APPENDICES

Appendix A. PCATS Study Protocol

PATIENT CENTERED ADAPTIVE TREATMENT STRATEGIES (PCATS) USING BAYESIAN CAUSAL INFERENCE

Sponsor: Investigator-Initiated

Study Sites: Cincinnati Children's Hospital Medical Center (CCHMC)

Principal Investigator: Bin Huang, PhD

Co-Investigators: Hermine Brunner, MD, MSc
Esi Morgan Dewitt, MD, MSCE
Michael Seid, PhD

Version: v.2.0
Date: December 12, 2018

STATEMENT OF COMPLIANCE

The study has been conducted in accordance with the International Conference on Harmonization guidelines for Good Clinical Practice (ICH E6) and the Code of Federal Regulations on the Protection of Human Subjects (45 CFR Part 46). All personnel involved in the conduct of this study have completed human subject protection training.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ATS	Adaptive Treatment Strategy
AIPTW	Augment Inverse Probability of Treatment Weighting
BART	Bayesian Additive Regression Tree
BF	Bayes Factor
bDMARD	Biologic Disease-Modifying Antiarheumatic Drug
CBPS	Covariate Balancing Propensity Score
CCHMC	Cincinnati Children's Hospital Medical Center
CER	Comparative Effectiveness Research
CHAQ	Childhood Health Assessment Questionnaire
cJADAS	Clinical Juvenile Disease Activity Score
CRF	Case Report Form
CTP	Consensus Treatment Plan
DBS	Double Balancing Score
DMARD	Disease-Modifying Antiarheumatic Drug
DR	Double Robust Method
ePAS	Electronic Protocols Application Software
GPMatch	Gaussian Process prior Matching
HIPAA	Health Insurance Portability and Accountability Act
HTE	Heterogeneous Treatment Effects
HRQoL	Health Related Quality Of Life
HCMM-LD	Hierarchically Coupled Mixture Model with Local Dependence
IBD	Inflammatory Bowel Disease
ILAR	International League Of Associations For Rheumatology
IPW	Inverse Probability Weighted Estimate
IRB	Institutional Review Board
JIA	Juvenile Idiopathic Arthritis
JIA-QoL	Juvenile Idiopathic Arthritis-Quality of Life
MD	Doctor of Medicine
MD	Mahalanobis distance
MICE	Multiple Imputation Chained Equation
MSM	Marginal Structural Model

nbDMARD	Non-biologic Disease-Modifying Antiarheumatic Drug
nb+bDMARD	Non-biologic and Biologic Disease-Modifying Antiarheumatic Drug
NIH	National Institutes of Health
NSAID	Nonsteroidal Anti-Inflammatory Drug
PCATS	Patient Centered Adaptive Treatment Strategies
pcJIA	Polyarticular Course Juvenile Idiopathic Arthritis
PCORI	Patient-Centered Outcomes Research Institute
PCORnet	Patient-Centered Clinical Research Network
PedsQL	Pediatric Quality Of Life Inventory
PHI	Protective Health Information
PPRN	Patient Powered Research Network
PR-COIN	Pediatric Rheumatology Care And Outcomes Improvement Network
PrS	Propensity Score
SAP	Stakeholder Advisory Panel
SAS	Statistical Analysis System
SMART	Sequential Multistage Adaptive Randomized Trial
SE	Squared-exponential
SUTVA	Stable Unit Treatment Value Assumption
SNM	Structural Nested Model
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
STTA	Single-Time Treatment Assignment

PROTOCOL SUMMARY

Title:	Patient centered adaptive treatment strategies (PCATS) using Bayesian causal inference
Study Sites	Cincinnati Children's Hospital Medical Center (CCHMC)
Study Design	Observational study with real world data collected during routine clinical care
Patients Population	Children who were 19 years of age or younger, newly diagnosed with polyarticular course of juvenile idiopathic arthritis, and treated on biologic and/or non-biologic Disease Modify anti Rheumatic Drugs (DMARDs)
Objective	*Primary*: To develop, refine and disseminate Bayesian causal inference methods for evaluating clinical effectiveness and for informing better PCATS *Secondary*: To evaluate clinical effectiveness of the newly recommended consensus treatment plans (CTPs) for pcJIA patients via conducting and analyzing a large inception cohort patient observational study.
Project duration	3 years, September 2015 - November 2018
Duration of Data Process Phase	2.5 years, February, 2016 - August, 2018
Minimum and Maximum length of follow up for any participants	1 month ~ 10 years
Sponsor	Patient Centered Outcome Research Institute (ME-1408-19894)
Principal Investigator	Bin Huang, PhD
Co-Investigators	Hermine Brunner, MD, MSc, Esi Morgan Dewitt, MD, MSCE Michael Seid, PhD

KEY ROLES AND CONTACT INFORMATION

Principal Investigator:	Bin Huang, PhD
PCORI Program Official:	Emily Evans
Co-Investigators:	Hermine Brunner, MD, MSc Esi Morgan Dewitt, MD, MSCE Michael Seid, PhD
Institution:	Cincinnati Children's Hospital Medical Center (CCHMC) *3333 Burnet Avenue, Cincinnati, Ohio 45229*
Other Key Personnel:	Database Manager: Tingting Qiu, MPH; Yin Zhang, MS Epic Specialist: Janet Zahner, Senior Database Administrator Project Manager: Anastasia Woeste, MA, CCRP Software System Architecture Specialist: Michal Kouril, PhD Statistician & Programmer: Chen Chen, PhD
Stakeholder Personnel:	Timothy Beukelman, MD (Patient Care Provider) Stafanie Gray (Parent) Jeff Guo, PhD (Health Policy and Outcome, Chair) Melanie Kohlheim (Parent) Daniel Lovell, MD (Patient Care Provider)
Past Team Members:	Michele Adams, PhD (Project and Database Manager) Jinzhou Liu, PhD (Graduate Student Research Assistant) Alivia Neace (Student Assistant) Jesse Pratt, MS (PR-COIN Database Manager) Huaiyu Zang, PhD student (Graduate Student Research Assistant)

1. INTRODUCTION

1.1. Background

Juvenile Idiopathic Arthritis

Juvenile idiopathic arthritis (JIA), a chronic inflammatory disease, is one of the most common autoimmune diseases and a major cause of childhood disability. The incident rate of JIA is approximately 10 per 100,000 for girls and 5.7 per 100,000 for boys[1]. JIA is a heterogeneous group of diseases, so the treatment plans are various across JIA subtypes. Systemic JIA requires the most distinctive treatment approaches compared to other types of JIA. Non-systemic JIA includes polyarticular JIA, oligoarticular JIA, psoriatic JIA, enthesitis-related JIA, and undifferentiated JIA. These are often refractory to treatment, and their courses alternate between relapse and remission. In the past two decades, various treatment options have been made available for JIA, such as non-biologic disease-modifying antirheumatic drugs (nbDMARD) and biologic DMARDs (bDMARD). These treatment options have the potential to induce disease quiescence in nearly half of patients as long as treatment is continued. However, it is unknown at the time of initial treatment which medication or medication combinations are the most effective to induce remission for a given individual. Additionally, for a patient who does not respond to a previous treatment, the next best option is often unknown. Unsurprisingly, such poorly guided treatment strategies produce inferior patient-reported outcomes, and despite advanced medical treatment, half of the patients experience a lower health-related quality of life (HRQoL)[2]. Therefore, another goal of informed medical decision making is to help optimize patients' HRQoL. In 2014, a panel of JIA experts developed three consensus treatment plans (CTPs) for polyarticular course JIA (pcJIA)[3] and called for evaluating these adaptive treatment plans through observational study. The CTPs for the systemic JIA are different than polyarticular course JIA. Polyarticular course JIA is referred to all JIA with arthritis in >4 joints, excluding systemic JIA[3].

In JIA research field, development of the CTPs represents the first step toward identifying optimal adaptive treatment strategies for children with JIA. In routine clinical care, treatments are adapted over time to patient's responses. This is particularly important for patients suffering from chronic illness. However, statistics causal inference methods are limited to evaluating the clinical effectiveness of adaptive treatment. This is a critical limitation. Lack of strong causal inference methods underscores the gap of evidence in many chronic or prolonged disease conditions. Motivated by the needs of filling the gap in caring for children with JIA, this method development is designed to improve methodology available for evaluating the clinical effectiveness of ATS.

Another development in the JIA field is the availability of real world data. The Patient-Centered Clinical Research Network (PCORnet) has been established and is actively growing in membership. The Pediatric Rheumatology Care and Outcomes Improvement Network (PR-COIN), which is part of a patient powered research network (PPRN), dedicated to children with a rheumatic disease such as JIA and Lupus. Similar efforts are ongoing for many other disease conditions. The electronic medical records system are implemented broadly in US health care institutes. These development bring us closer to transforming how we deliver the best possible care to patients with JIA and other diseases. However, the answers to comparative effectiveness of adaptive time-varying treatment strategies remain unclear. Thus, there are urgent needs in developing advanced analytic methods that allow for the best evaluation of the clinical

effectiveness of adaptive treatment strategies, and for identifying the optimal adaptive treatment assignment.

