease activities at the baseline, they are no longer statistically different at the 6 and 12 month months. The median (25%-tile, 75%-tile) of cJADAS scores in patients initiated on the early combination and step-up treatment are: 16.0 (11.0, 22.0) vs. 12.0 (8.0, 17.0) (KW rank-sum test P<0.0001) at the baseline; 4.75 (1.5, 8.75) vs. 6.0 (1.0, 10.0) (KW rank-sum test P=0.29) at 6 month; and 3.0 (0.5, 6.5) vs. 3.0 (0.5, 7.5) (KW rank-sum test P=0.87) at 12 month. Due to missing data, only 194, 137 and 95 of step-up patients and 91, 64

and 44 of early combination patients have non-missing cJADAS score at the baseline, 6 and 12 months respectively.

Table 2. Non-missing cJADAS at Study Visits

	Step-up (N=283)	Early combination (N=124)
Baseline non-missing cJADAS	194/283 (68.55%)	91/124 (73.39%)
6 months non-missing cJADAS	137/266 (51.50%)	64/117 (54.70%)
12 months non-missing cJADAS	95/245 (38.78%)	48/110 (43.64%)

Note: Numbers reported in the table are $\frac{\#\ non-missing\ cJADAS}{\#\ patients\ had\ study\ visits}$.

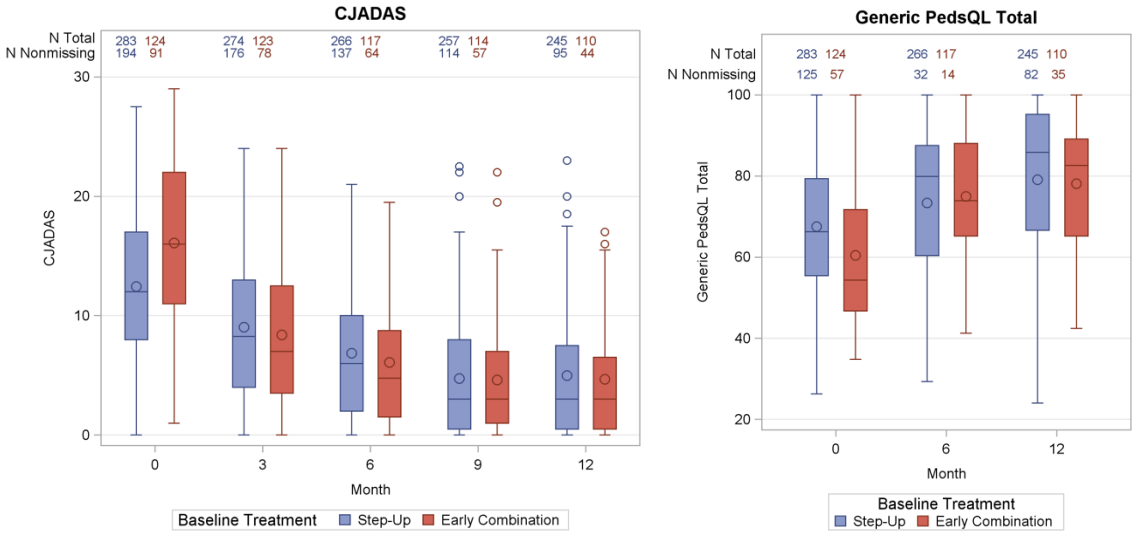


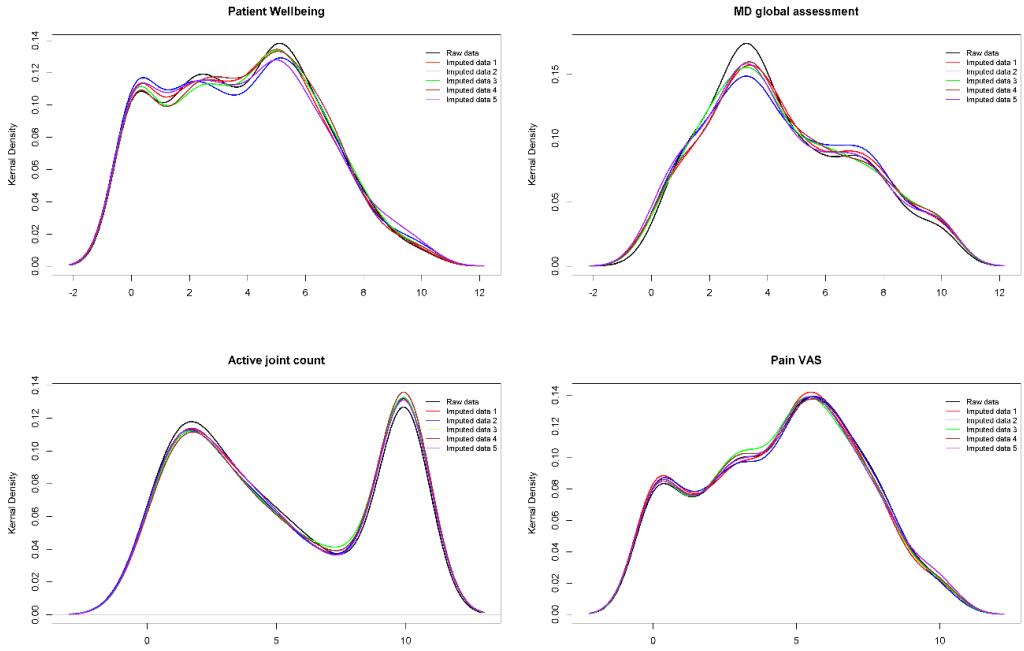
Figure 4 Box-Whisker Plots of the Study Outcomes

Figure 4 box-whisker plot of the PedsQL generic total score by study arm shows patient quality of life improves in both arms. Similarly, patients prescribed on the early combination presented much poor quality of life at the baseline than the nbDMARD patients, but they show similar distributions at the 6 and 12 month follow up. The median (25%-tile, 75%-tile) of PedsQL generic scores in patients initiated on the early combination and step-up treatment are: 54.3 (46.7, 71.3) vs. 66.3 (55.4, 79.3) (KW rank-sum test P=0.005) at the baseline; 73.9 (65.2, 88.0) vs. 80.0 (60.3, 87.5) (KW rank-sum test P=0.97) at 6 month; and 82.6 (65.2, 89.1) vs. 85.8 (66.7,

95.) (KW rank-sum test P=0.39) at 12 month. Only 46 (32 in step-up, 14 in early combination) and 117 (82 in step-up, 35 in early combination) patients had at PedsQL score at 6 and 12 month respectively.

