Supplemental Material

Bayesian comparative effectiveness study of four consensus treatment plans for initial management of systemic juvenile idiopathic arthritis: FiRst Line Options for Systemic juvenile idiopathic arthritis Treatment (FROST)

Peter A. Nigrovic, Timothy Beukelman, George Tomlinson, Brian M. Feldman, Laura E. Schanberg, and Yukiko Kimura and the Childhood Arthritis and Rheumatology Research Alliance Systemic juvenile idiopathic arthritis consensus treatment plan Workgroup

Inclusion criteria

Entry criteria for FROST are as follows:

1. Age 6 months to 18 years of age.
2. Enrollment in the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Registry at a site participating in FROST.
3. A diagnosis of systemic juvenile idiopathic arthritis as per the CARRA modified definition of the International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: onset before the 18th birthday, fever for at least 2 weeks, arthritis in one or more joints for at least 10 days, plus at last one of: evanescent erythematous rash, generalized lymphadenopathy, hepatomegaly and/or splenomegaly, and pericarditis, pleuritis and/or peritonitis.[1] Daily fever is not required, but fever must at some point exhibit a quotidian pattern, defined as fever that rises to ≥39°C at least once a day and returns to ≤37°C between fever peaks. Arthritis is defined as swelling within a joint, or limitation in the range of joint movement with joint pain or tenderness, observed by a physician and not due to primarily mechanical disorders or other identifiable causes.

4. No prior treatment for systemic juvenile idiopathic arthritis except NSAIDs and/or short-course systemic glucocorticoids, defined as <2 weeks total including ≤ 3 pulse doses of methylprednisolone. Prior treatment with intravenous immunoglobulin is permitted in patients previously considered to have Kawasaki Disease.

Exclusion criteria

Patients are excluded for any relative or absolute contraindication to biologic therapy. Examples include 1) concomitant active or recurrent chronic bacterial, fungal or viral infection, for example a positive screening test for tuberculosis without documented past treatment; 2) active or past malignancy; and 3) immunization with live virus vaccines within 4 weeks prior to enrollment. Consent for biosample collection is not required for participation. Since the onset of systemic juvenile idiopathic arthritis is sometimes difficult to pinpoint, disease duration is not an exclusion criterion.

Outcomes

Given the observational nature of the FROST study and the planned Bayesian analysis, formal primary and secondary outcomes are not defined. However, the outcome of principal interest used for determination of sample size is attainment of clinically inactive disease off glucocorticoids at 9 months. Clinically inactive disease is defined as no active arthritis, a physician's global assessment of disease activity score of 0, ESR and/or CRP in the normal range, no extra-articular features of systemic juvenile idiopathic arthritis (fever, rash, serositis, splenomegaly, or generalized lymphadenopathy), no uveitis, and duration of morning stiffness <15 minutes.[2] Discontinuation of glucocorticoids was included in the outcome because of the high morbidity associated with sustained treatment in children, including with respect to linear growth, bone mineralization, and other consequences. The main intended comparison is

between non-biologic (Consensus Treatment Plans 1+2) and biologic (Consensus Treatment Plans 3+4) strategies.

Additional outcomes being assessed include elapsed duration of time to achieve clinically inactive disease off glucocorticoids; glucocorticoid-free status at 6 months; cumulative glucocorticoids dose; pain, global assessment, quality of life and other PROs (**Table 2**); American College of Rheumatology (ACR) pediatric 50, 70, 90 and 100 percent improvement scores at 6 and 9 months, modified to require absence of fever;[3] patient-reported medication side effects; Serious Adverse Events (SAEs), Common Terminology Criteria for Adverse Events (CTCAE v4.0) Grade III or higher, and episodes of macrophage activation syndrome and other pre-defined Events of Special Interest not meeting SAE or CTCAE grade III criteria; inadequate response and/or intolerance of initial consensus treatment plan at 3, 6 and 9 months; and consensus treatment plan switching and reasons for switching. Further exploratory endpoints include comparison of IL-1 vs. IL-6 inhibition with respect to attainment of clinically inactive disease off glucocorticoids at 9 and 24 months, active joint count, control of systemic symptoms, and lab values.