Statistical Methods

Prior to reviewing the methodology background, it is necessary to distinguish a few terminologies. For time-varying treatment, we must distinguish the dynamic treatment regime from the static treatment regime. The dynamic treatment regime is determined at the point-of-care during the course of treatment; while the static treatment regime is pre-determined ahead of time. For example, a randomized trial may randomize a patient to a predetermined treatment sequence of treatment A followed by treatment B. Although a patient may receive different treatments at different times, the decision was predetermined at the time of randomization, thus it is static, not dynamic. A randomized clinical trial is an example of a single-time treatment assignment (STTA), which is the common setting considered in analytic methods. On the other hand, adaptive treatment strategy (ATS) stresses the treatment plan is adaptive to both treatment history and the patient's responses to the previous treatment, which is routinely used clinical care. In this proposed study, we focus on developing analytic methods for an optimal patient centered adaptive treatment strategy (PCATS).

The current method development addressing ATS. The development in study design has primarily been the sequential multistage adaptive randomized trial (SMART) design[4,5]. Data analysis methods can be classified into two large categories. The first includes time-varying inverse probability weighted estimate (IPW;[6]) the marginal structural model (MSM;[6–8]) and the structural nested model (SNM;[8,9]). The difference between the MSM and the SNM lies in the different causal effects of interest. The MSM estimates average causal effect between two ATS; while the SNM answers the causal question, "What if we stopped treatment at time t as opposed to at time $t-1$?". Daniel et al.[10] reviewed the idea behind the MSM, SNM and IPW methods while considering a simple 2-staged ATS case. The second class of analytic approach is reinforcement learning, represented by Q-learning [11,12] and A-learning [13,14]. Q-learning and A-learning are designed to identify an optimal decision rule on the next ATS. The issue with Q-learning is that the asymptotic distribution of the treatment effect estimator does not have uniform coverage over the parameter space which creates the issue of non-regularity for utilizing standard statistical inference methods[14,15]. Neither A- or Q-learning methods may fail when model is miss-specified. Latest reviews of these methods are found in Moodie, Chakraborty et. al.[16] and Wallace & Moodie[17]. In summary, the current existing analytical methods for ATS are limited: 1) these methods are vulnerable to model misspecification, which could seriously undermine the validity of these methods[18–20], 2) the performances of these existing methods have not been carefully evaluated for ATS under more realistic setting where the true functional forms for the outcome and treatment models are unknown, and 3) while some tools for SNM have been offered in R[21], there is a lack of robust, general and easy to use analytic tools.

Statistics causal inference methods that are commonly used for comparative effectiveness research (CER) assume STTA. Under the STTA, let's assume we could treat a patient twice,

once under the active treatment ($a = 1$) and the other under the control treatment ($a = 0$). In denoting the corresponding outcomes under each treatment by $Y_i(a)$, for $a = 0,1$, the causal effect would be simply obtained by $Y_i(1) - Y_i(0)$. However, a patient can only take a treatment once at the time of diagnosis or at a given time point, his or her outcomes under the treatment not taken is unknown. The "fundamental challenge" in causal statistics inference is that one of the potential outcome is always unobserved.

During the routine clinical care, treatments are assigned for deliberate reasons, thus factors that determine who takes the treatment are often related to the outcome. This is the treatment-by-indication-bias. Under the STTA assumption, causal inference methods such as propensity score (PrS) and prognostic score (PgS) methods have been adopted for removing the treatment-by-indication-bias. The validity of causal inference methods rests on three fundamental causal assumptions[22].

- Stable Unit Treatment Value Assumption (SUTVA): the potential outcomes of one experiment unit do not change despite how the treatment was assigned, nor related to the potential outcomes of the other experiment unit;
- Strong Ignorable Treatment Assignment Assumption (SIA): the potential outcomes and the treatment assignment (A) are independent conditional on the observed covariates, i.e. $(Y^{(0)}, Y^{(1)}) \perp A|(X, V)$, where X denotes prognostic factors and V denotes confounding factors. The assumption requires there is no unmeasured confounder.
- Positivity Assumption: this assumption ensures every unit has non zero probability of being assigned into either one of the treatment arms, i.e. $0 < Pr(A_i|X_i, V_i) < 1$.

CER study with real world data must confront challenges in causal inferences, including model misspecification, uncertainty in the estimation of PrS and PgS, complexed type of treatment such as time-varying adaptive treatment, missing data, and unmeasured confounders. These challenges could seriously threaten the validity of causal inference results. For guard against model misspecification, the double balancing score (DBS) and double robust method (DR) have been proposed[23,24] =. However these methods may suffer from poor statistics operational characteristics, under the setting of dual miss-specification, subsequently, it could produce more erroneous estimates with increased sample size[25].

The complexity of ATS mandates advanced causal inference methods. Because of the adaptive assignment process, patients responding better (or worse) are likely to end up on the same ATS, thus are confounded not only by the baseline covariate but also by time-varying covariates. Since the missing potential responses increase exponentially with the increased number of decision time points and treatment choices, the "fundamental challenge" of statistical causal inference is heightened in an ATS setting and demands a more rigorous approach. Few causal inference methods have been proposed for time-varying treatment. Longitudinal IPW has been the most commonly used method[10,26]. The longitudinal PrS[27] method was proposed but does not allow for the evolving covariate history to inform the next treatment decision, and therefore is not adaptive. Bayesian framework is particularly suitable for comparative effectiveness research, for its ability to incorporate model uncertainty and prior knowledge into the updating of the analyses results. However, no Bayesian causal inference methods have been proposed for ATS.

1.2. Rationale

Motivated by the gap in evidence for treating JIA, the PCATS study recognizes limitations in the existing statistical causal inference methods used in comparative effectiveness researches when analyzing real world data, particularly within the Bayesian's framework. Therefore, the study aims to develop a rigorous Bayesian casual inference method that is robust to model misspecification and can provide an accurate estimate of comparative effectiveness of ATS such as the consensus treatment plan for JIA patients.

1.3. Choice of Comparators

The consensus treatment plans (CTPs) for children with newly diagnosed polyarticular course JIA (pcJIA) involve three time-varying adaptive treatment plans. The step-up plan starts with non-biologic DMARDs (nbDMARDs) as the first line treatment, then step-up with more aggressive approaches, including change to a different nbDMARDs or initiate biologic DMARDs if needed. The early combination treatment plan starts with the combination of non-biologic and biologic DMARDS (nb+bDMARDs) first, then taper down or adjust the treatment depending on how the disease responses. The biologic only CPT starts with the biologic DMARDs (bDMARDs) only therapy first, then consider adding or adjust DMARDs based on the disease responses. The step-up CTP is a more conservative and most commonly adopted DMARD treatment approach. It is expected that patients who receive the early combination plan are likely to find matching patients who receive the step-up plan. Thus, the comparator is the step-up CTP group.

1.4. Objectives

The long term objective is to improve the health outcomes and experiences in patients with chronic illness by enabling evidence-based shared decision-making tools at the point-of-care through rigorous analytic method development, such that: 1) at the initiation of the treatment, it can provide analyses and recommendations of optimal treatment based on the patient specified treatment goals; and 2) during the course of the treatment, it can incorporate the patient's responses/experiences with previous treatment, patient specified treatment goals, and provide analyses and recommendations for the next optimal treatment choice.

1.4.1. Specific Aims and Hypotheses

Recognizing the needs for methodology development and the needs to fill in the gap of knowledge in understanding effectiveness of CTPs for pcJIA patients, we proposed the patient centered adaptive treatment strategy (PCATS) study. The study has two specific aims:

Aim1. To develop, refine and disseminate Bayesian causal inference methods for evaluating clinical effectiveness and for informing better PCATS:

H1.1. To develop and refine Bayesian causal inference methods for PCATS that is general, robust and efficient for analyzing observational data, and provides better ways of addressing model uncertainty.

H1.2. To disseminate the study results with the development of a publicly accessible "PCATS" R shinny package, a graphic user interface interactive online application, which allows

for easy application of the newly developed methods using user supplied data in many other settings and disease conditions.

Aim2. To evaluate clinical effectiveness of the newly recommended ATS for pcJIA patients via conducting and analyzing a large new patient registry study in collaboration with PR-COIN, a participant in a PCORI funded PCORnet.

H2.1. The early combination plan is more effective and leads to better clinical and HRQoL outcomes than the step-up plan for the pcJIA patient population; the treatment effect may differ by patients disease subtypes.

H2.2 In pcJIA patients who fail to respond to non-biologic DMARD treatment, adding biologic DMARD treatment as 2nd line approach is more effective and leads to better clinical and HRQoL outcomes than non-escalated approach. The effect may differ by patients' disease subtypes.

1.5. Study Design

The sequential multi-stage adaptive randomized trial (SMART) is a design that adapts treatment decisions over time. Ideally, without time, resource and logistic constrain, we may wish to evaluate the effectiveness of early combination CTP vs. the step-up CTP by conducting a SMART as illustrated in the schematic presentation in Figure 1.

