7 February 2024
EMA/OD/0000123138
EMADOC-1700519818-1188310
Committee for Orphan Medicinal Products

Orphan Maintenance Assessment Report

Bylvay (odevixibat)
Treatment of Alagille syndrome
EU/3/12/1040

Sponsor: Albireo AB

Note

Assessment report as adopted by the COMP with all information of a commercially confidential nature deleted.

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Table of contents

1. Product and administrative information

Product	
Designated active substance(s)	(2S)-2-{[(2R)-2-[({[3,3-dibutyl-7-(methylthio)-1,1-dioxido-5-phenyl-2,3,4,5-tetrahydro- 1,2,5-benzothiadiazepin-8-yl]oxy}acetyl)amino]-2-(4-hydroxyphenyl)acetyl]amino}butanoic acid
Other name(s)	Bylvay
International Non-Proprietary Name	Odevixibat
Tradename	Bylvay
Orphan condition	Treatment of Alagille syndrome
Sponsor's details:	Albireo AB Arvid Wallgrens Backe 20 Goteborgs Annedal 413 46 Goteborg Vastra Gotalands Lan Sweden
Orphan medicinal product designation procedural history	
Sponsor/applicant	Albireo AB
COMP opinion	18 July 2012
EC decision	9 August 2012
EC registration number	EU/3/12/1040
Type II variation procedural history	
Rapporteur / Co-rapporteur	Johann Lodewijk Hillege / Jayne Crowe
Applicant	Albireo AB
Application submission	29 November 2022
Procedure start	31 December 2022
Procedure number	EMEA/H/C/004691/II/0011
Invented name	Bylvay
Proposed therapeutic indication	Bylvay is indicated for the treatment of cholestasis and pruritus in Alagille syndrome (ALGS) in patients aged 6 months or older. Further information can be found in the European public assessment report (EPAR) on the Agency's website https://www.ema.europa.eu/en/medicines/human/EPAR/Bylvay
CHMP opinion	20 July 2023
COMP review of orphan medicinal product designation procedural history	
COMP rapporteur(s)	Zsofia Gyulai / Joao Rocha
Sponsor's report submission	13 January 2023
COMP discussion and adoption of list of questions	13-15 June 2023
Oral explanation	11 July 2023
COMP opinion (adoption via written procedure)	31 July 2023
Appeal to the COMP opinion procedural history	
COMP rapporteur	Elisabeth Johanne Rook / Ines Alves

Appeal submission	15 September 2023
Appeal oral explanation	3 October 2023
COMP final opinion	5 October 2023

2. Grounds for the COMP opinion

Orphan medicinal product designation

The COMP opinion that was the basis for the initial orphan medicinal product designation in 2012 was based on the following grounds:

"Whereas, the Committee for Orphan Medicinal Products (COMP), having examined the application, concluded:

- Alagille syndrome (hereinafter referred to as "the condition") was estimated to be affecting not more than 0.3 in 10,000 persons in the European Union, at the time the application was made; this is based on a literature search conducted by the sponsor.
- The condition is chronically debilitating due to hepatic and cardiac dysfunction. Pruritus is a common symptom associated with cholestasis. Portal hypertension develops in up to one third of patients. In 75% of patients life expectancy is around 20years and death is associated with liver failure, cardiac problems and blood vessel abnormalities.
- There is, at present, no satisfactory method of treatment that has been authorised in the European Union for patients affected by the condition.

The COMP recommends the designation of this medicinal product, containing (2S)-2-{[(2R)-2-[({[3,3-dibutyl-7-(methylthio)-1,1-dioxido-5-phenyl-2,3,4,5-tetrahydro- 1,2,5-benzothiadiazepin-8-yl]oxy}acetyl)amino]-2-(4-hydroxyphenyl)acetyl]amino}butanoic acid, as an orphan medicinal product for the orphan indication: treatment of Alagille syndrome".

3. Review of criteria for orphan designation at the time of type II variation

Article 3(1)(a) of Regulation (EC) No 141/2000

Intention to diagnose, prevent or treat a life-threatening or chronically debilitating condition affecting not more than five in 10 thousand people in the Community when the application is made

Condition

Condition Alagille syndrome is a highly heterogenous, autosomal dominant multisystem condition and is caused by mutations in one of two genes: JAG1 and NOTCH2. It was initially described as a hepatic disease, but molecular testing has shown that individuals with ALGS and JAG1 or NOTCH2 mutations may present without overt liver disease.