3.4. Missing Data Imputation

Presented below in Figure 5 are the distributions (kernel fit to continues and bar plot of the categorical variables) of five imputed datasets using HCMM_LD overlaying with the observed data before imputation. The results show nearly identical distributions before and after imputation.



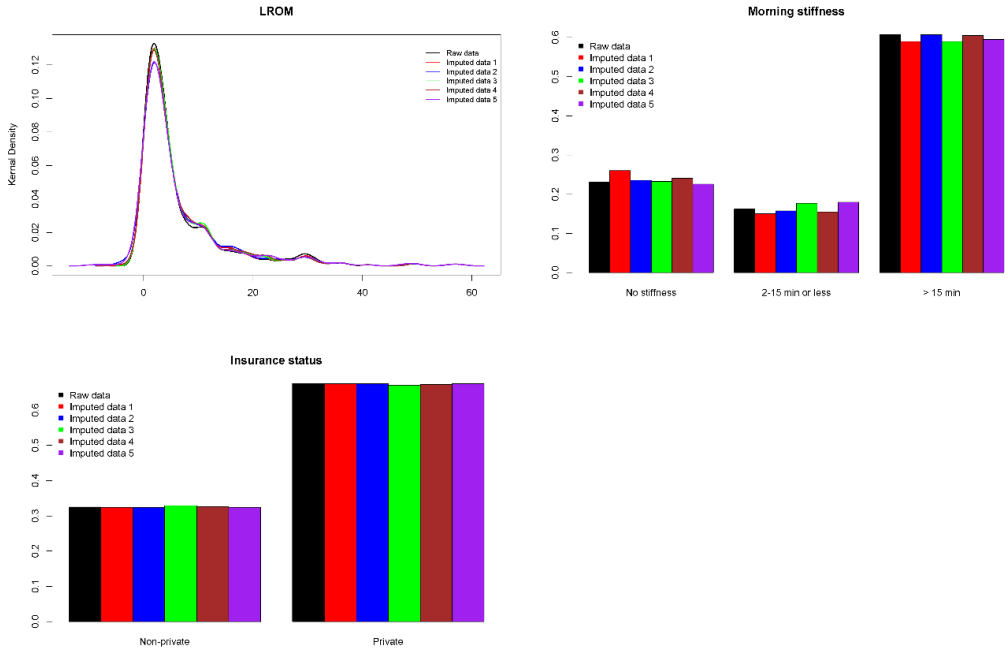


Figure 5 Kernel Plots of Distributions of Covariates before and after Imputation

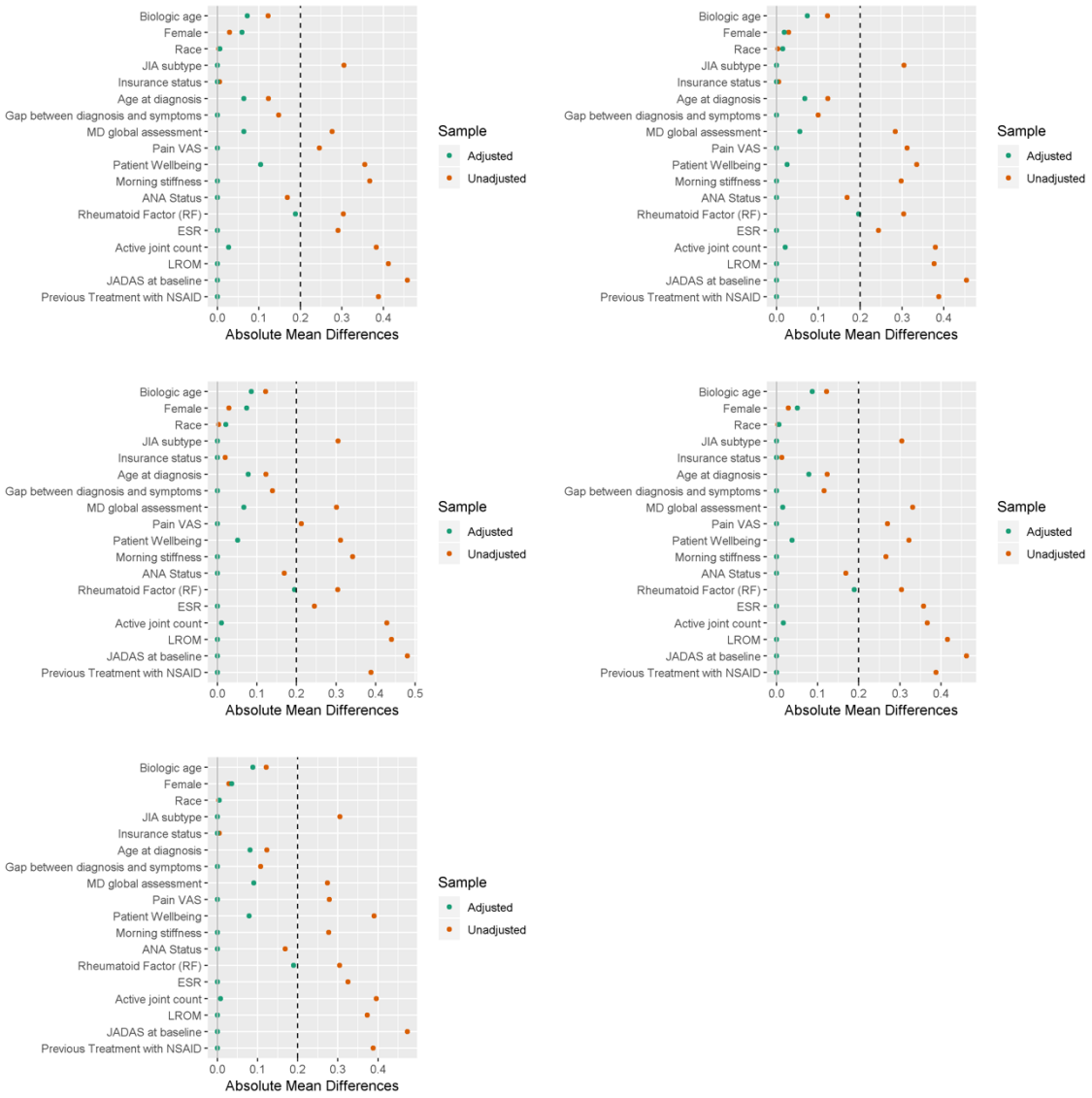


Figure 6 Balance Check of the Propensity Scores for Each of the Five Imputed Data

3.5. Propensity Score Estimation and Balance Checking

The propensity scores were derived using the CBPS method applied to the pre-determined important baseline confounders, for each simulated dataset. The CBPS are able to achieve desired covariate balance within the 0.2 absolute standardized mean difference, and comparable distributions in all pre-defined important confounders for each of the five imputed datasets (Figure 6).

The propensity score methods were used for estimating the effectiveness of the 1st line treatment approach in children newly diagnosed with pcJIA, for each one of the five simulated datasets. The CER results from PS methods are summarized over all five simulation sets, which are reported in the sections below.

3.6. Comparative Effectiveness of the Primary Endpoint

After 6 month of treatment, compared to the step-up treatment plan, the patient initiated on early combination is expected to gain more reduction in disease activity (see Table 3). The GPMatch result suggested suggest both CTP are effective in improving data activities, predicting 6.7±0.48 and 4.7±0.66 expected mean cJADAS score by 6 months if treated on the step-up and early combination respectively. Early combination CTP led to a significant -1.98 more reduction in cJADAS with 95% CI estimate of (-3.55, -0.40).