As presented in Figure 1, newly diagnosed patients have the option to be allocated into one of the two treatment arms, the nbDMARD (comparator arm, initial stage of the step-up CTP) and the nb+bDMARD (the experiment arm, initial stage of the early aggressive arm). The disease response will be evaluated after 6 months of treatment, and the decision of the 2nd line treatment will be evaluated based on the patients' response to the initial 1st line treatment assignment. If the patient is doing well, then the patient will continue on the same treatment. If not, then the treatment will be adjusted. The CTP recommends that a patient is considered doing well if his or her physician's global evaluation of disease activity is 2 or less. Following another six months of treatment, the study endpoints will be evaluated at 12 months of the follow-up visit.

Utilizing existing data observed from real clinical encounters, by applying causal treatment methods, PCATS study is designed to emulate the SMART design for evaluating the effectiveness of CTP treatment.

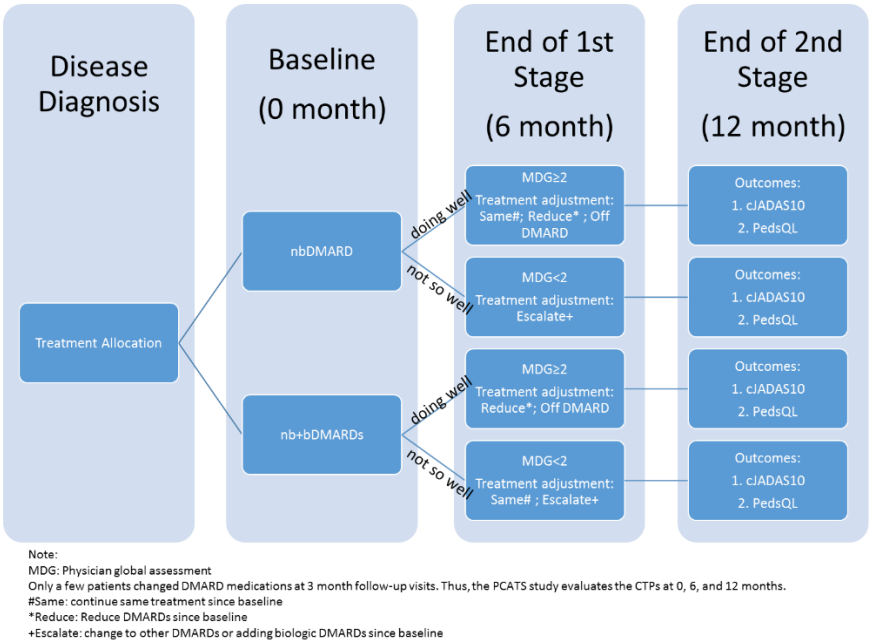


Figure 1 Emulated Sequential Multi-stage Adaptive Randomized Trial (SMART) Design

2. METHODS: Participants, Interventions, Outcomes

2.1. Study Setting

This statistical methods development and CER study will analyze data collected in three data sources:

1. Pediatric Rheumatology Care and Outcomes Network (PR-COIN) registry, a multi-center patient centered network for children with JIA;
2. CCHM JIA patient registry at the division of pediatric rheumatology embedded as part of the CCHMC Epic™ electronic medical record system; and
3. The JIA-QoL research cohort.

The majority of participants of JIA-QoL cohort study are also part of the CCHMC JIA patient registry. Thus, the JIA-QoL data could be used as external validation samples for data quality control. Table 1 provides an overview of the three data sources. Details related to these three data sources please reference Section 3.1 and Section 3.2.

Table 1 Overview of Data Sources

Data Sources	Enrollment Period	# of Participating Centers	# of JIA Patients in Databases
PR-COIN Registry	2011-2016	15 (including CCHMC)	4681
CCHMC JIA Cohort (Epic™ EMR system)	2009-2017	1 (CCHMC)	1750
CCHMC JIA-QoL Cohort	2008-2011	1 (CCHMC)	220

2.2. Eligibility Criteria

The study inclusion and exclusion criteria (Table 2) and data elements (Table 4) are designed closely following the CTP paper published by the Childhood Arthritis and Rheumatology Research Alliance (CARRA)[3].

Table 2 Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
1) Age ≤ 19 years at baseline;	1) Systematic JIA patient according to the ILAR code[*]
2) Diagnosed with polyarticular course of juvenile idiopathic arthritis (pcJIA) in accordance with the operational definition of pcJIA presented in Table 2 of CTP paper[3].	2) Patients with comorbid diagnoses of IBD, celiac disease, trisomy 21[**]
3) Diagnosed with pcJIA no more than 6 months at the first clinical encounter captured in the database	
4) Taken DMARDs no more than 9 months after diagnosed with pcJIA	

Note:
[*] *The CARRA operational case definition of polyarticular course juvenile idiopathic arthritis (pcJIA) defines pcJIA patients may have any types of JIA (e.g. RF positive polyarticular JIA, RF negative polyarticular JIA, and extended oligoarticular JIA) except systemic JIA*[3].
[**] *This exclusion criteria only applies to the Epic data. The PR-COIN registry does not collect comorbidities on IBD, celiac disease, trisomy 21 before May 2016.*

2.3. Treatment

In 2014, a panel of JIA experts developed three consensus treatment plans (CTPs) for pcJIA[3], with intentions to standardize treatment approach and improve clinic outcomes. The step-up plan starts with non-biologic DMARDs (nbDMARDs) first, then step-up with more aggressive approaches, including change to a different nbDMARDs or initiate biologic DMARDs if needed. The early combination treatment plan starts with the combination of non-biologic and biologic DMARDS (nb+bDMARDs) first, then taper down or adjust the treatment depending on how the disease responses. The biologic only CPT starts with the biologic DMARDs (bDMARDs) only therapy first, then consider adding or adjust DMARDs based on the disease responses.

PCATS study will collect data retrospectively from the electronic health records and registry studies where data from the routing clinical care are recorded. In order to identify the CTP treatment patterns, all concurrent medication prescriptions recorded in the Epic, including medication names, dosage, route, start and end dates, for the eligible patients during the study period will be extracted for each patient clinical encounters. The medication prescriptions will be

classified into biological DMARDs and non-biological DMARDs based on the recorded prescription names (Table 3). If the patient received both biological and none biological DMARDs within 2 months, then we consider the patient receiving a combination of non-biological and biological DMARDs (nb+bDMARDs), and allocate the patient into early combination CTP. If the patient initiated on nbDMARD and not initiated bDMARD within 2 months following the initiation of nbDMARD, then the patient is allocated into the step-up CTP For both step-up and early combination CTP, the starting date of the 1st DMARD is marked as the beginning of the treatment course. Following the first treatment course, if medication prescriptions are in the same DMARDs class over adjacent clinical encounters, then they are considered into the same treatment course. A new treatment course starts when any changes in the DMARD class at any given clinical encounter.

Table 3 Disease-Modifying Antirheumatic Drugs List

Medication Categories	Simple Generic Medications Record Name
Non-biologic DMARDs	Azathioprine, Cyclosporine, Hydroxychloroquine Sulfate, Lenalidomide, Methotrexate, Sulfasalazine, Minocycline, Cyclophosphamide, Tofacitinib Citrate*, Apremilast
Biologic DMARDs	Abatacept, Adalimumab, Anakinra, Canakinumab, Certolizumab Pegol,Etanercept, Efalizumab, Golimumab, Infliximab, Rilonacept, Rituximab, Tocilizumab, Ustekinumab

*Note: *Tofacitinib Citrate is in the small molecule pharmacologic class.*

2.4. Outcomes

The primary outcome is the clinical Juvenile Arthritis Disease Activity Score (cJADAS-10) at the 6 months and 12 months of follow-up visits. The cJADAS is a widely adopted clinical outcomes measures in patients care[28,29]. The cJADAS-10 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[29]. These three core measures reflect different prospective of disease progression, which are evaluated routinely during the clinical encounters. The secondary outcome is PedsQL generic total score[30]. At the CCHMC, patients or their parent also reported their health related quality of life by filling out the PedsQL generic module on an annual basis. Both of the cJADAS and PedsQL scores are bounded scores. cJADAS is bounded between 0 and 30, with a higher score indicating more disease severity. PedsQL scores are bounded between 0 and 100, with a higher score indicating better quality of life.

2.5. Participant Timeline

PCATS study starts on September 2015 and ends on November 2018. The study retrospectively collects data on patients who visited CCHMC from 2009-2017 or enrolled in the PR-COIN registry from 2011-2016. The baseline visit is defined as the date when patient initiated on DMARDs prescription after diagnosis. The follow-up visits are determined relative to the index

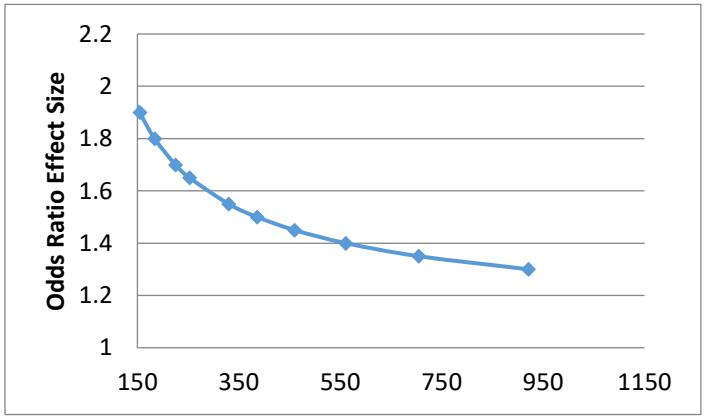
visit. The follow-up visits are identified from the clinical encounters that fall within the specified time window. The timing of the 3 months follow up visits may vary by patients, determined according to the ending date of the first treatment course. If the first medication course ended within the 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 visit is 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 are determined similarly. The time duration after the index visit is calculated for each patient. The asymmetric -1 and +2 month window is used to accommodate the possible delay in patients taking up medication after given prescription, as well as potential legged treatment effect.