Table 1. Summary of molecular genetic testing used in ALGS

Gene	Proportion of ALGS attributed to mutation of this gene	Test method	Mutations detected
JAG1	89%	Sequence analysis/mutation scanning	Sequence variants
	~5%–7%	Deletion/duplication analysis (including FISH and MLPA)	Deletion and duplication of exon(s) and entire gene deletion
NOTCH2	1%–2%	Sequence analysis	Sequence variants
	Unknown	Deletion/duplication analysis	Unknown, none reported

ALGS is a multisystem disorder with a wide spectrum of clinical signs and symptoms ranging from lifethreatening liver or congenital cardiac defects to only subclinical manifestations, such as mildly abnormal liver enzymes, a heart murmur, butterfly vertebrae, posterior embryotoxon (a thickening of the Schwalbe's line of the cornea), or characteristic facial features. This variability is present even among individuals from the same family sharing the same mutation.

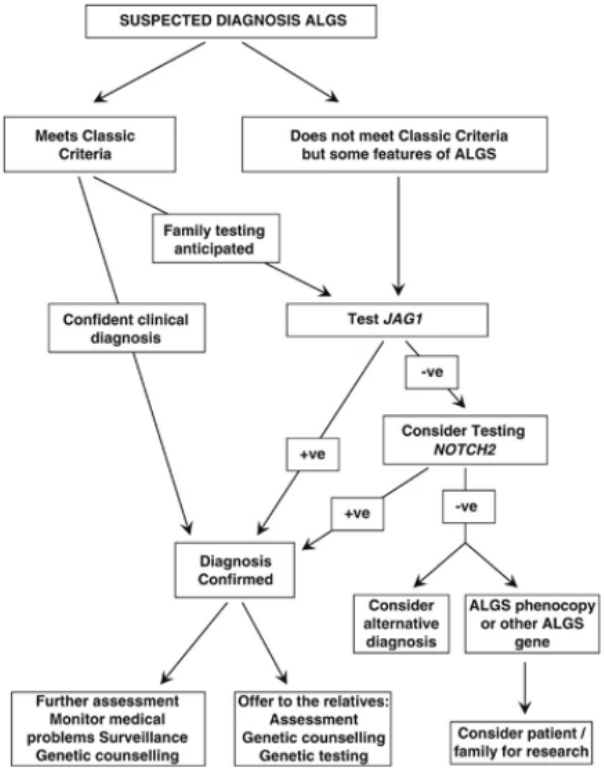
Table 2. A summary of the clinical features and the frequency reported among individuals with ALGS

Common system involved in ALGS	Feature	Overall frequency in ALGS	Frequency of finding in *JAG1* (+) ALGS	Frequency of finding in *NOTCH2* (+) ALGS
Hepatic	Paucity of biliary duct, conjugated hyperbilirubinemia, and liver failure	Up to 100%	100%	100%
Cardiac	Structural changes, pulmonary stenosis, and tetralogy of Fallot	90%–97%, 60%–67%, and 7%–16%	100%	60%
Facial features	Prominent forehead, deep-set eyes with moderate hypertelorism, pointed chin, and saddle or straight nose with a bulbous tip	20%–97%	97%	20%
Eye	Posterior embryotoxon	78%–89%	75%	60%
Skeletal	Vertebral anomalies (hemivertebra and butterfly vertebra)	33%–93%	64%	10%
Renal	Ureteropelvic obstruction and renal tubular acidosis	39%	40%	40%

In the majority of cases, individuals with ALGS present in infancy with cholestasis (conjugated hyperbilirubinemia with high GGT, increased serum bile acids, and elevated cholesterol and triglycerides), which manifest as jaundice, intense pruritus, xanthomas (fatty deposits on the extensor surfaces), and failure to thrive due to fat malabsorption. Cardiac findings ranging from benign heart murmurs to significant structural defects occur in 90%–97% of individuals with ALGS. Pulmonic stenosis (peripheral and branch) is the most common cardiac finding (60-67%). The most common complex cardiac defect is tetralogy of Fallot, which is seen in 7%–16% of individuals. Other cardiac malformations include ventricular septal defect, atrial septal defect, aortic stenosis, and coarctation of the aorta (in order of decreasing frequency). The typical facial features (see Table 2 above) are almost universally present in ALGS due to JAG1 mutations. The typical facial features do not seem to be as prevalent in individuals with ALGS carrying a NOTCH2 mutation. The mortality is ~10%, with vascular accidents, cardiac disease, and liver disease being the most frequent cause of death (The Application of Clinical Genetics 2016:9 75–82). Alagille syndrome is diagnosed when an individual has three out of seven major clinical features. See Figure 1 below for diagnostic framework. Individuals with an affected first-degree relative and who do not meet full clinical criteria but with the presence of one or more clinical features should be diagnosed with Alagille syndrome. Infants younger than 6 months of age may not present with a marked paucity of the bile ducts or even present with ductal proliferation that could lead to a misdiagnosis of biliary atresia. Traditionally, the clinical diagnostic criteria for ALGS included liver histology showing bile duct paucity (an increased portal tract-to-bile duct ratio) and three of five major clinical features: cholestasis; ophthalmologic abnormalities (commonly posterior