Table 3 Averaged Treatment Effect of Early Combination vs. Step-up CTP on cJADAS at 6 Months

	Est	SD	95%LL	95%UL
Direct Modeling	-1.472	0.685	-2.815	-0.129
Regression by PS quintile stratification	-1.450	0.771	-2.961	0.060
Regression with PS	-1.573	0.716	-2.976	-0.171
Regression Adjustment with spline fit PS	-1.642	0.720	-3.052	-0.232
PS weighted regression	-1.760	0.504	-2.749	-0.771
AIPTW	-1.379	1.199	-3.730	0.972
Propensity score quintile sub classification	-1.331	0.712	-2.726	0.065
IPTW	-1.299	0.658	-2.589	-0.009
BART	-0.990	0.856	-2.668	0.688
GPMatch	-1.975	0.805	-3.552	-0.398

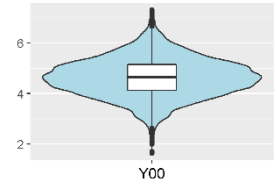
Note: Negative estimate indicating early aggressive treatment result in lower cJADAS score

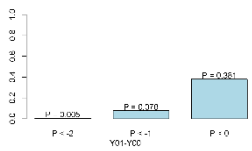
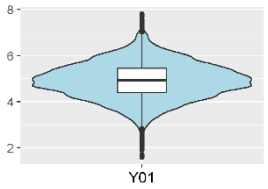
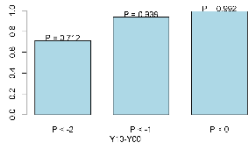
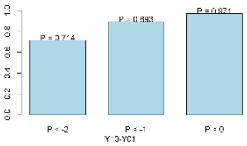
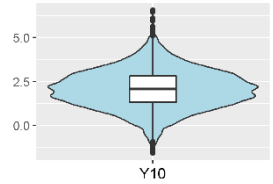
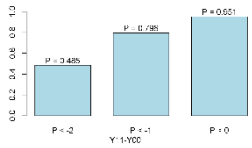
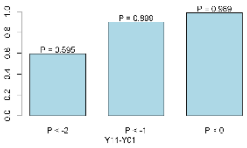
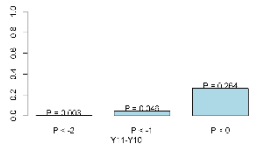
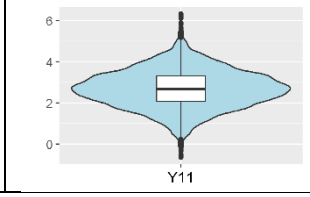
The propensity score methods, including PS quantile matching IPTW and AIPTW method led to similar but statistics non-significant results. While the regression with PS adjustment, linear and spline fit of PS, AIPTW methods reported approximate -1.5 points statistically significant improvements in cJADAS. Interestingly, the BART results the least effect size -0.70 (95% CI –

2.67, 0.69). The potential heterogeneous treatment effect were considered in BART and GPMatch models, by including potential treatment by JIA subtype and baseline JADAS score interactions. The Bayesian factor (BF) of 1.04, suggests no strong evidence to support subgroup treatment effect by baseline cJADAS score. Detailed results of subgroup analyses are reported in the supplemental material section S2.