2.6. Sample Size

The study sample size is established based on the effect size reported in TREAT study [31]. TREAT is an NIH funded multicenter, randomized, and double blinded, placebo controlled clinical trial. The study enrolled 85 pJIA patients from 15 clinical sites with newly onset of disease; 42 patients were randomized to the early aggressive therapy arm, i.e. methotrexate (MTX) plus etanercept and prednisolone, and 43 patients were randomized to the control arm treated with MTX plus etanercept placebo and prednisolone placebo. The characteristics of the TREAT patients were reported in Table 1 of TREAT study [31]. Patient participating TREAT study may not reflect the general patient population. For this reason, we compared the patient baseline characteristics reported in the QoL study with those reported in TREAT study.

This comparison suggests that at CCHMC, participants from both studies had similar demographics (age, gender and race). However, participants of JIA-QoL study present much fewer disease activities than the participants in TREAT study. In the CCHMC JIA-QoL study, the mean number of active joint counts (mean±SD: 5.8±7.3 vs. 21.9±13.3), number of joints with limited motion (4.6±6.4 vs. 15±12.5), CHAQ score (0.6±0.7 vs. 1.2±0.7), patient/parent reported assessment of well-being (2.7±5.5 vs. 5.3±2.5), and physician assessment of disease activity score (3.3±2.6 vs. 6.9±1.8) were less than those in TREAT study. The differences could be due to the inherent differences between the patients enrolled in a randomized controlled trial vs. the general patient population. In TREAT study, patients are referred by their treating physicians and motivated to participate in the experiment; these patients are likely being more active in their disease. While the participants of the CCHMC JIA-QoL study are representing a general clinical population in day to day clinical care. In addition, TREAT study excluded any patients who are not polyarticular subtype of JIA, while CCHMC JIA-QoL study included patients from oligoarticular and other none systematic subtypes of JIA.

Out of the 104 participants of JIA-Qol study who are meeting the inclusion and exclusion criteria of the PCATS study and finished 12 months follow up, 18 (17.3%) achieved ACR70 by the 12-month visit. Based on the medication information at the baseline, we found 70 patients took non-biologic DMARD therapy, and 50 patients took combinations of non-biologic and biologic DMARD. By 12 month, 12% (7/58) of the patients treated by non-biologic DMARD vs. 24% (11/46) patients treated by combination therapy achieved ACR70. The corresponding OR is estimated at 2.29 (95%CI of 0.81- 6.48). Given the nature of the observational study, the two groups are different in their baseline disease characteristics. To correct for the confounding by indication bias, we did a quick propensity score analyses. We applied both inverse propensity weighting (IPW) analyses and propensity score matching approaches. The IPW approach estimated OR of 2.95 with 95% CI of (0.85, 10.19). The matching approach estimated OR of 2.69 with 95% CI of (0.74, 9.76). By assuming an odds ratio of 1.75, a Mantel-Haenszel test of H0: odds ratio =1 for 2x2 tables in 5 strata (by quintiles of principal score strata) will have 80% power using a two-sided 0.050 level test, when the sample size in each group is 203 and the stratum and response proportions are assumed to be {0.40, 0.425, 0.45, .425, .40} respectively to each of the five stratum. The detectable effect sizes are plotted against the sample size given 80% power and 2-sided alpha value of 0.05.



Additional sample size is estimated for continuous type of outcome, e.g. cJADAS and PedsQL scores. A sample size of 200 in each group will have 80% power to detect an effect size of 0.281, or a difference in means of 1.685 assuming that the common standard deviation is 6.000 using a two group t-test with a 0.050 two-sided significance level.

2.7. Recruitment

The PCATS will use secondary data that have already collected during routine clinic visits and a JIA registry for analysis. No patient recruitment will be involved in the PCATS study.

3. METHODS: Data Collection and Management

3.1. Data Collection

3.1.1. Case Report Forms

PR-COIN legacy data (2011-2016 years) had three CRFs: patient registration form, patient reported form, and patient encounter form. The patient registration form was used to report patient demographics. The patient reported form was used for patient-reported outcomes. The patient encounter form included more details about patients' medication prescriptions, lab results, and other clinical measures.

PCATS CRFs are created based on recommendations from the consensus treatment plan[3] and the existing CRFs in the PR-COIN. The CRFs are developed in REDCap by the data managers and

are reviewed and approved by the investigator team and data management team. The study data elements are also reviewed and discussed with the Stakeholder Advisory Panel (SAP).

Data elements in the PCATS study include (Table 4): participant demographics, disease characteristics and manifestation, clinical, genetic, and physiological measures, physical exams, medication assignment, health related quality of life, and comorbidity. These data elements are included in the study based on the current literature reviews[3] and expert inputs from stakeholders and clinical co-investigators who determined that they are necessary for accomplishing the goal of the CER aim of the study (Aim2).

Table 4 PCATS Data Elements

Data Fields	Data Elements	Baseline	Follow -up
Demographics and social economics	age, gender, race/ethnicity, JIA subtype (ILAR code)	x	
	insurance type	x	x
Disease characteristics	symptom onset date, date of diagnosis	x	
	disease duration	x	x
Clinical & Functional measures	MD Global assessment (MDG), Pain VAS, patient well-being, CHAQ, morning stiffness, Shober, change since last visits of disease activity	x	x
Genetic measure	HLA-B27	x	
Physical examine	Height, weight, temperature, BP	x	x
Physiological Measures	active joint count (AJC)*, joints with limited range of motion (LOM)	x	x
Health Related Quality of Life	PedsQL generic total score; PedsQL Rheumatology module total score	x	x
Medication Assignment	generic name, class, begin and end date of prescription	x	x
Lab	Erythrocyte sedimentation rate (ESR), Antinuclear antibody (ANA), Rheumatoid factor (RF), Cyclic citrullinated peptide (CCP), C-reactive protein (CRP), Aspartate aminotransferase (AST),	x	x

	Alanine aminotransferase (ALT), Ferritin, Creatinine, Blood urea nitrogen (BUN), Complete blood count (CBC) test, Lipid test, Neutrophil,		
Disease manifestation in the past 2 weeks	rash	x	x
Comorbidity	uveitis, IBD, Celiac Disease, Trisomy 21	x	

3.1.2. PR-COIN Registry

PR-COIN is a multicenter learning network that develops and evaluates JIA disease management strategies to increase remission rates, improve functional status, and improve quality of life. The PR-COIN registry was established in 2011, and by May 2016 it included 15 participating centers. Their data management platform was sponsored by the American College of Rheumatology (ACR). However, in June 2016, the registry data migrate into a different data management system. Due to complications of data migration, only data collected prior to migration were included into the PCATS study (referred to as PR-COIN legacy data). Patients who were cared for at the participating centers, and were diagnosed by a pediatric rheumatologist with JIA following the International League of Associations for Rheumatology (ILAR) criteria were eligible in the PR-COIN registry. Patient demographics, disease characteristics, clinical examination, medication prescription, and patient report outcomes at each clinical encounter were entered into the PR-COIN registry by each participating center for all enrolled patients. Most of the centers entered the data manually, except one center poured the extracted data from its EMR system into the PR-COIN registry. A data committee, composed of PR-COIN members, oversees data quality and policy standards to ensure high quality data.

3.1.3. CCHMC JIA Cohort: Epic™ EMR Data Capture

The Division of Rheumatology at Cincinnati Children's Hospital Medical Center (CCHMC) maintains a registry database for every patients' clinic visit dating back to 2003. To be included into the JIA registry, the patient must have been diagnosed with JIA following the ILAR criteria by at least two pediatric rheumatologists at two distinct clinic visits. Prior to 2007, the data were keyed into a relational database, and the data began to migrate into the CCHMC Epic EMR system in 2007. By 2008, all JIA registry data were embedded as part of the CCHMC Epic system.

Epic™, an electronic medical record system, is primarily designed for clinical care purposes. It is integrated into the providers' clinical workflow and captures real-time patient information, including vital signs, physical examinations, diagnoses, patient-reported outcomes, treatment plans, and clinical procedures. The Epic database allows real-time reporting of data immediately entered by users. Data stored in the Epic data warehouse are transformed into Clarity, which is a relational analytic data repository for Epic™ EMR system. The research data will be extracted from the Epic Clarity database and transformed into the PCATS study specific data model using the procedure language (PL)/structured query language (SQL).

The Biomedical Informatics (BMI) team at CCHMC makes a copy of the Epic Clarity database every night to facilitate the needs of using EMR data for research. The PCATS research team and the BMI team will work closely together in designing filters and mapping rules specifically for the PCATS study (Table 2). After finalization, the data extraction algorithms (Table 5) will be applied to the Clarity Data Mart quarterly using the Oracle 12c. The senior database administrator will be responsible for extracting data from Epic Clarity for the PCATS study. Figure 2 provides an illustration of the Epic data flow. Access to the research datasets will be limited to study staff with privilege depending on their assigned study role.