embryotoxon); characteristic facial features; cardiac defect (see Table 2); and skeletal abnormalities (commonly butterfly vertebrae). The five criteria have been increased to seven which are: cardiac defects, hepatic manifestation, renal abnormalities, skeletal abnormalities, ophthalmologic manifestations, dysmorphic facies and vasculature abnormalities. Bile duct paucity on liver histology is no longer considered mandatory for the diagnosis of Alagille syndrome, the presence of cholestasis can be used instead. (Alagille Syndrome Article - StatPearlshttps://www.statpearls.com › Article Library › view article 14 Aug 2022 https://www.statpearls.com/ArticleLibrary/viewarticle/17321).

Figure 1. Flow diagram of genetic investigations and management for suspected ALGS patients



European Journal of Human Genetics volume 20, pages251–257 (2012)

The approved therapeutic indication "*Bylvay is indicated for the treatment of cholestasis and pruritus in Alagille syndrome (ALGS) in patients aged 6 months or older"* falls within the scope of the designated orphan condition "treatment of Alagille syndrome".

Intention to diagnose, prevent or treat

The medical plausibility has yet to be confirmed by the positive benefit/risk assessment of the CHMP.

Chronically debilitating and/or life-threatening nature

The condition is chronically debilitating and life threatening due to chronic cholestasis due to paucity of intrahepatic bile ducts, progressing to portal hypertension and liver failure. Pruritus (itching) is reported as the most bothersome symptom of ALGS across all ages by patients and caregivers (Kamath et al. 2018b), which is difficult to treat, leads to cutaneous mutilation, mood disturbances, disruption of sleep and school performance, and has negative impact on physical and psychosocial health (Elisofon et al. 2010; Kamath et al. 2015, Kamath et al. 2018b), as well as overall Quality of Life (Abetz-Webb et al. 2014). Twenty-one percent to 31% of patients require liver transplantation

during childhood, including approximately 50% of those diagnosed in infancy. Cardiac dysfunction also occurs due to cardiac defects which are reported in greater than 90% of patients and include peripheral pulmonic stenosis (60-67%), tetralogy of Fallot (7-16%), ventricular septal defect, atrial septal defect, aortic stenosis, and coarctation of the aorta. Associated with cardiac defects are vasculature abnormalities which when present, are often associated with neurovascular abnormalities such as aneurysms, Moyamoya syndrome, abnormalities in cerebral arteries, reno-vascular abnormalities, and middle aortic syndrome.

Intracranial bleeding can also occur. Additional involvement of other systems/organs contributes to the chronically debilitating nature of the disease, including butterfly vertebrae, posterior embryotoxon and/or anterior segment abnormalities of the eyes, pigmentary retinopathy, and dysplastic kidney. The reported mortality is ~10%, with vascular accidents, cardiac disease, and liver disease being the most frequent cause of death. In a retrospective analysis of 1,154 children from 25 countries with a clinically and/or genetically confirmed ALGS diagnosis the 18-year survival was 88.6% (Vandriel et al. 2020). Mortality is high in the very young age because of the concurrent severe cardiac malformations.

Number of people affected or at risk.

The sponsor has submitted a prevalence estimate based on the original designation. The reasoning is that no new data has been published since the initial submission.

There is still very limited information on the birth incidence and prevalence of ALGS. In 2014, the estimate by Danks in 1977 was revised to 1 in 30,000 to 50,000 live births owing to more recent evidence showing that a notable proportion of individuals with the JAG1 mutation would not have been diagnosed with ALGS on the basis of clinical characteristics alone [Leonard 2014]. This incidence has later been cited by Kamath et al [Kamath 2018 and 2020a], Ayoub et al [Ayoub 2020], and Vandriel et al [Vandriel 2023]. Two other publications have been identified, and neither include new analyses. Saleh et al [Saleh 2016] refer to Gene Reviews [Spinner 2019] and claim that the incidence is 1/30,000. However, the Gene Reviews on ALGS refer back to Saleh 2016 when claiming that the incidence is 1/30,000. It is likely that ultimately, they both refer to the same Kamath publication already cited in the initial application.