Depends on the disease progression at the end of 1st line treatment, medication are adjusted correspondingly, and the cJADAS outcome are measured at the end of 2nd line of treatment. For the patient initiated on the nbDMARD, they may escalate their initial nbDMARD medication by changing to a different nbDMARD or adding bDMARD medication, we label their treatment by (01); alternatively they may remain on the same prescription or getting off from DMARD, their treatments is denoted by (00). For the early combination patient, patients may remain on the same aggressive treatment, remain on the combination or change one of either bDMARD or nbDMARD prescription, denoted by (11); or they may getting off bDMARD and/or nbDMARD (10). GPMatch method estimates the posterior distribution of cJADAS outcome, had the patient gone through each one of the four possible two-staged adaptive treatment, for each patient. The predicted mean±SD potential outcomes are: Y00 = 4.6±0.77, Y01 = 4.9±0.79, Y10 = 2.1±1.11, & Y11=2.7±0.93. On the diagonal of the Table 4, the violin plot of the estimated potential outcomes are presented under each of the four possible 1st and 2nd line treatment course. The upper triangle of the Table 4 and Table 5 presents the mean (95% CL) of the averaged treatment effect contrasted between each pair of the potential outcomes. The lower triangle presents the probability of achieving at least 0, 1 and 2 points improvement contrasting the two alternative potential outcomes. The averaged treatment effects were also calculated using the extended two-staged BART and presented in the upper triangle of the Table 4 and Table 5 in gray fonts. The results are summarized over all subtypes of pcJIA, no heterogeneous treatment effect was identified for any subtypes of pcJIA. The results are further estimated in the subset of patients who are considered non-responders to the 1st line treatment as determined by MD global assess >2 after six months of treatment (see Table 5).

Table 4 GPMatch &BART CER Results of 1st and 2nd line Treatment for all patients

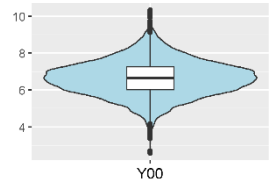
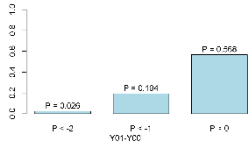
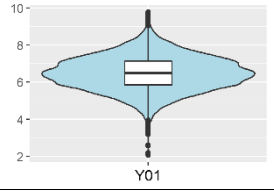
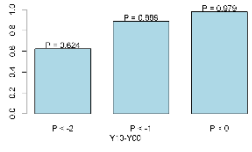
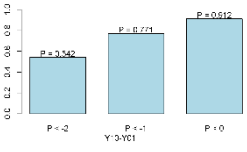
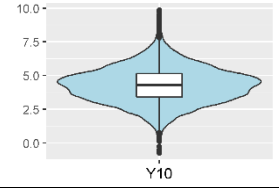
cJADAS	Y00	Y01	Y10	Y11
Y00		0.27 (-1.48, 2.02) 0.54 (-1.12, 2.21)	-2.57 (-4.58, -0.56) -1.85 (-3.83, 0.13)	-1.96 (-4.25, 0.34) -1.41 (-3.60, 0.78)

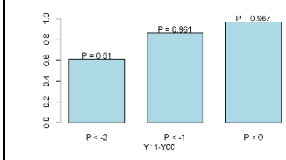
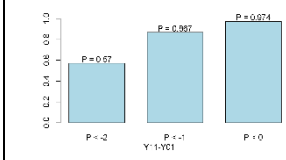
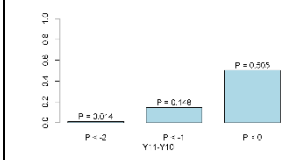
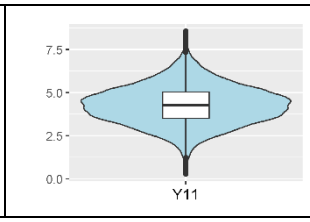
Y01			-2.84 (-5.74, 0.07) -2.40 (-5.12, 0.33)	-2.23 (-4.14, -0.31) -1.96 (-3.85, -0.07)
Y10				0.61 (-1.25, 2.48) 0.44 (-1.37, 2.25)
Y11				

Note:

- *Negative treatment effect indicating the column potential outcome is better than the row potential outcome in reducing cJADAS value.*
- *Results of two-staged BART are reported in gray fonts. All other reported estimates are due to GPMatch method.*

Table 5 GPMatch &BART CER Results of 1st and 2nd line Treatment among the non-responders

cJADAS	Y00	Y01	Y10	Y11
Y00		-0.16 (-2.04,1.71) 0.44 (-1.32, 2.19)	-2.34 (-4.51, -0.18) -1.83 (-4.14, 0.47)	-2.34 (-4.84, 0.13) -1.53 (-4.06, 0.99)
Y01			-2.18 (-5.29, 0.94) -2.27 (-5.32, 0.77)	-2.19 (-4.33, -0.06) -1.97 (-4.19, 0.25)
Y10				-0.01 (-1.84, 1.82) 0.30 (-1.63, 2.23)

Y11					

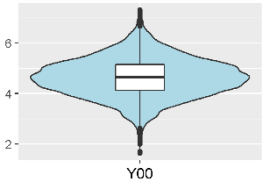
Note:
- *Negative treatment effect indicating the column potential outcome is better than the row potential outcome in reducing cJADAS value.*
- *Results of two-staged BART are reported in gray fonts. All other reported estimates are due to GPMatch method.*