Table 5 Epic EMR Data Extraction Design

Time / Dataset	**June 2017-current**
_DEMOGRAPHICS	all pts who are part of the CCHMC JIA Registry or CCHMC JIA-QoL project
_ENCOUNTERS	Epic EHR encounters, including canceled and no-show appointments[+]. Exclusions: • Epic "history" encounters *(These are dummy encounters created for administrative purposes and do not represent an actual patient visit.)* • Encounter date prior to 1/1/2008[*]
_DIAGS	Disease diagnoses. Exclusion: • Diagnosis date prior to 1/1/2008[*]
_FLOWSHEETS	Flowsheet name starts with "RHE" *(filters out all non-rheumatology flowsheets)* Exclusion: • flowsheet recorded time prior to 1/1/2008[*]
_LABS	Lab results. Exclusion: • ordered date prior to 1/1/2008[*]
_MEDS	Exclusion: • medication ordered date prior to 1/1/2008[*]
_QUESTIONS	Question name starts with "RHE" *(filters out all non-rheumatology questions)* Exclusion: • question response date prior to 1/1/2008[*]
_SMART_DATA	Data recorded in Epic Smart Data locations. Exclusion: • entry recorded date prior to 1/1/2008[*]
_INSURANCE	Insurance coverage info at the encounter level. Exclusions: • coverage for encounters prior to 1/1/2008[*]

Note: [+] The canceled and no-show appoints are not used in the CER analyses.

[*] The data collected prior to 1/1/2009 are not used in the CER analyses.

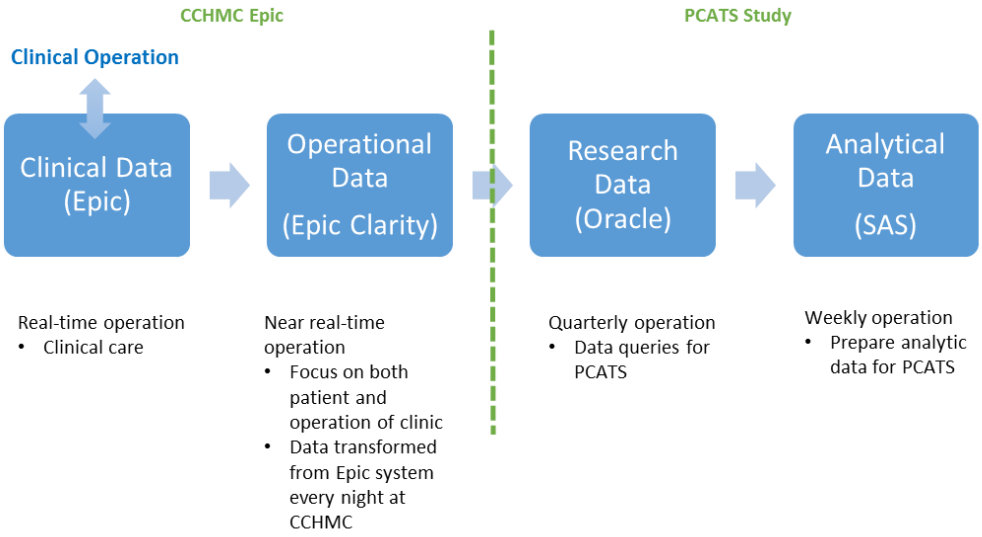


Figure 2 Epic EMR Data Flow

3.1.4. CCHMC JIA-QoL Cohort

Out of the 220 participants of the CCHMC JIA-QoL research, some of them are eligible for PCATS study. These patients will also be part of the larger CCHMC JIA cohort, but these CCHMC JIA-QoL patients will contribute more detailed information on patient personal (e.g. self-efficacy, coping, social support), parental (e.g. parental stress), and family environmental measures (e.g. family coherence) via in-person interviews. The additional data information could be used to augment causal inference analyses. Additional, since JIA-QoL study prospective collected some data fields that overlap with the PCATS study and had gone through careful data quality checking, these data could be used for quality assurance purpose. Most of the study team members are also IRB approved investigators on the JIA-QoL study; therefore they already have access to the PHI.

3.2. Data Management

After collecting data from all three registries, the data managers will perform data cleaning and prepare analytical datasets. Even though structured data fields are extracted from the Epic electronic medical records, data quality issues remain challenging. The data managers will perform data validation by using SAS programming or/and review of patients electronic medical records. Also, the data managers will derive de-identified analytical datasets for the study investigators.

3.2.1. PR-COIN Registry

Limited datasets will be extracted out of the PR-COIN registry database by a PR-COIN designated data manager. The dataset will be reviewed and transferred into research datasets following the database design specified in PCATS CRF. Based on the registry, patients who meet the PCATS enrollment criteria will be identified and included into the PCATS study.

3.2.2. CCHMC JIA Cohort: Epic EMR

To ensure data obtained from a mature Epic system, PCATS study only considers patients in the CCHMC JIA cohort who had clinical encounters during the time period of January 1st, 2009 - December 31st, 2017. The end date is determined to ensure PCATS study have sufficient time to process, quality checking and analyzing data for the specific Aim 2 of the study.

For the CCHMC Epic EMR data, an Epic database administrator will create relational databases for PCATS project that contains research data extracted from Epic Clarity database, and then the data will be passed to the data managers for quality controls.

3.2.3. CCHMC JIA-QoL Cohort

In order to assure the quality of data used in analyses, CCHMC JIAQoL cohort will be used for quality checking. This NIH R01 funded research study[2] is an inception prospective follow-up cohort study, which was designed for understanding quality of life in children with newly diagnosed JIA. It includes 220 patients who were treated at the same medical center (CCHMC) as the Epic data. Many of the data fields overlap with the PCATS data collection. Therefore, this dataset is feasible for validating purposes and can help ensure the quality of the data derived from Epic. We refer to this quality control dataset as JIA-QoL data.

3.2.4. Data Validation

The validation procedures consist of comparing Epic data with the JIA-QoL data on the same patient-encounters and their overlapping data fields. The Epic data extraction algorithms will be revised based on the findings, and the revised algorithms will be quality checked again following the same procedure. The processes will be performed repeatedly until the results yielded minor or no discrepancies.

3.2.5. Data Quality

Figure 3 is a quality check procedure for the Epic EMR data. The data managers are IRB approved unblinded study team member. They will access identifiable data extracted from Epic Clarity for the purpose of verifying the dataset for accuracy and completeness by reviewing a small subset of participants' electronic medical records. No other members of the study team will receive or have access to this identifiable data from the Epic Clarity.

After receiving extracted EMR data from the BMI team, the Data Management team will perform data queries using SAS and quality control data to ensure the completeness, correctness, plausibility, and concordance of the data. The data queries will be programmed and validated by the lead data manager in the SAS. Quality checks will be completed before the database is locked. Manual EMR chart reviews maybe performed if abnormalities or outliers are found. Table 6 presents a list of data quality control queries.

After resolving all outstanding issues concerning data quality, the PCATS team will transform the Oracle database into a SAS relational database following the PCATS CRF.