Table 3. Prevalence Resources used in the Initial Application

REFERENCE	BIRTH INCIDENCE	RESULTING PREVALENCE	COMMENT
Danks 1977	1/70,000	0.14/10,000	11 out of 790,385 children in Victoria, Australia from 1963-74
Powell 2003	1/27654	0.36/10,000	Neonatal testing in Birmingham, UK
Kamath 2003 Kamath 2010b	1/30,000	0.33/10,000	Extrapolation based on assumed underdiagnosis

A birth incidence of 0.3/10 000 was proposed by the sponsor. This gives a total of 123 ALGS patients born/year, corresponding to a yearly birth incidence of 0.003/10 000 (123/ (447 000 000x10 000)). (4.1 million babies born in the EU in 2020; https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Fertility_statistics)

Considering that most liver transplants occur in childhood, a conservative average life span of 70 years for patients with ALGS was used. Based on this, the prevalence can be calculated as follows using the incidence x duration formula: 0.003/10 000 x 70 years = 0.2/10 000.

Despite the lower number achieved (0.2) taking life expectancy into account, a final estimate of 0.3 in 10,000 is proposed by the sponsor to be the current prevalence in the EU. The COMP accepted the proposed prevalence of 0.3 in 10,000.

Article 3(1)(b) of Regulation (EC) No 141/2000

Existence of no satisfactory methods of diagnosis prevention or treatment of the condition in question, or, if such methods exist, the medicinal product will be of significant benefit to those affected by the condition.

Existing methods

The sponsor has identified one centrally authorised product: maralixibat (Livmarli). Livmarli is indicated for the treatment of cholestatic pruritus in patients with Alagille syndrome (ALGS) 2 months of age and older. As Livmarli covers the same patient population as Bylvay, it is considered a satisfactory method and will have to be discussed in the significant benefit section.

It was noted by the COMP that off-label use with the following products has been reported: ursodeoxycholic acid, cholestyramine, rifampicin and naltrexone. These products are used in the management of pruritus and xanthoma but generally have limited success (Kronsten et al. J Pediatr Gastroenterol Nutr. 2013 Aug;57(2):149-54). Surgical interventions are used to treat cholestatic disease in ALGS and include surgical biliary diversion (SBD) procedures and liver transplantation. SBD procedures may not be fully effective and can lead to electrolyte disturbances and dehydration (Emerick et al. 1999; Kamath et al. 2018). Liver transplantation is associated with considerable safety risks, and donors are limited available.

There are no formal European Guidelines regarding the management and treatment of these patients.

Significant benefit

Protocol Assistance was sought by the sponsor but no question on significant benefit was asked as an approved product was only recently authorised (Livmarli, December 2022).

Significant benefit based on a clinically relevant advantage of Bylvay over Livmarli

To establish significant benefit a Matching-Adjusted Indirect Comparison or MAIC between maralixibat and odevixibat has been submitted.

A comparison of the study design features of the pivotal odevixibat and maralixibat studies is provided in Table 4.

Table 4. Study design features of the pivotal odevixibat and maralixibat

Feature	Study A4250-012 Odevixibat	Study LUM001-304 Maralixibat
Number analysed for primary and key secondary efficacy endpoints	Serum bile acid: Odevixibat: 35 Placebo: 17 Pruritus Severity: Odevixibat: 35 Placebo: 17	Serum bile acid, primary analysis: Maralixibat: 5 Placebo: 10 Serum bile acid and pruritus severity secondary analysis: Maralixibat: 13 Placebo: 16