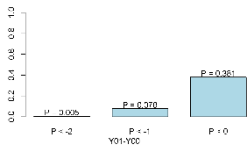
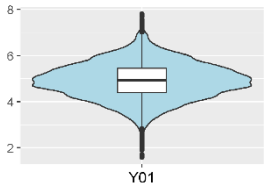
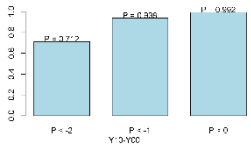
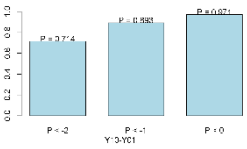
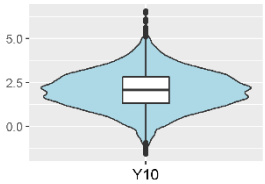
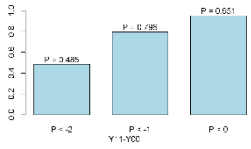
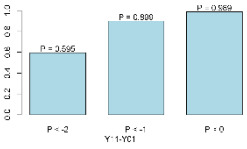
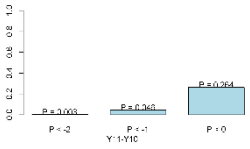
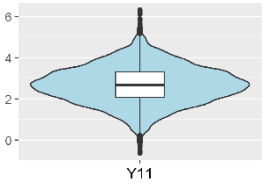
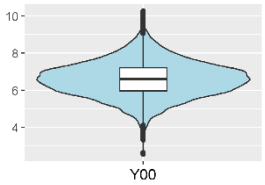
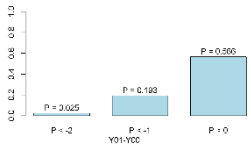
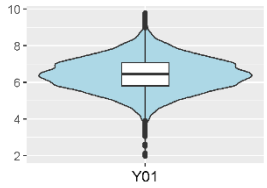
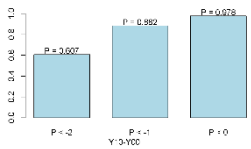
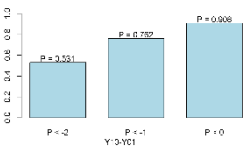
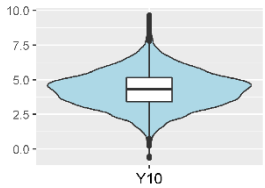
3.7. Sensitivity Analyses

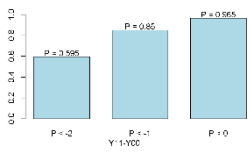
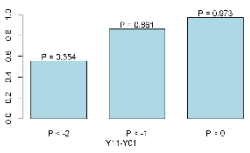
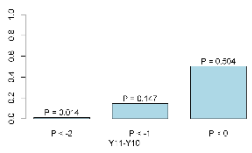
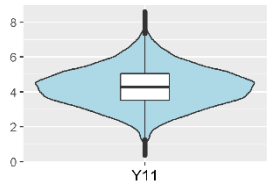
A major concern with the causal inference is the potential unmeasured confounders. In this study, we assume that important factors considered into Physician's decision in medication prescription are fully captured in the EMR data. While this may be a reasonable assumption, the patient specific determinants may also confound the medication prescription decision. To take into account of potential unmeasured confounders due to patient's quality of life, we conducted sensitivity analyses by including PedsQL, both generic and rheumatology disease specific module, into the analyses.

As a sensitivity analyses, we considered the PedsQL measures into the CER analyses. Since the PedsQL forms are filled out by patients on an annual basis for generic score and semi-annually for the Rheumatology module, which are completed at a convenient time rather a pre-specified visit. As the results, missing PedsQL measure could be an issue. Less than half (182, 45%) of patients completed PedsQL form at the baseline. Therefore, the sensitivity analyses first imputed missing PedsQL scores at the baseline, using all available demographic, clinical and patient reported measures at the baseline. Then, analyses were repeated and the comparative effectiveness results are nearly identical to primary analyses (see Table 6). The full report of the results of this sensitivity analyses are reported in the supplemental material section S4.

Table 6 Sensitivity Analyses of GPMatch & BART CER Results

	cJADAS	Y00	Y01	Y10	Y11
All Patients	Y00		0.27 (-1.48,2.02) 0.55 (-1.12, 2.21)	-2.56 (-4.57, -0.55) -1.85 (-3.83, 0.13)	-1.95 (-4.24, 0.35) -1.41 (-3.59, 0.78)

	Y01			-2.84 (-5.74, 0.07) -2.40 (-5.12, 0.33)	-2.22 (-4.14, -0.30) -1.96 (-3.85, -0.06)
	Y10				0.62 (-1.25, 2.48) 0.44 (-1.37, 2.25)
	Y11				
None responders	Y00		-0.16 (-2.03,1.71) 0.44 (-1.32, 2.19)	-2.29 (-4.45, -0.14) -1.82 (-4.12, 0.48)	-2.30 (-4.77, 0.17) -1.52 (-4.04,1.00)
	Y01			-2.13 (-5.24, 0.97) -2.26 (-5.30, 0.78)	-2.15 (-4.27, -0.03) -1.96 (-4.18, 0.26)
	Y10				-0.01 (-1.84, 1.82) 0.30 (-1.63, 2.23)

	Y11				

Note:
- *Negative treatment effect indicating the column potential outcome is better than the row potential outcome in reducing cJADAS value.*
- *Results of two-staged BART are reported in gray fonts. All other reported estimates are due to GPMatch method.*

3.8. Comparative Effectiveness of the Secondary Endpoint

There were 182 patients had PedsQL scores at the baseline, 46 and 117 patients at the 6 and 12 month follow-up. Since patients were asked to complete PedsQL generic module on an annual basis for generic module, only 9 patients had both baseline and six month scores. Given the large number of missing, comparative effectiveness analyses could only be performed for the 12 month PedsQL outcome on 1st line of treatment. The GPMatch result suggest both CTP are effective in improving PedsQL score, reporting 74.8±2.0 and 80.4±3.7 by 12 months if treated on the step-up and early combination respectively. The results of averaged treatment effect comparing the early combination vs. step-up CTP are reported in Table 7 below are summarized over the five simulated datasets. The estimated treatment effect size ranges from 1.8 to 7.7 points improvement in PedsQL generic scores, but with small sample size, none of the estimate are statistically significant.