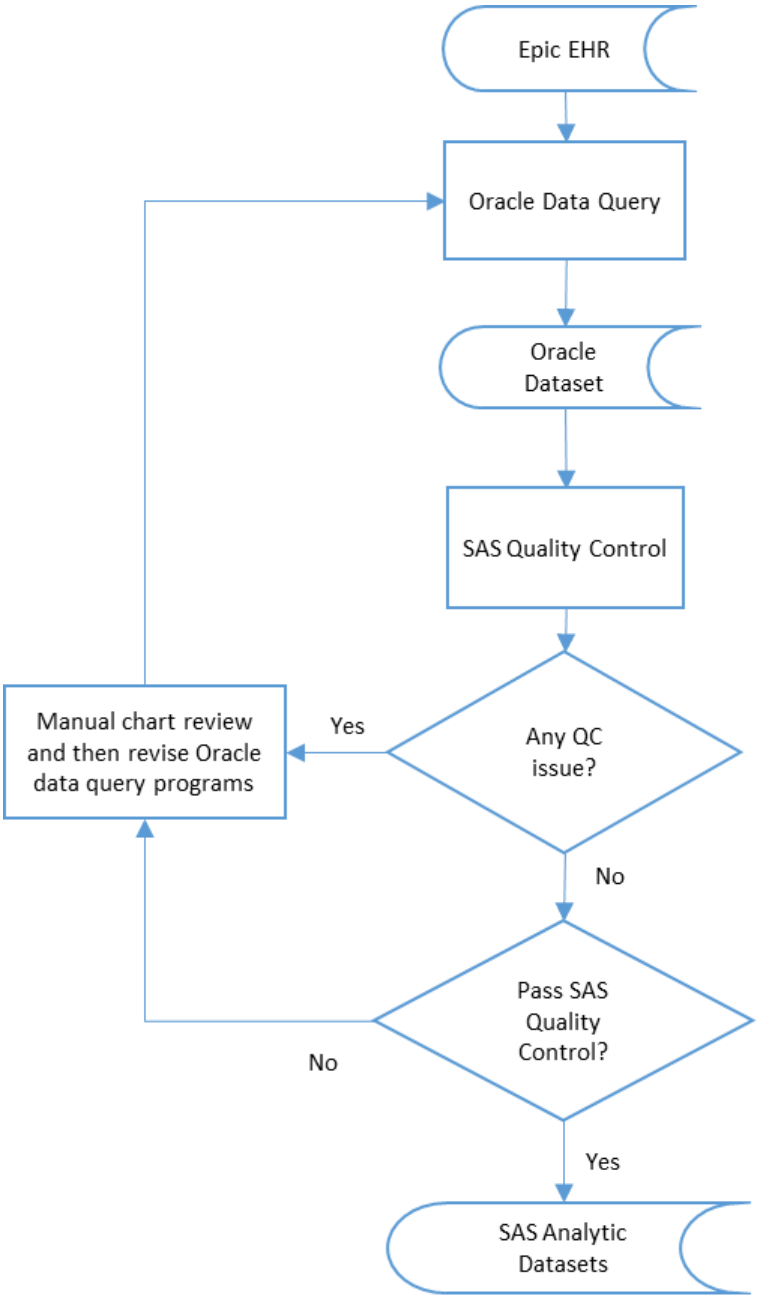


Figure 3 EMR Data Quality Control Process

Table 6 EMR Data Quality Control Queries

Quality Controls	Queries	Actions
Completeness	How much data are missing for each variable?	If the missing are substantial, then manual chart review and further investigations should be performed to identify causes of the missing data. If the causes are due to Oracle query programs, then we communicate with the BMI team about the issues, and revision of query programs will be further validated and revised as needed.
Correctness	Does data value exceed the defined ranges?	If yes, then potential causes are investigated. If deemed illegitimate, then set the out of range value as missing.
	Is the same measure unit used for the give data field?	If not, they should be converted into a standard unit of measure.
Plausibility	Are dates followed this temporal order: date of birth < date of symptom onset ≤ date of diagnosis ≤ date of first DMARD prescription?	If not, manual chart review should be performed to identify all relevant information for reconciliation. If deemed not plausible, set the dates to missing.
Concordance	Are there duplicate records? If so, are the duplicate records consistent for patient-encounters?	If duplicate records are inconsistent for patient-encounters, then communicate with physicians to understand possible causes of duplicate data and resolve duplicate data accordingly.
	Are there multiple dates of diagnosis and dates of symptom onset per patient?	If yes, then use the earliest date.
	Are there multiple data values per patient per encounter date for clinical measures, i.e. height, weight, temperature?	If yes, then calculate and record mean values after ensuring consistency of unit and plausibility.
Concurrency	NA	Study design ensures concurrency

3.2.6. Data Security

Only limited de-identified datasets will be available to the study personnel and be used for all analyses in this study, and the identification of patients will be untraceable and unknown to investigators. Since the PI and other study staff already have access to the full CCHMC JIA-QoL dataset, those data will remain identified until the end of data collection.

Once all data has been collected, Protected Health Information (PHI) will be organized into limited datasets by the data managers. The data managers will remove all PHI from the data, except date of visits and date of diagnosis. The date of visits and date of diagnosis are essential for the CER analyses, thus will remain in the limited datasets. Each patient will be assigned with an individual study ID number. The data managers will maintain the linkage between the study ID and the patient identity and they will also responsible for maintaining the storage and integrity of the PHI.

3.2.7. Data Closure

Lock Procedures

The study will initiate a database lock after data cleaning, validations, and reconciliations are completed and upon agreement of the database quality amongst members of the data management team. The data managers are responsible for locking the database. Final analysis of study data must always be carried out on a locked version of the data.

The following list of tasks will be completed prior to database lock:

- All queries have been resolved.
- Coding has been reviewed for completeness and consistency.
- Review of quality and consistency check output has been conducted.
- Frequencies have been reviewed.

Unlock Procedures

There may be a need for the database to be unlocked if correction of validation errors are required. The PI must initiate the database unlock request. The data management team will make necessary amendments and record the status of the database as partially locked.

3.2.8. Data Storage

Data are stored on the designated project drives at CCHMC which are backed up by the Department of Information Services (IS) every 24 hours. All data kept in electronic form are password protected by a username and password.

4. METHODS: Analysis

4.1. Aim 1: Improving Causal Inference Methods

To improve causal inference method that is robust to model miss-specification, GPMatch is proposed as a Bayesian's nonparametric causal inference method. The GPMatch utilizes Gaussian process (GP) prior as the matching tool, where the GP prior is formulated in such a way that, for each individual patient in the sample, it allocates a weight ranges from 0 to 1 to patients in the dataset, based on the similarity defined by the squared-exponential (SE) covariance matrix.

$$\mathrm{K}(v_i, v_j) = \sigma_f^2 \exp\left(-\sum_{k=1}^{q} \frac{|v_{ki} - v_{kj}|^2}{\phi_k} \right).$$

The SE covariance function is used for its ability to fit smoothed response surface. By including confounding variables (denoted by V) into the covariance function, the GP prior specifies that patients with same values of all confounder variables are matched completely, i.e. assigning a weight of 1. The GP prior assigns a smaller weight to patients who are less similar and zero weight to patients who are sufficiently different. As a consequence, GP prior accomplishes "matching" for each individual patient. It then estimates expected potential outcomes for a given patient by utilizing data information from other "matched" patients who are sufficiently similar. The matching, weighting, and estimation processes are accomplished in a single step of Bayesian GP regression modeling. Therefore, it accomplishes matching and outcome modeling in one single step. The GPMatch method can easily incorporate different types of treatments. For example, continuous treatment and its potential higher order terms could be included in modeling treatment. The heterogeneous treatment can be evaluated by including the treatment by covariate interactions. Higher order terms could also be included to model treatment effect as a nonlinear function of a continuous variable.

The Mahalanobis distance (MD) matching is a much better performed matching method than the propensity score matching method, in the sense that it approaches the true treatment value with increased accuracy as the method achieves better matching. Whereas matching on propensity score could present more erroneous paradoxical behavior as the matching becomes more precise[32]. The MD is defined by

$$MD_{ij} = \begin{cases} \sqrt{(\boldsymbol{v}_i - \boldsymbol{v}_j)' S^{-1} (\boldsymbol{v}_i - \boldsymbol{v}_j)} & \text{, if } |v_{ik} - v_{jk}| < cs_k, \text{for } k = 1,2,\dots,q, \\ \infty & otherwise, \end{cases}$$

where $c \in R^+$ is the caliper, S is the sample variance covariance matrix of confounding variables $\boldsymbol{v}$. Of note, the MD matching requires specification of a caliper. Smaller c leads to more precise matching but often results in a serious reduction in sample size after matching.

Compared with the MD matching, GPMatch does not require arbitrary specification of a caliper; instead, it estimates length scale parameters (ϕ_k) which governs the extent to which the data points are matched on. The length scale parameter also distinguishes confounding variables by their relative importance in their role of matching. The variables with larger value of ϕ_k parameter are considered more important than those with smaller values. The performance of GPMatch will be compared against the MD matching in simulation studies.

Additional simulation studies will further compare the performances of GPMatch method against the many other widely used causal inference methods, including propensity score (PS) sub-classification matching, linear regression with PS adjustment, linear regression with spline fit PS adjustment[33], augment inverse probability of treatment weighting (AIPTW)[34], and the Bayesian additive regression tree (BART)[35]. The simulations are designed to reflect more realistic real world CER settings which are plagued by modeling uncertainty about the treatment assignment and the outcome model. In particular, the well-known Kang and Schafer's simulation design are considered. The simulation will also consider cases of heterogeneous treatment effect. The evaluation criteria will focus on statistics operational characters, including bias, absolute median

error, root mean square error and coverage rate of interval estimates. Different sample sizes will be considered to evaluate if the performances improve with increased sample size.

For evaluating causal treatment effect of the time-varying adaptive treatment, we will extend both GPMatch and BART methods for considering multiple treatment decision points, taking Bayesian's g-computation approach. Under the assumptions of sequential ignorable treatment assignment, stable unit treatment value assumption (SUTVA) and positivity of treatment propensity assumption, the g-computation method factorize the joint likelihood of outcomes into a product of multiple likelihood of outcome models at each of the follow-up time point, given the past history of treatment and covariates, up to final study endpoint, $for\ t = 1, 2, ..T$. The g-computation method will be used in conjunction with flexible Bayesian nonparametric method of GPMatch and BART, allowing for predicting the potential outcomes are each follow-up point for each potential treatment trajectories. Because BART and GPMatch are able to achieve more accurate estimation of potential outcomes and treatment effect at each time point, both are expected to perform well over time. The simulation studies will consider two-staged time-varying adaptive treatment, including settings with considering heterogeneous treatment effect. The same operational characteristics will be used for evaluating performances of the methods. Different sample sizes will be considered.

The online application will be developed using R shiny, implementing both GPMatch and BART. It will allow users to upload their own data, specify outcome and treatment variables, the confounder and prognostic variables. The output will present a table comparing two treatment arms side by side on the selected confounding and prognostic variables, and estimates of averaged treatment effect and predicted potential outcomes will be presented in both tables and figures. If heterogeneous treatment effect option is selected, then the heterogeneous treatment effect estimates will also be presented. A detailed user's guide will be developed and made available. Both the App and the user's guide will provide examples and step-by-step instructions for some commonly encountered CER problem settings, including continuous, multi-level, categorical and mixed composite types of treatment, either statics over time or adaptive. These examples will facilitate better user experiences. The App will be made publicly available.