Study LUM001-304 was a Phase 2, long-term, open-label study with a randomised, double-blind, placebo-controlled drug withdrawal phase in patients with ALGS. The study was designed to evaluate the safety and efficacy of maralixibat and was comprised of an 18-week open-label run-in phase (6-week dose escalation and 12-week stable maralixibat dosing), followed by a 4-week randomised withdrawal (RWD) phase. To improve gastrointestinal tolerability, patients were required to dose escalate gradually over the initial 6-week treatment period from a starting dose of 14 µg/kg/day to 400 µg/kg/day. Following the RWD, patients were treated for 26 weeks in a stable-dosing period at doses up to 400 µg/kg/day, and in an optional long-term treatment period (long-term extension phase; LTE) [Gonzales 2021]. Placebo control was utilised in the 4-week RWD phase only. The RWD design has limitations due to the short-term treatment assessment period, evaluation of an enriched population, potential carry-over effects, and difficulty in extrapolation of the treatment effect to the general population. Overall, 31 patients within the age range of 1 and 15 years (inclusive) were treated in Study LUM001-304. The primary endpoint of the study was the difference in the change in mean fasting serum bile acids at the end of the RWD phase compared to Week 18 of open-label treatment (i.e. just prior to the RWD phase) as analysed in the modified intent to treat population (mITT). Because the mITT population was enriched, including only those participants who previously achieved a serum bile acid reduction of ≥ 50% during the open-label run-in phase, the primary efficacy analysis was limited to a small sample of the enrolled population, including only 15 of the 31 patients (placebo n=10, maralixibat n=5) [Gonzales 2021]. The key secondary efficacy endpoint was the change from Week 18 to Week 22 in pruritus severity as measured by the weekly morning average Itch Reported Outcome (Observer) (ItchRO[Obs]) score (AM score), maralixibat vs placebo, during the RWD phase. A total of 29 patients who completed Week 18 of the open-label period were included in the pruritus analysis.

In contrast, Study A4250-012 was a Phase 3, multicentre, multinational, randomised, double-blind, placebo-controlled, parallel group study conducted in 52 patients with a genetically confirmed diagnosis of ALGS. Randomisation was 2:1 (odevixibat:placebo). The study evaluated the efficacy and safety of odevixibat administered at a dose of 120 µg/kg/day compared to placebo for 24 weeks. The primary efficacy endpoint for the study was the change from baseline to Month 6 (Weeks 21-24) in average AM and PM scratching severity scores as measured by the Albireo ObsRO caregiver instrument in the full analysis set (FAS) comprised of all 52 randomised patients. The key secondary efficacy endpoint was the change from baseline to the average of Week 20 and Week 24 in serum bile acid levels in the FAS population.

For the indirect comparison of results from Study LUM001-304 and Study A4250-012, only unanchored comparisons could be performed due to the difference in study designs, including different durations of the placebo-controlled period (4 weeks vs 24 weeks, respectively) in each trial. The endpoints for comparison that were chosen depended on availability of published data from the maralixibat

Study LUM001-304 and the nearest equivalent data from Study A4250-012. Based on this, serum bile acids and pruritus score, which were assessed at Week 22 of maralixibat Study LUM001-304, were chosen as the most appropriate endpoints for comparison to the odevixibat Study A4250-012.

Clinically relevant advantage based on efficacy

At Week 22, a greater reduction from baseline in serum bile acids was observed for odevixibat compared with maralixibat with mean absolute change from baseline of -141.9 µmol/L and -101.7 µmol/L, respectively, and a difference of -40.2 µmol/L.

Similarly, at Week 22, treatment with odevixibat led to greater reductions in pruritis severity scores compared to maralixibat with LS mean differences of -0.39 for nighttime score, -0.23 for daytime score, -0.30 for average AM+PM score, and -0.18 for worst AM+PM score.

Table 5. Matching-Adjusted Indirect Comparison: Mean (95% CI) Changes from Baseline in Serum Bile Acids and Pruritus Severity Score to Week 22 (Study LUM001-304 and Study A4250-012)

Parameter	Maralixibat[a] (N=13)	Odevixibat (Adjusted)[b,c] (N=35)	Comparison
			Difference[d] (95% CI)
sBA (µmol/L)	-101.7 (-193.9, -9.4)	-141.9 (-184.3, -99.6)	-40.2 (-126.3, 45.8)
ItchRO(obs) or ObsRO scratching score			
Nighttime (AM) Score	-1.3 (-1.9, -0.8)	-1.69 (-1.99, -1.39)	-0.39 (-1.02, 0.24)
Daytime (PM) Score	-1.3 (-1.8, -0.8)	-1.53 (-1.84, -1.22)	-0.23 (-0.81, 0.36)
Average AM+PM Score	-1.3 (-1.8, -0.8)	-1.60 (-1.89, -1.30)	-0.30 (-0.87, 0.28)
Worst Daily (AM+PM) Score	-1.5 (-2.0, -1.0)	-1.68 (-1.96, -1.39)	-0.18 (-0.75, 0.40)

CI: confidence interval; sBA: serum bile acids; ItchRO: itch reported outcome-observer (maralixibat); LS: least square; ObsRO: observer-reported outcome (odevixibat)
[a] sBA: Gonzales, et al. 2021 (mean [95% CI]); ItchRO: CDER Integrated Review, Table 16 (LS mean [95% CI])
[b] Adjusted for baseline covariates: age, sex, sBA, ObsRO
[c] For odevixibat, data are the average of Weeks 20 and 24.
[d] Comparison is odevixibat – maralixibat; values < 0 favour odevixibat.

