Supplemental Document 1 - Full Experimental Procedures

EXPERIMENTAL PROCEDURES

Study Population and Design

This analysis pools data from four clinical trials (IMAGO, ITCH, IMAGINE, and IMAGINE II). The primary results of ITCH have been published (1). IMAGO and its extension study, IMAGINE, were conducted by evolving sponsors concluding with Mirum Pharmaceuticals at three sites in the United Kingdom, while ITCH and its extension study, IMAGINE II, were conducted by the NIDDK-funded Childhood Liver Disease Research Network (ChiLDReN) in collaboration and through a cooperative research and development agreement (CRADA) between the NIDDK and the same evolving sponsors who developed this IBATi (Lumena Pharmaceuticals and Shire Pharmaceuticals [LUM001] and Mirum Pharmaceuticals [maralixibat – referred to as maralixibat going forward]). The studies conducted in the UK and by ChiLDReN were very similar in design, with earlier initiation of the studies in the UK (Supplemental Table 1). Inclusion and exclusion criteria for IMAGO and ITCH were essentially the same (Supplemental Tables 2A and 2B), enrolling children between the ages of 2 and 18 inclusive, who had ALGS, evidence of cholestasis, intractable pruritus and compensated liver disease. Entry criteria included significant pruritus as assessed by the ItchRO instrument with a requirement of a mean daily ItchRO(Obs) score of $\geq$2 for two consecutive weeks as previously described (1, 2). IMAGO and ITCH were randomized placebo-controlled trials to investigate the safety and efficacy of maralixibat using multiple dosing regimens (based upon the weight of maralixibat chloride, although reported as maralixibat; 280 µg of maralixibat chloride is equivalent to 266 µg of maralixibat free base), which have been previously described for ITCH and were similar in IMAGO (1) (Figure 1, Supplemental Table 1 and Supplemental Figures 1A and 1B). Participants who completed IMAGO and ITCH were provided the opportunity to enroll in follow-up studies of long-term safety and durability of response in IMAGINE and IMAGINE II, respectively. The complex pattern of adjustment in dosing of maralixibat is seen in Supplemental Figures 1A and 1B, which captures dose escalation and dose optimization, which was completed by week 12. Participants who had received placebo started maralixibat in a dose-escalation manner in the extension studies, while those on maralixibat remained on their final dose and were escalated to their maximum tolerated doses in the extension studies. Dosing of maralixibat remained blinded throughout these studies; knowledge of the original randomized dose and the actual dose of maralixibat during the extension study were blinded to the patient and study personnel. The majority of participants received a stable dose of 280 mcg/kg/day. Study drug administration continued according to protocol for 45 of the 57 participants through week 48. Due to procedural issues in the transition of the protocols to the long-term follow-up, some participants had a pause in study drug administration, which was followed by a blinded dose-escalation over 4 weeks at the re-initiation of study drug administration (Supplemental Figures 1A and 1B). A priori, enhanced monitoring criteria and stopping guidelines for total bilirubin (TB) and alanine aminotransferase (ALT) levels were established for potential drug induced liver injury in the setting of pre-existing liver disease (Supplemental Table 3).

Written informed consent was obtained from caregivers, and assent was obtained when appropriate from the child according to local Institutional Review Board (IRB) rules. These studies

were approved by local IRBs and Ethics Committees and complied with the Declaration of Helsinki and Good Clinical Practice Guidelines. The studies were registered at ClinicalTrials.gov. Mirum was not involved in the data analysis or the interpretation and reporting of the results, which was independently performed by ChiLDReN in collaboration with the UK investigators.

A number of endpoints were examined during the 13-week double-blind trials and up to 220 (IMAGINE II) or 288 weeks (IMAGINE) in the extension studies at approximately 12-week intervals. Week 0 was set at the initiation of IMAGINE or IMAGINE II, therefore the initiation of study drug administration in ITCH and IMAGO occurred at week -13. Endpoints included change in pruritus as measured by ItchRO(Obs) and the Clinician Scratch Scale (CSS). Change in quality of life was assessed using PedsQL total score-parent and subscores for the Multidimension Fatigue (MFS) and Family Impact total scale modules, which may be major factors impacted by pruritus in ALGS (3). A priori, clinically significant improvements in ItchRO(Obs) and CSS were set at $\leq$ -1 and at $\geq$ +10 for the quality of life measurements. Other endpoints included total serum bile acids (SBA), total serum cholesterol, alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transpeptidase (GGT), total bilirubin (TB), platelet count (plt), the AST to platelet ratio index (APRI) and albumin. Growth and nutriture were assessed by examining changes in weight, height and BMI z-scores, along with fat soluble vitamin levels assessed as previously described (4).

Treatment-emergent adverse events (AEs), serious adverse events (SAEs), treatment discontinuations due to AEs, and AEs of special interest (e.g., gastrointestinal (GI) symptoms, liver injury and fat-soluble vitamin level abnormalities) were analyzed to characterize the safety and tolerability of maralixibat.