4.2. Aim 2: Analysis of the CER study

The baseline patient's demographic, insurance, and disease characteristics will be compared between the two treatment arms. The propensity score is derived using the covariate balancing propensity score (CBPS)[36] 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, physician global assessment, pain VAS, patient well-being, 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 will be used for 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 weighting by the CBPS.

For comparing effectiveness of the 1st line treatment on the 6-monts endpoint, GPMatch along with all the other existing causal inference methods considered in the simulation study will be applied and compared. 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[37]. The comparative effectiveness of early-combination CTP compared to the step-up CTP will be reported at 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, patients 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 will be 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. However, we make the missing at random assumption, assuming that patients who missed the follow-up visit will have similar treatment effect as their peers who had the visit, conditioning on they share the similar baseline characteristics. To handle missing data at baseline, we applied Bayesian multivariate missing data imputation technique, hierarchically coupled mixture model with local dependence (HCMM-LD)[38] 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)[39,40]. The diagnosis of missing data imputation will be presented by presenting kernel fit of the distributions of variables before and after imputation. For sensitive analyses, missing data will also be 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[41].

It is perceivable that patient with different disease subtype, and disease activities at the baseline, and duration of treatment may have different treatment effects. 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 will be repeated by including additional baseline covariates. Specifically, patient reported health related quality of life at the baseline is not included in the primary analyses for

the concern over the large percentage of missing at the baseline. The sensitivity analyses will include 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 and estimated potential outcomes will be also reported for the 6 and 12-month outcomes.

The analyses report will follow the Patient-Centered Outcomes Research Institute (PCORI) methodology guideline (https://www.pcori.org) and strengthening the reporting of observational studies in epidemiology (STROBE) recommendations (https://www.strobe-statement.org). All analyses will be performed in SAS (https://www.sas.com) and R (https://www.r-project.org/).

5. METHODS: Monitoring

5.1. Data monitoring

A Data Safety Monitoring Board is not required for this research because this involves no more than minimal risk. The Principal Investigator and Project Managers will periodically review data to ensure that the participants enrolled in the study meeting the inclusion and exclusion criteria.

5.2. Stakeholder Advisory Panel

The Stakeholder Advisory Panel consists of representatives of patients and clinician care providers and health policy and outcome researchers, who have a commitment to patient-centeredness through active participation in the study. The primary mission of the Stakeholder Advisory Panel is to ensure the project is patient-centered. In particular, to confirm that the method development can adequately answer the specific patient-centered questions and that the method development is understandable, meaningful and generalizable to broader patient populations and stakeholders. The secondary mission of the stakeholder panel is to monitor and advise study progress, provide approval and facilitate the dissemination of the study results. Upon receiving the award, we will have an initial in-person Stakeholder Advisory Panel meeting, where a chair will be elected and a charter will be developed and approved. The initial meeting will be in-person at CCHMC; however, the remaining meeting that will be conducted twice per year (or more often as necessary) will utilize conference call and email communications. The panel has been involved in the design, development, and approval of study protocols. They will continue to be involved in building and maintaining the study database; reviewing and advising the analysis plan and interpretation of the results; advising and helping with the dissemination of the study results, such as crafting introductory educational materials, and advising the best approaches to share the study results to larger patient and stakeholder populations. They will receive the report of the CER analyses results.

In summary, our patients and clinician stakeholders have been and will be fully committed to engaging in every step of the study by serving on the Stakeholder Advisor Panel and on the study team. We have already established close relationships as a team; all engagement will be interactive and co-learning, based on the principals of trust, transparency, and honesty. For compensating their effort, members of the Stakeholder Advisory Panel will each receive a $1,500 honorarium for the entire study period, and they will be compensated for the travel expenses during the initial face-to-face meeting.

5.3. Harms

This study involved minimal risks, and it does not require consenting participants.

6. ETHICS AND DISSEMINATION

6.1. Ethics

The investigator will ensure that this study is conducted in full conformity with the principles set forth in The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, as drafted by the US National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research and codified in 45 CFR Part 46 and/or the ICH E6.

6.2. Protocol Amendments

Protocol amendments will be submitted to IRB if there are changes to eligibilities criteria, outcomes, or staffs.

6.3. Consent/Assent

The IRB protocol has been approved. This study involved minimal risks, and it does not require consenting participants. No consent/assent form is created or needed approval from the IRB.

6.4. Privacy and Confidentiality

Participant confidentiality is strictly held in trust by the investigators, study staff, and the sponsor(s) and their agents. The study protocol, documentation, data, and all other information generated will be held in strict confidence. No information concerning the study or the data will be released to any unauthorized third party without prior written approval of the sponsor.

The privacy and confidentiality of patient information will be maintained in accordance with the Health Insurance Portability and Accountability Act (HIPAA) regulations. All research personnel who work on this study must complete HIPPA and the Collaborative Institutional Training Initiative module on human research with direct subject interaction. No PHI will be used in any publications that are a result from this work.

6.5. Declaration of Interests

There is no conflict of interests involved in the study.

6.6. Access to Data

The principal investigator (PI) leads the project data team, consisting of a blinded and un-blinded statisticians, data managers, and research assistants. The team works closely with the BMI collaborator, Epic specialist, Epic project manager, and the PR-COIN research team actively involved in managing and securing the data. Following project completion, the Epic portion of the PCATS data will be uploaded into a REDCap database. Upon the close out of the PCATS IRB, the project manager will submit to the IRB for the data to be managed under a data repository.

The de-identified data will be used for future research. The identifiable data will only be viewed by the data managers who are unblinded team members. All the data will be kept in a secure server with restricted access.

6.7. Dissemination Policy

During and following the funding period we plan to actively work on disseminating and implementing the results of the proposed project through presentations, courses, developing an online application and through ongoing scholarly activities. The goal of dissemination will be to contribute to stronger and more reproducible comparative effectiveness research and to help improve patient centered health care utilizing real world data.

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. 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.

3. 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.

4. Murphy SA, Bingham D. Screening Experiments for Developing Dynamic Treatment Regimes. Journal of the American Statistical Association. 2009 Mar 1;104(458):391–408.

5. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). Clinical Trials: Journal of the Society for Clinical Trials. 2014 Aug 5;11(4):426–34.

6. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling. 1986 Jan 1;7(9–12):1393–512.

7. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: proofs of results. The international journal of biostatistics. 2010 Mar 3;6(2):Article 9.

8. Robins JM. Marginal Structural Models versus Structural nested Models as Tools for Causal inference. In Springer, New York, NY; 2000. p. 95–133.

9. Young JG, Hernán MA, Picciotto S, Robins JM. Relation between three classes of structural models for the effect of a time-varying exposure on survival. Lifetime data analysis. 2010 Jan;16(1):71–84.

10. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. Statistics in Medicine. 2013 Apr 30;32(9):1584–618.

11. Murphy SA. A Generalization Error for Q-Learning. Journal of machine learning research : JMLR. 2005 Jul;6:1073–97.

12. Watkins & Dayan (1992) [Internet]. [cited 2018 Apr 23]. Available from: http://www.gatsby.ucl.ac.uk/~dayan/papers/wd92.html

13. Murphy SA. Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2003 May 1;65(2):331–55.

14. Robins JM. Optimal Structural Nested Models for Optimal Sequential Decisions. In Springer, New York, NY; 2004. p. 189–326.

15. Hirano K, Porter JR. Impossibility Results for Nondifferentiable Functionals. Econometrica. 2012 Jul 1;80(4):1769–90.

16. Moodie EEM, Chakraborty B, Kramer MS. Q-learning for estimating optimal dynamic treatment rules from observational data. The Canadian journal of statistics = Revue canadienne de statistique. 2012 Dec 1;40(4):629–45.

17. Wallace MP, Moodie EEM. Personalizing medicine: a review of adaptive treatment strategies. Pharmacoepidemiology and Drug Safety. 2014 Jun 1;23(6):580–5.

18. Chakraborty B, Moodie EEM. Statistical methods for dynamic treatment regimes : reinforcement learning, causal inference, and personalized medicine. 204 p.

19. Zhang B, Tsiatis AA, Laber EB, Davidian M. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. Biometrika. 2013;100(3).

20. Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Statistics in medicine. 2009 Nov 20;28(26):3294–315.

21. R Development Core Team. R: a language and environment for statistical computing [Internet]. 2011 [cited 2018 Apr 23]. Available from: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing

22. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.

23. Hu Z, Follmann DA, Qin J. Semiparametric Double Balancing Score Estimation for Incomplete Data With Ignorable Missingness. Journal of the American Statistical Association. 2012 Mar;107(497):247–57.

24. Lunceford JK, Davidian M, Tsiatis AA. Estimation of Survival Distributions of Treatment Policies in Two-Stage Randomization Designs in Clinical Trials. Biometrics. 2002 Mar 1;58(1):48–57.

25. Kang JDY, Schafer JL. Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. Statistical Science. 2008 Nov;22(4):523–39.