The sponsor concludes that the MAIC results support evidence of improved efficacy of odevixibat versus maralixibat. They argue that there were greater mean reductions in serum bile acid levels and pruritus severity scores were consistently seen in all endpoints, in favour of odevixibat. The between-group differences in mean values did not reach statistical significance that would indicate superiority; however, this result was expected given the sample sizes (n=13 for maralixibat and n=35 for odevixibat).

The COMP noted that unanchored MAICs come with limitations where it is assumed that the adjusted analysis accounts for all important prognostic and predictive factors.

For Livmarli the comparison includes all patients who were randomized to Livmarli (i.e. not restricted to responders during run-in phase as used for the primary efficacy analysis), which is reasonable.

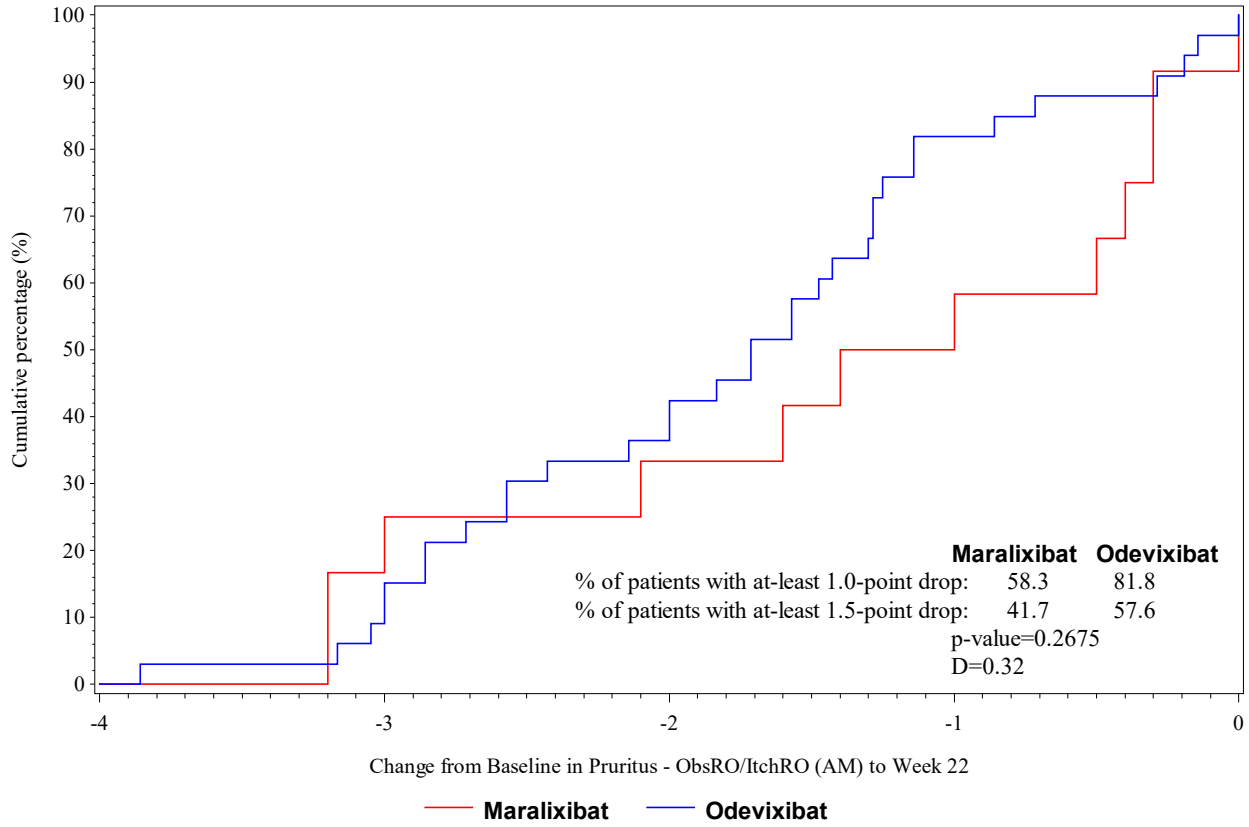
The unadjusted differences in mean serum bile acid at BL seem most prominent with higher values in the Livmarli (maralixibat) arm (mean of 318 vs 237.4). Adjusted estimated differences for sBA favour Bylvay but are not significant. For the AM scores, different dimensions of pruritus were assessed in both trials which is a limitation for comparability. The COMP noted that there was missing information on the **effective sample sizes** following adjustment. In addition to the MAIC, naive unweighted indirect comparisons using available IPD were conducted using data from the maralixibat-treated

patients in Study LUM001-304 and the odevixibat-treated patients in Study A4250-012. This was done for pruritus severity, serum bile acids and xanthomas, but in this report only the severity of pruritus is commented upon.