Statistical Methods

The focus of analyses is on the long-term effect of maralixibat on efficacy and safety outcomes during the extension studies, thus the majority of summaries pool data from all four studies across all doses of maralixibat (ranging from 140 to 560 µg/kg/d) without consideration of the original randomized treatment given during the 13-week double-blind studies. All participants who were enrolled and received at least one dose of the study drug were included in analyses.

Changes from baseline (pre-treatment at Week -13) to Weeks 48, 72, and end of treatment (after week 48 = EOT) in the extension studies in efficacy outcomes are summarized descriptively and graphically using observed cases. These three time points were selected because (1) Week 48 reflects the approximate one year experience on maralixibat, avoids drug interruptions because of protocol administrative delays which affected 18 participants, and allows sufficient maralixibat drug exposure for participants originally randomized to placebo to respond; (2) Week 72 reflects the end of the original follow-up period in IMAGINE after which a number of participants declined further participation, and the middle of the second follow-up period in IMAGINE II; and (3) end of treatment reflects the experience of participants who have at least 48 weeks of treatment in IMAGINE or IMAGINE II and provides the most long-term measures of drug response.

To estimate and test the significance of the change from baseline to Week 48, we fit linear mixed effects models with random subject-specific intercepts and slopes for changes from baseline controlling for study (ITCH/IMAGINE II or IMAGO/IMAGINE), week (0, 2, 4, 8, 12, 24, 36, 48), age at baseline (years), and baseline level of the dependent variable for each efficacy outcome. Least square (adjusted) mean, standard error (SE), 95% confidence interval and p- value are reported for each outcome. Multiple imputation was used to address missing data. Ten imputed datasets were generated using the multivariate normal distribution method for continuous measures and the fully conditional specification method for categorical methods (5). The multiple imputation model included all characteristics at weeks -13, 0, 2, 4, 8, 12, 24, 36, and 48, study (ITCH/IMAGINE II or IMAGO/IMAGINE), age at baseline (years), and sex. For each linear mixed effects model, the results obtained from each imputed dataset were combined for the inference using Rubin's rule (6).

For the safety analysis, treatment-emergent adverse events (AEs), serious AEs, AEs resulting in early discontinuations, and AEs of special interest are summarized as the number of events, number of participants with at least one event, and rates (per person-year of follow-up) during the course of the studies. Treatment-emergent is defined as AEs where the start of the event is on or after the first dose date, and within 14 days of the last dose date. Potential impact on markers of liver injury were assessed by examining TB and ALT over time and by the application and modification of evaluation of Drug-Induced Serious Hepatotoxicity (DISH) plots (7). To address the large number of participants in the studies with elevated TB and ALT at baseline, we developed a novel DISH approach which uses change in peak ALT vs the corresponding change in TB or the change in peak TB vs the corresponding change in ALT.

Three different approaches to DISH plots were used, 1) eDISH based upon peak ALT or TB levels as measured by multiples of the upper limit of normal, 2) mDISH based upon peak ALT or TB as measured by multiples of the baseline values of ALT or TB and 3) hDISH based upon changes in ALT or TB as measured as multiples of the upper limit of normal above the baseline values of ALT or TB. To address the relationship of pruritus measures and serum bile acid scatterplots and Spearman correlations (with p-value based on Fisher's z transformation) were provided.

We performed inferential analyses on the effect of maralixibat only through Week 48 because of the validity and completeness of the data; however, no power or sample size calculations were performed and thus conclusions relied on clinical importance and statistical significance. We used descriptive methods after Week 48 because of the drug interruptions and attrition. Thus, these results should be interpreted with caution and considered exploratory.

Mean ± standard deviation (SD) are reported unless otherwise specified. Analyses were performed using SAS 9.4 (Cary, NC) or R 4.0.5 (Vienna, Austria).

References

1. Shneider BL, Spino C, Kamath BM, Magee JC, Bass LM, Setchell KD, Miethke A, et al. Placebo-Controlled Randomized Trial of an Intestinal Bile Salt Transport Inhibitor for Pruritus in Alagille Syndrome. Hepatol Commun 2018;2:1184-1198.
2. Kamath BM, Abetz-Webb L, Kennedy C, Hepburn B, Gauthier M, Johnson N, Medendorp S, et al. Development of a Novel Tool to Assess the Impact of Itching in Pediatric Cholestasis. Patient 2018;11:69-82.
3. Kamath BM, Spino C, McLain R, Magee JC, Fredericks EM, Setchell KD, Miethke A, et al. Unraveling the Relationship Between Itching, Scratch Scales, and Biomarkers in Children With Alagille Syndrome. Hepatol Commun 2020;4:1012-1018.

4. Shneider BL, Magee JC, Bezerra JA, Haber B, Karpen SJ, Raghunathan T, Rosenthal P, et al. Efficacy of fat-soluble vitamin supplementation in infants with biliary atresia. Pediatrics 2012;130:e607-614.

5. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007;16:219-242.

6. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York City: John Wiley & Sons, 1987.

7. Merz M, Lee KR, Kullak-Ublick GA, Brueckner A, Watkins PB. Methodology to assess clinical liver safety data. Drug Saf 2014;37 Suppl 1:S33-45.