26. Hogan JW, Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. Statistical Methods in Medical Research. 2004 Feb 2;13(1):17–48.

27. Achy-Brou AC, Frangakis CE, Griswold M. Estimating treatment effects of longitudinal designs using regression models on propensity scores. Biometrics. 2010 Sep;66(3):824–33.

28. Swart JF, van Dijkhuizen EHP, Wulffraat NM, de Roock S. Clinical Juvenile Arthritis Disease Activity Score proves to be a useful tool in treat-to-target therapy in juvenile idiopathic arthritis. Annals of the rheumatic diseases. 2018 Mar 1;77(3):336–42.

29. 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.

30. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL^{TM}* 4.0 as a Pediatric Population

Health Measure: Feasibility, Reliability, and Validity. Ambulatory Pediatrics. 2003 Nov 1;3(6):329–41.

31. 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. 2012 Jun;64(6).

32. King G, Nielsen R. Why Propensity Scores Should Not Be Used for Matching. Political Analysis. 2018;

33. Little R, An H. ROBUST LIKELIHOOD-BASED ANALYSIS OF MULTIVARIATE DATA WITH MISSING VALUES. Vol. 14, Statistica Sinica. Institute of Statistical Science, Academia Sinica; p. 949–68.

34. Bang H, Robins JM. Doubly Robust Estimation in Missing Data and Causal Inference Models. Biometrics. 2005 Dec 1;61(4):962–73.

35. Xu Y, Müller P, Wahed AS, Thall PF. Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. Journal of the American Statistical Association. 2016;1459(October):1–46.

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

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

38. 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.

39. Murray JS. Multiple Imputation: A Review of Practical and Theoretical Findings. Statistical Science. 2018 May;33(2):142–59.

40. Meng X-L. Multiple-Imputation Inferences with Uncongenial Sources of Input. Vol. 9, Statistical Science. Institute of Mathematical Statistics; p. 538–58.

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

PCATS Case Report Forms

Demographic and Critical Dates

Record ID	_____
Patient ID	_____
Date of Birth	_____ (D-M-Y)
Age when first taken DMARD	_____ (in Year)
Gender	◯ Male ◯ Female
Race	◯ American Indian and Alaska Native ◯ Asian ◯ Native Hawaiian and Other Pacific Islander ◯ Black or African American ◯ White ◯ Multi-race ◯ Other ◯ Unknown
Ethnicity	◯ Hispanic ◯ Non-Hispanic ◯ Unknown
Insurance	◯ Public ◯ Private ◯ Multiple-type ◯ Other
Date of Symptom Onset *	_____ (D-M-Y)
Date of Diagnosis	_____ (D-M-Y)
Date of First DMARD Rx	_____ (D-M-Y)

Note:
* These variables are not available in PR-COIN registry.

PCATS Enrollment Criteria

Patient ID	_____
MRN	_____
1. Polyarticular course JIA	◯ Yes ◯ No
2. Have Clinical Encounter within 6 Months after Diagnosis	◯ Yes ◯ No
3. Ever Have DMARDs Prescription	◯ Yes ◯ No
4. Received DMARD Prescription as a New Patient	◯ Yes ◯ No
5. Age at Initiation of First DMARD Prescription	◯ < or =19 years old ◯ >19 years old
6. Have Comorbidity: IBD, celiac disease, or trisomy 21 *	◯ Yes ◯ No
Inflammatory Bowel Disease *	◯ Yes ◯ No
Celiac Disease *	◯ Yes ◯ No
Trisomy 21 *	◯ Yes ◯ No

Vitals

Patient ID

_____

Visit Date

_____
(D-M-Y)

Height (cm)

_____
(Centimeters (cm))

Weight (kg)

_____
(kilogram (kg))

Temperature *

_____
(Fahrenheit (F))

Blood Pressure (Diastolic) *

_____

Blood Pressure (Systolic) *

_____

Clinical Assessment

Patient ID

_____

Visit Date

_____
(D-M-Y)

ILAR Code

◯ Systemic JIA
◯ Polyarticular RF (-)
◯ Polyarticular RF (+)
◯ Oligoarticular (unspecified)
◯ Oligoarticular persistent
◯ Oligoarticular extended
◯ Psoriatic Arthritis
◯ Enthesitis Related Arthritis
◯ Undifferentiated Arthritis
◯ Other

pJIA Subtype

◯ Polyarticular RF (-)
◯ Polyarticular RF (+)
◯ Oligoarticular
◯ Other

Clinical Assessment

Physician Global Assessment

_____
(range 0-10)

Disease Activity- change since last visit *

◯ Much better
◯ Same
◯ Somewhat better
◯ Somewhat worse
◯ Worse

Patient Overall Well-being

_____
(range 0-10)

Patient Reported Pain Level

_____
(range 0-10)

Morning Stiffness

◯ No stiffness
◯ 15 minutes or less
◯ >15-30 minutes
◯ >30 minutes-1 hour
◯ >1-2 hours
◯ >2-4 hours
◯ >4 hours

Number of Joint with Limited Range of Motion *

_____

Active Joint Count

_____
(range 0-71)

Schober Test Score *

_____

JADAS Score *

_____

cJADAS10 Score

_____
(range 0-30)

Review of System

Uveitis

◯ Yes
◯ No

Rash *

◯ Yes
◯ No

Labs

Patient ID	_____
Visit Date	_____ (D-M-Y)
Erythrocyte sedimentation rate (ESR) +	_____ (mm/hr)
ESR record date	_____ (D-M-Y)
Antinuclear antibody (ANA)	◯ Positive ◯ Negative
ANA record date	_____ (D-M-Y)
Rheumatoid Factor *	◯ Positive ◯ Negative
RF record date *	_____ (D-M-Y)
Anti-CCP *	◯ Positive ◯ Negative
Anti-CCP record date *	_____ (D-M-Y)
CRP +	_____ (mg/L)
CRP record date	_____ (D-M-Y)
AST +	_____ (unit/L)
AST record date	_____ (D-M-Y)
ALT +	_____ (unit/L)

ALT record date

_____ (D-M-Y)

Ferritin *

_____ (ng/mL)

Ferritin record date *

_____ (D-M-Y)

Human leukocyte antigen B-27 (HLA-B27) * ◯ Positive ◯ Negative

Creatinine +

_____ (mg/dL)

Creatinine record date +

_____ (D-M-Y)

Blood urea nitrogen (BUN) +

_____ (mg/dL)

BUN record date +

_____ (D-M-Y)

CBC: Red Blood Count (RBC) +

_____ (M/mcL)

RBC record date +

_____

CBC: White Blood Count (WBC) +

_____ (K/mcL)

CBC: Hemoglobin (HGB) +

_____ (gm/dL)

CBC: Hematocrit (HCT) +

_____ (%)

HCT record date +

_____ (D-M-Y)

CBC: Platelet +

_____ (K/mcL)

Platelet record date +

_____
(D-M-Y)

Lipid: Total Cholesterol *

_____

Total Cholesterol record date *

_____
(D-M-Y)

Lipid: Low density lipoprotein (LDL) *

_____
(mg/dL)

LDL record date *

_____
(D-M-Y)

Lipid: High density lipoprotein (HDL) *

_____
(mg/dL)

HDL record date *

_____
(D-M-Y)

Lipid: Triglycerides *

_____
(mg/dL)

Triglycerides record date *

_____
(D-M-Y)

Absolute neutrophil count (ANC) *

_____
(K/mcL)

Neutrophil record date *

_____
(D-M-Y)

Note:
* These variables are not available in PR-COIN registry.
+ These variables are reported differently in PR-COIN registry. [Lab values were reported as dichotomized variables in PR-COIN. For example, ESR was reported as elevated (yes/no). Creatinine and BUN were reported as one variable.]

Peds QL Questionnaire

Patient ID

_____

Visit Date

_____
(D-M-Y)

PedsQL Generic Score - Child Reported

PedsQL Generic Questionnaire Record Date

_____
(D-M-Y)

Physical Functioning *

_____

Emotional Functioning *

_____

Social Functioning *

_____

School Functioning *

_____

Generic Total Score

_____

PedsQL Generic Score - Parent Reported

PedsQL Generic Questionnaire Record Date

_____
(D-M-Y)

Physical Functioning *

_____

Emotional Function *

_____

Social Functioning *

_____

School Functioning *

_____

Generic Total Score

_____

PedsQL Rheumatology Module - Child Reported

PedsQL Rheumatology Module Record Date

_____
(D-M-Y)

Pain and Hurt *

_____

Daily Activities *

_____

Treatment *

_____

Worry *

_____

Communication *

_____

PedsQL Rheumatology Module - Parent Reported

PedsQL Rheumatology Module Record Date

_____
(D-M-Y)

Pain and Hurt *

_____

Daily Activities *

_____

Treatment *

_____

Worry *

_____

Communication *

_____

Medication Prescription

Patient ID

_____

Medication Generic Name

_____

Pharmacologic Class

◯ Non-Biologic DMARDs
◯ Biologic DMARDs
◯ Glucocorticoid
◯ NSAID
◯ Eye Drop
◯ Joint Injection

Prescription Start Date

_____
(D-M-Y)

Prescription End Date

_____
(D-M-Y)