- *Comparison of Changes in Pruritus Severity*

As an initial investigation of the efficacy of odevixibat compared to maralixibat, an evaluation of the change in pruritus severity across the 2 studies was conducted. Specifically, an indirect comparison of naïve unweighted data based on IPD from both trials was performed using empirical distribution functions for change from baseline in the maralixibat ItchRO(Obs) using published data from Study LUM001-304 by Gonzales, et al [Gonzales 2021] and the odevixibat ObsRO scratching severity weekly average AM score at Week 22; the results were plotted and analysed (Figure 2). A clear upward separation from -2.5 to -0.5 of the odevixibat group strongly supported better efficacy. Although the maximum proportional difference of 0.32 in favour of odevixibat was impressive, the two-sample Kolmogorov–Smirnov test failed to reject the null hypothesis; however, this result was unsurprising given the small number of data points in the maralixibat sample (n = 13).

Figure 2. Empirical Distribution Functions for Change from Baseline to Week 22 in Pruritus Severity (AM Score) (Study LUM001-304 and Study A4250-012)



ItchRO(Obs): itch reported outcome, observer; ObsRO: observer-reported outcome Note: p-value from exact Komlmogorov-Smirnov (KS) test; D is the maximum absolute difference in the empirical cumulative distribution function between maralixibat and odevixibat.
Sample size: maralixibat n = 13; odevixibat n = 35

Based on the results of the empirical distribution function, further analyses were conducted. Table 6 presents the results for changes from baseline to Week 22 in pruritus severity AM score for the ItchRO(obs) as reported for maralixibat and the ObsRO as reported for odevixibat; results were consistent with the MAIC analyses.

Baseline pruritus severity score was similar for odevixibat and maralixibat with mean (SD) of 2.94 (0.572) and 2.87 (0.534), respectively. At Week 22, a greater reduction from baseline in pruritus severity score was observed for odevixibat compared to maralixibat with LS mean changes from baseline of -1.76 and -1.49, respectively, and an LS mean difference of -0.27.

Pruritus responder analyses were also conducted, assessing a ≥ 1-point reduction from baseline as conducted in the pivotal maralixibat study (Table 7; Figure 3). In this analysis, the percentage of patients achieving a ≥ 1-point reduction from baseline was higher for odevixibat compared to maralixibat with 81.8% (27 of 33) of odevixibat patients meeting this responder definition compared to 58.3% (7 of 12) of maralixibat patients (odds ratio [95% CI]: 3.04 [0.70, 13.18]). Although the odds ratio was robust at 3.04 favouring odevixibat, the lower bound of the 95% CI was just below 1.0 (at 0.7) indicating that the difference was not statistically significant (p = 0.1379). The difference would have been statistically significant if the maralixibat pivotal study had evaluated the same number of patients as the odevixibat study (n = 33), assuming a response rate of 58% similar to what was observed in the maralixibat study (Table 7) (chi-squared test p-value = 0.0321).

Table 6. Naive Unweighted Indirect Comparison: Change from Baseline to Week 22 in Nightime Scratching Severity (Average Weekly AM Score) from the ItchRO(Obs) and ObsRO (Study LUM001-304 and Study A4250-012)

Visit Statistic	Maralixibat (N=13)[a]	Odevixibat (N=35)[a]
Baseline		
n Mean (SD) Min, Max	13 2.87 (0.534) 1.9, 3.7	35 2.94 (0.555) 2.0, 4.0
Week 22		
n Mean (SD) Min, Max	12 1.38 (0.935) 0.1, 2.8	33 1.15 (0.977) 0.0, 3.1
LS Mean Change[b]		
n LS mean (SE) 95% CI	12 -1.49 (0.281) -2.1, -0.9	33 -1.76 (0.169) -2.1, -1.4
LS mean difference[c] (SE) 95% CI p-value[d]	-0.27 (0.328) -0.94, 0.39 0.4100	