Table 7 Averaged Treatment Effectiveness of Early Combination vs. Step-up CTP

	PedsQL Generic Total Score at 12 month			
	Est	SD	95%LL	95%UL
Direct Modeling	3.530	3.643	-3.611	10.670
Regression by PS stratification	1.979	4.595	-7.030	10.988
Regression with PS	4.021	3.729	-3.287	11.329
Regression Adjustment with spline fit PS	3.382	3.761	-3.989	10.754
PS weighted regression	1.764	2.436	-3.011	6.540
AIPTW	4.171	3.145	-1.994	10.336
Propensity score based stratification method	2.289	4.374	-6.286	10.864
IPTW	7.670	5.650	-3.409	18.749
BART	3.822	3.645	-3.323	10.968
GPMatch	5.612	4.849	-3.892	15.117

To facilitate the interpretation of the study results, we conducted additional analyses evaluating the improvement needed for achieving minimum clinical important difference (MCID) in PedsQL. Previous study has established the MCID of 4.4 in PedsQL generic score[18]. We evaluated MCID in cJADAS score utilizing all existing data on cJADAS and PedsQL measure. The results suggest a linear relationship between the cJADAS and PedQL generic total scores. For every one unit decrease in cJADAS, we expect to see 0.99 increase in PedsQL generic score. Therefore, for achieving the MCID of 4.4 points increase in PedsQL, treatment needs to achieve is 4.5 points decrease in cJADAS. This results is consistent with the previous study[19], which reported JADAS score is responsive to changes and reflect clinical meaningful changes. The study estimated 5.5 reduction in JADAS27 anchoring the ACR30 criteria.

4. CONCLUSION AND DISCUSSION

4.1. Key results

In this study, we evaluated the comparative effectiveness of the early combination treatment plan in children with newly diagnosed pcJIA, compared to the more conventional approach of step-up plan, on the clinical and quality of life outcomes at the 6 and 12 months of treatment. To the best of our knowledge, this is the first study that applies causal inference methods to evaluate comparative effectiveness of early combination vs. step-up CTP using EMR data.

The results suggest both CTP are effective in improving disease activities, reporting expected mean cJADAS score of 6.7 and 4.7 by 6 months and 4.8 and 2.4 by 12 months if treated on the step-up and early combination respectively. Early combination treatment on average produce a significant 2 points more reduction with averaged treatment effect of -1.98, 95%CL of (-3.55, -0.40) in cJADAS score by 6 months, which sustained up to 12 months. Due to the limited data available for the PedsQL generic scores recorded in the Epic database, the averaged treatment effect of 5.6 is associated with large variance, with 95%CL of (-3.9, 15.1). The study estimated expected potential outcomes of 74.8 and 80.4 by the end of 12 months, if treated on the step-up and early combination respectively.

We also estimated MCID of 4.5 points decrease in cJADAS is required for meaningful improvement in PedsQL. This results is consistent with the previous study[19], which reported JADAS score is responsive to changes and reflect clinical meaningful changes. The study estimated 5.5 reduction in JADAS27 anchoring the ACR30 criteria. Taking it together with the estimated potential outcomes and the averaged treatment effect, these results suggest both early combination and step-up CTP achieves meaningful important improvement by 6 months of treatment, with early combination achieve statistically significantly more improvement compared to the step-up treatment.

The study is carefully designed to emulate a randomized trial following CTP. The TREAT study is the only randomized trial evaluating early aggressive treatment vs. step-up approach[4], and the

study found that early aggressive treatment worked better than the step-up treatment in reducing disease activities. There are some differences in study designs worth noting comparing to the TREAT study. Here, we imposed stronger requirement on the new patient definition. Only patients diagnose with JIA for <= 6 month are considered as new patient in our study, on the other hand, TREAT study included patients within 12 months of diagnose. Following the CTP, we excluded patients with comorbid condition of IBD, celiac, and trisomy 21. Whereas, TREAT study only excluded patients with uveitis. TREAT allows patients to be treated on methotrexate before enrollment, while we request participants were naïve to DMARDs before the baseline. Despite these differences, our study lead to consistent study results, confirming clinical effectiveness of the early combination treatment.

Electronic medical record systems are widely implemented in health care organizations, capturing interactions between patients and care providers, on their examinations, treatment approaches, laboratory tests, medical imagines, and other diagnostic decision making processes, as well as patient reported outcomes. Within an established EMR system, such interactions could be tracked from the first date of diagnosis throughout the courses of disease progression and the treatment, particularly for patients with chronic conditions. Therefore, it offers invaluable data sources for evaluating the effectiveness of alternative treatment choices, understanding potential treatment heterogeneity, and subsequently guiding evidence based treatment decisions. The study demonstrates that EMR could be used for better understanding of treatment effectiveness.

Widely adopted currently, EMR is becoming increasingly sophisticated and mature. The possibility of using EMR for understanding effectiveness of treatment approaches in real world setting has sparked great enthusiasms. However, many have raised concerns over the quality of data and methodology complications for research. The issues of incomplete data, inconsistent, and sometime duplicate incoherent data are the major barriers to utilizing EMR for comparative effectiveness research. In this study, to ensure the data quality, we have taken multiple steps. We identified and utilized an existing prospective follow cohort recently completed in a subset of the same patient population as the quality assurance (QA) data. With the QA dataset, we were able to cross valid and derive rigorous data extraction algorithms. Further, during the real clinical encounters, patients follow different follow-up schedules, and gone through different treatment courses. For the purpose of this CER, rigorous data management process were taken, in order to identify for treatment courses and outcome measures mimicking to a controlled trial protocol. These data management effort help ensure the data quality, and feasibility for evaluating comparative effectiveness of the adaptive treatment effect using data collected from the real clinical encounters.