Consensus treatment plan switching

The observational nature of FROST allows for patients to switch to a different consensus treatment plan at the discretion of the treating clinician. We will address these scenarios as follows (**Table S1**). Patients will be assigned for the purpose of analysis to the consensus treatment plan actually begun at enrollment. With respect to the outcome of clinically inactive disease off glucocorticoids at 9 months, patients will be considered failures if they switch groups (non-biologic vs. biologic), including addition of methotrexate to biologic therapy. Within-group switches (non-biologic to non-biologic, biologic to biologic) will be considered failures for the 4

consensus treatment plans but not for the group, with the exception of switches between IL-1 blockers for reason of intolerance or convenience. Separately, we will assess elapsed duration of time from enrollment to first achievement of clinically inactive disease off glucocorticoids. In the analysis, all observed outcomes will be attributed to the consensus treatment plan actually begun, irrespective of subsequent changes. The analysis will evaluate the effect of the initial choice of consensus treatment plan on outcomes, reflecting the choice physicians must make in typical clinical practice.

Missing data

All study data are collected within the CARRA Registry. Through the use of programmed validity and consistency checks at the point of data entry and specific data queries to the clinical sites, we anticipate no substantial difficulties achieving data for required fields.[4] Where data are missing, they will be imputed using a multiple imputation strategy; this assumes that they are missing at random or that missing values are ignorable given observed data. Missing outcomes, especially on patients who have dropped out are potentially more problematic and in addition to the straightforward multiple imputation strategy we will carry out sensitivity analyses that assume various proportions of missing outcomes are treatment failures.

Statistical analysis

We will use the two-stage methods outlined in Kaplan et al.[5] and generate 1000 samples of the propensity score for each subject, which will be used in both stratified and matched Bayesian models for estimation of the difference in treatment response, averaged over the uncertainty in the computation of the propensity score. In the first stage, the propensity score will itself be generated from a logistic regression of initial treatment group on variables known to be associated with treatment choice. We anticipate that the priors in this model will be minimally

informative – for example, limiting the odds ratio relating a dichotomous factor to treatment group to have most of its prior weight on values between 0.25 and 4.0. The choice of the variables to enter the propensity score model can affect both the quality of the matching and the final estimate of the treatment effect on; a standard propensity model generates one propensity score per subject. One advantage of fully Bayesian model for the propensity score estimation is the ability to represent model uncertainty in the generation of these propensity scores. In one of the proposed approaches (Bayesian model averaging), the 1000 sets of propensity scores will come from models with different sets of predictors, so that the propensity model for treatment assignment in addition to pure parameter uncertainty, also reflects the uncertainty in choosing this model. In the second stage of this approach, the Bayesian outcome model will be a logistic regression with clinically inactive disease off glucocorticoids as the outcome. The main analysis will use a minimally informative prior for the effect of treatment on having clinically inactive disease off glucocorticoids. We will also run the model with archetypal priors that Spiegelhalter calls skeptical and enthusiastic.[6] In this case, since we do not have placebo or control treatments *per se*, the skeptical prior for the effect of the biologic consensus treatment plans will be centered at an odds ratio of 1 and put 95% prior weight on odds ratios between 0.5 (biologic consensus treatment plans worse than non-biologic consensus treatment plans) and 2 (biologic consensus treatment plans better). The enthusiastic prior for the biologic consensus treatment plans will be centered at an odds ratio of 2 and extend so it puts only 5% prior probability on values below 1; enthusiasm for the non-biologic consensus treatment plans will represented by a prior for the effect of the biologic consensus treatment plans centered at an odds ratio of 0.5 and extending so it puts only 5% prior probability on values above 1.