4.2. Interpretations of analyses results

Randomized controlled trial, by ensuring the internal validity, has been considered the gold standard for evaluate treatment effect. However, the RCT could compromise the external validity

due to consent process and trial logistic. Further, the RCT could still suffer from the attrition, missing data, and other complications, and the highly controlled setting may not reflect what really happened in the real world. Using data obtained from real clinical encounters, during the course of treatment, comparative effectiveness study evaluate what happens in the real world setting. Utilizing electronic health record, via rigorous study design and causal inference analyses, this study suggests similar results as observed previous in an RCT that the early combination approach is expected to lead to better disease outcome after 12 month of the treatment than step-up approach.

Different causal inference methods may report somewhat different effect size and with different level of accuracy. For getting the method right, we have conducted extensive studies comparing different methods under the most realistic setting, that is when neither of the propensity score or the outcome modeling are not correctly specified. Our investigations confirm with concerns raised by others over some widely used causal inference methods under dual miss-specified model[20,21]. The newly proposed Bayesian' GPMatch demonstrate superior performances compared with the most widely used methods[22]. Future studies may wish to further investigate the performances of GPMatch under wider scenarios.

Missing data remains to be an important challenge. Here, we took multiple missing data imputation approach for missing baseline covariates, under the assumption of missing at random. The assumption is reasonable as the missing baseline covariates are most likely due to the reasons that are unrelated with the patients' outcomes. The missing outcome, however, is likely related to the patients' disease activities/progression. It is perceivable that patients who are unsatisfied with their disease activity/progression are more likely to come back to see doctor; while those who are making good progress may not come back as frequently. Thus, the timing of follow up visit and missing repose at six month is likely related with the treatment and the outcomes. For dealing with the issue, our GPMatch model and other regression type of model included the time of follow up in the analyses. Conditional on the baseline covariates and the treatment assignment, the modeling approach acknowledge that the patients' responses will vary over the time of treatment. The missing response at the six month follow-up are modeled based on the functional relationship between the time of treatment and the responses utilizing the data obtained from the patients with the six month follow-up. Here, the AIPTW and PS sub classification approaches did not consider modeling treatment effect as a function of time, which could contribute to the somewhat lower treatment effects produced by both methods. We took the Bayesian HCMM-LD approach for multiple imputation of the missing baseline covariates, for preserving the joint distributions among all covariates. As the sensitivity analyses, MICE is also used, and the results are nearly identical. (The MICE results are reported in the supplemental material section S3.)

The cJADAS is a bounded summary score. The BART approach did not accounted for such bounded nature, thus may under estimated treatment effect. In addition, BART is a highly

flexible modeling approach, considering outcome modeling only. It does not include regularization or prior knowledge to account for the treatment by indication selection bias. Therefore, it could be vulnerable to the potential model miss-specification, and suffer when there is a lack of overlapping in covariate space, which may also contributed to the different result reported by BART. The GPMatch and other regression type of approach considered the bounded outcome by Tobit regression. Other than specifically addressing baseline confounding by formulating GP prior as a matching tool, the GPMatch is able to address lack of overlapping in covariate space by down weighting those data points presenting little or no similarity. For the CER of adaptive 2-staged CTP, BART and GPMatch reported similar results, with BART showing somewhat weaker estimate of effect sizes and bigger variance estimates.

4.3. Limitations and Generalizability

This study compared the early combination vs. step-up CTP. Biological only is another recommended CTP, but it was not considered in this study for a number of reasons. Only 71 were treatment on the Biological only CTP. Half of the patients treated on the biological only CTP are ERA patients, they are more likely to be male, RF-, ANA- and more likely having HLA-B27 present. Thus, it could represent a different group of patients than those being treated on the step-up and the early combination CTPs. (Detailed baseline covariate distributions for patients treated on biological only CTP are presented in the supplemental material section S1.) For the considerations of equipoise and sample size, this study did not including the biological only CTP. Future CER studies on the biological only CTP only may consider revising the study inclusion and exclusion criteria to ensure equipoise of the comparison, when evaluating the effect biological only CTP vs. other treatment approaches. Patients with systematic subtype are treated different, a different CTP were recommended. The current study was focused on evaluate CTP for the pcJIA. Future study could apply the same method directly to the systematic JIA patients.

The study has some major methodology limitations. First, the treatment were determined by medication prescription recorded in electronic medical records. Medication dispense and adherence are not available. The treatment effect may vary by medication dose, formulation, and route. Such information could not be considered in the current study, due to the limitation of the electronic medical records. Second, patients do not follow the predetermined schedule of follow-up, making it challenging to evaluate the CER at a given time points. This study developed an algorithm to define the timing of follow up visit for each patients, which is determined based on their treatment courses. Further the outcome measures, given the nature of EMR could be missing, and are assigned utilizing the records from nearby visits within one month window. The analyses results might be sensitive the specification of the time window. Sensitivity analyses may consider varying time window, and other decisions rule used in the shaping EMR data for research purpose. At last, while we conducted sensitivity analyses evaluating the effect of not accounting for quality of life measures at baseline as a treatment-by-indication confounder, it is

perceivable other unmeasured confounders may exists. While the potential unmeasured confounder remain to be a critical challenge, the sensitive analyses results suggest that inclusion of additional covariates to the large set of covariates considered in this causal inference analyses results to nearly identical results. The study did not identify heterogeneous treatment effectiveness by JIA subtype, or the baseline cJADAS value.

The study is limited to a single medical center study. Patient from different center may represent somewhat different patient population in their demographics and disease subtypes. Clinicians from different center may also engage different practices in treatment assignment. Our study did not find significant subgroup treatment effect, suggesting the results is generalizable. On the other hand, however, MD global and patient reported well-being could subject to individual and center variations, thus the effect size may different by clinical center. Future study should consider multiple center study.

In conclusion, both early combination and step up CTP are effective in reducing disease activities, where early combination CTP is more effective leading to better cJADAS score by 6 and 12 months of treatment after diagnosis. Future studies are needed to investigate comparative effectiveness of CTP on the longer term of outcomes in cJADAS, including inactive disease and health related quality of life outcomes.

ACKNOWLEDGEMENT

Funding support from the Patient Centered Outcome Research Institute (PCORI ME-1408-19894; Patient Center Adaptive Treatment Strategy – PCATS with Bayesian's Causal Inference)

REFERENCES

1. Thierry S, Fautrel B, Lemelle I, Guillemin F. Prevalence and incidence of juvenile idiopathic arthritis: A systematic review. Joint Bone Spine. 2014 Mar 1;81(2):112–7.
2. Beukelman T, Kimura Y, Ilowite NT, Mieszkalski K, Natter MD, Burrell G, et al. The new Childhood Arthritis and Rheumatology Research Alliance (CARRA) registry: design, rationale, and characteristics of patients enrolled in the first 12 months. Pediatric Rheumatology. 2017 Dec 17;15(1):30.
3. Donahue KE, Gartlehner G, Schulman ER, Jonas B, Coker-Schwimmer E, Patel SV, Weber RP, Lohr KN, Bann C VM. Drug Therapy for Early Rheumatoid Arthritis: A Systematic Review Update. Rockville; 2018.
4. Wallace CA, Giannini EH, Spalding SJ, Hashkes PJ, O'Neil KM, Zeft AS, et al. Trial of Early Aggressive Therapy in Polyarticular Juvenile Idiopathic Arthritis. Arthritis Rheum. 2012 Jun;64(6).
5. Nigrovic PA, Mannion M, Prince FHM, Zeft A, Rabinovich CE, van Rossum MAJ, et al. Anakinra as first-line disease-modifying therapy in systemic juvenile idiopathic arthritis: Report of forty-six patients from an international multicenter series. Arthritis & Rheumatism. 2011 Feb 1;63(2):545–55.
6. Vastert SJ, De Jager W, Noordman BJ, Holzinger D, Kuis W, Prakken BJ, et al. Effectiveness of First-Line Treatment With Recombinant Interleukin-1 Receptor Antagonist in Steroid-Naive Patients With New-Onset Systemic Juvenile Idiopathic Arthritis Results of a Prospective Cohort Study. ARTHRITIS & RHEUMATOLOGY. 2014;66(4):1034–43.
7. Ravelli A, Consolaro A, Horneff G, Laxer RM, Lovell DJ, Wulffraat NM, et al. Treating juvenile idiopathic arthritis to target: recommendations of an international task force. Annals of the rheumatic diseases. 2018 Apr 11;2018–213030.
8. Stoll ML, Cron RQ. Treatment of juvenile idiopathic arthritis: a revolution in care. Pediatric rheumatology online journal. 2014;12:13.
9. van Rossum MAJ, van Soesbergen RM, Boers M, Zwinderman AH, Fiselier TJW, Franssen MJAM, et al. Long-term outcome of juvenile idiopathic arthritis following a placebo-controlled trial: sustained benefits of early sulfasalazine treatment. Annals of the rheumatic diseases. 2007 Nov;66(11):1518–24.
10. Ringold S, Weiss PF, Colbert RA, DeWitt EM, Lee T, Onel K, et al. Childhood Arthritis and Rheumatology Research Alliance consensus treatment plans for new-onset polyarticular juvenile idiopathic arthritis. Arthritis Care & Research. 2014 Jul;66(7):1063–72.
11. PCORI. PCORI METHODOLOGY STANDARDS. 2018.

12. Seid M, Huang B, Niehaus S, Brunner HI, Lovell DJ. Determinants of health-related quality of life in children newly diagnosed with juvenile idiopathic arthritis. Arthritis Care & Research. 2014 Feb;66(2):263–9.

13. Consolaro A, Giancane G, Schiappapietra B, Davì S, Calandra S, Lanni S, et al. Clinical outcome measures in juvenile idiopathic arthritis. Pediatric Rheumatology. 2016 Dec 18;14(1):23.

14. Imai K, Ratkovic M. Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014;76(1):243–63.

15. Chib S. Bayes inference in the Tobit censored regression model. Journal of Econometrics. 1992 Jan 1;51(1–2):79–99.

16. Murray JS, Reiter JP. Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models With Local Dependence. Journal of the American Statistical Association. 2016 Oct 4;111(516):1466–79.

17. Rubin DB. Multiple imputation for nonresponse in surveys. Wiley-Interscience; 2004. 287 p.

18. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL™* 4.0 as a Pediatric Population Health Measure: Feasibility, Reliability, and Validity. Ambulatory Pediatrics. 2003 Nov 1;3(6):329–41.

19. Bulatovic Calasan M, de Vries LD, Vastert SJ, Heijstek MW, Wulffraat NM. Interpretation of the Juvenile Arthritis Disease Activity Score: responsiveness, clinically important differences and levels of disease activity in prospective cohorts of patients with juvenile idiopathic arthritis. Rheumatology. 2014 Feb 1;53(2):307–12.

20. Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical Science. 2007;22(4):523–39.

21. Gutman R, Rubin DB. Estimation of causal effects of binary treatments in unconfounded studies with one continuous covariate: https://doi.org/101177/0962280215570722. 2015 Feb 24;

22. Huang, Bin; Chen C et al. Estimating Causal Treatment Effect with Bayesian's Marginal Structural Model using Gaussian Process Prior as a matching tool. In Progress.