Flat priors were used in the analysis of simulated data in the sample size calculation, so that the results depend almost entirely on the simulated data. In particular, t-distributions with 1 degree

of freedom and a scale of 2.5 were used for the intercept and each of the 4 propensity-category coefficients. Our main analyses will also use uninformative priors, but we will use archetypal and expert-based informative priors in sensitivity analyses. The experience-based priors will represent pooled expert opinion on the relative effectiveness of the 4 consensus treatment plans obtained previously from a group of CARRA-affiliated rheumatologists who were asked, for a specific clinical scenario, to state their beliefs about the probability of clinically inactive disease at 9 months under each consensus treatment plan. These prior distributions can be used directly in unadjusted comparisons of the proportion achieving clinically inactive disease. With some manipulation of the priors on probabilities of clinically inactive disease with each consensus treatment plan, we will also create priors on the probability of clinically inactive disease with biologic and non-biologic consensus treatment plans. Finally, we will reframe this pair of priors as a single prior on the relative risk or odds ratio between biologic and non-biologic consensus treatment plans to use in adjusted analyses.

We will adhere to the Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines for observational studies (www.strobe-statement.org), and write a detailed statistical analysis plan before beginning data analysis. We will adhere to this plan, and document and report any deviations.

Biospecimen collection

Patients enrolled in FROST have new-onset systemic juvenile idiopathic arthritis without prior exposure to disease-modifying anti-rheumatic drugs, and will be closely phenotyped and followed longitudinally to establish long-term course creating a unique clinical research resource. Accordingly, biosamples will be collected from FROST patients providing appropriate consent to enable biomarker studies directed at understanding predictors of disease course

(e.g. monophasic vs. persistent course, risk for macrophage activation syndrome) and response to therapy, as well as investigations into disease mechanism. Samples collected at baseline and 6 months include plasma, DNA, serum, peripheral blood mononuclear cells, and whole blood RNA (P100, SST and CPT tubes, Becton Dickenson; Tempus, Thermo Fisher). Clinical sites without centrifugation capacity will provide only SST and Tempus tubes. All samples will be shipped at room temperature according to an established protocol to one of two sites for processing and storage (US, Cincinnati Children's Hospital Medical Center; Canada, Toronto SickKids Hospital).[7] Access to biosamples and linked de-identified clinical data will be available through the CARRA Data and Sample Share Committee, as per regular CARRA procedures (https://carragroup.org/).

Table S1. Outcome determination in FROST

Initial consensus treatment plan	Change	Outcome Assessment	
		Biologic vs. Non-biologic	Individual consensus treatment plans
GC	Begin MTX	As observed	Failure
GC	Begin IL-1 or IL-6	Failure	Failure
MTX	Begin IL-1 or IL-6	Failure	Failure
IL-1	Add MTX	Failure	Failure
IL-1	Switch IL-1	As observed	Failure*
IL-1	Begin IL-6	As observed	Failure
IL-6	Add MTX	Failure	Failure
IL-6	Begin IL-1	As observed	Failure

* If switching due to patient intolerance of medication administration, then outcome will be assessed as observed.

References for Supplementary Material

1. DeWitt EM, Kimura Y, Beukelman T, Nigrovic PA, Onel K, Prahalad S, et al. Consensus treatment plans for new-onset systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken). 2012; **64**(7): 1001-10.
2. Wallace CA, Ruperto N, Giannini E. Preliminary criteria for clinical remission for select categories of juvenile idiopathic arthritis. The Journal of rheumatology. 2004; **31**(11): 2290-4.
3. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis and rheumatism. 1997; **40**(7): 1202-9.
4. Beukelman T, Kimura Y, Ilowite NT, Mieszkalski K, Natter MD, Burrell G, et al. The new Childhood Arthritis and Rheumatology Research Alliance (Childhood Arthritis and Rheumatology Research Alliance) registry: design, rationale, and characteristics of patients enrolled in the first 12 months. Pediatric rheumatology online journal. 2017; **15**(1): 30.
5. Kaplan D, Chen J. A Two-Step Bayesian Approach for Propensity Score Analysis: Simulations and Case Study. Psychometrika. 2012; **77**(3): 581-609.
6. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health care evaluation. Chichester ; Hoboken, NJ: Wiley; 2004.
7. Matheson LA, Duong TT, Rosenberg AM, Yeung RS. Assessment of sample collection and storage methods for multicenter immunologic research in children. J Immunol Methods. 2008; **339**(1): 82-9.