CI: confidence interval; FAS: full analysis set; ItchRO(Obs): itch reported outcome, observer; LS: least square; ObsRO: observer-reported outcome; RWD: randomised withdrawal; SD: standard deviation; SE: standard error
a Maralixibat analysis based on 13 patients who received maralixibat through Week 22 (i.e. continued active treatment through the RWD); odevixibat analysis based on 35 patients randomised into study A4250-012 (FAS).
b From an analysis of covariance model with baseline pruritus AM score as covariate and treatment group as fixed effect.
c (Odevixibat – Maralixibat)
d Two-sided p-value vs maralixibat

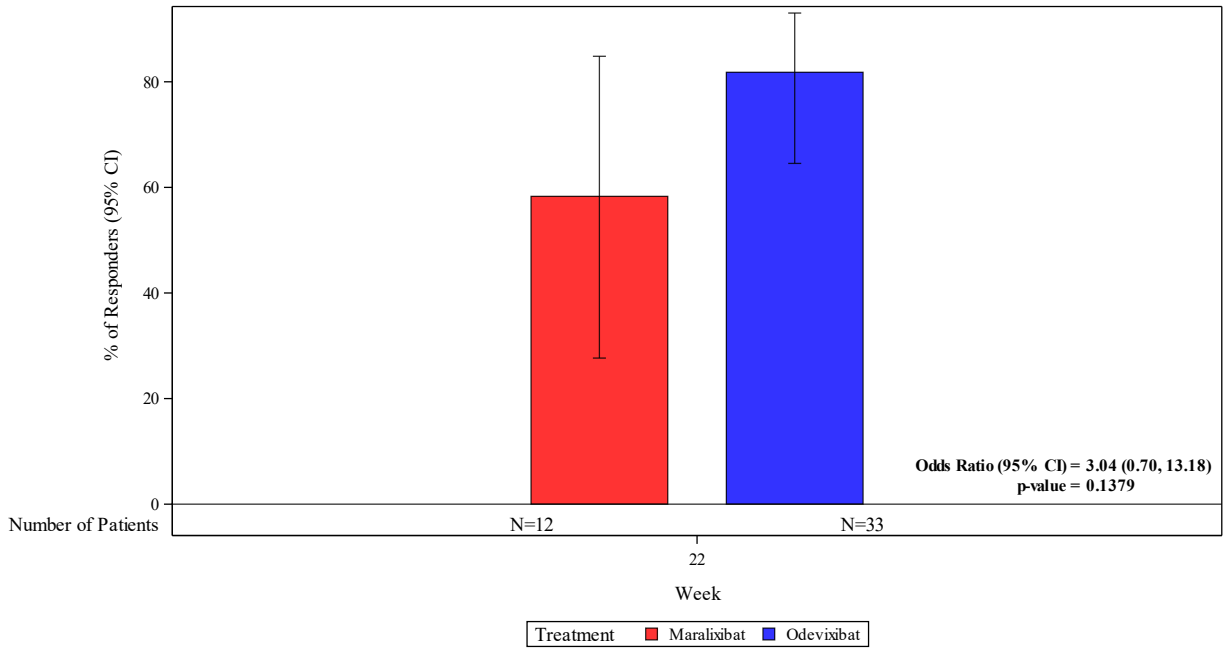
Table 7. Naïve Unweighted Indirect Comparison: Percent of Patients Achieving a ≥1.0-point Reduction in Nighttime Scratching Severity (AM Score) from Baseline to Week 22 (Study LUM001-304 and Study A4250-012)

Parameter Statistic	Maralixibat (N=13)[a]	Odevixibat (N=35)[a]
At least a ≥1.0-point drop, n/N1[b] (%)	7/12 (58.3)	27/33 (81.8)

Parameter Statistic	Maralixibat (N=13)[a]	Odevixibat (N=35)[a]
Odds Ratio (95% CI)[c] p-value[d]	3.04 (0.70, 13.18) 0.1379	

CI: confidence interval; FAS: full analysis set; RWD: randomised withdrawal
a Maralixibat analysis based on 13 patients who received maralixibat through Week 22 (i.e. continued active treatment through the RWD); odevixibat analysis based on 33 patients randomised into study A4250-012 (FAS) with assessments at Week 22.
b n is number of patients who responded, N1 is number of patients with non-missing data
c The confidence interval and two-sided p-value are from the logistic regression that models the probability of being a responder versus not with treatment and baseline as explanatory variables.

Figure 3. Comparison of the Percent of Patients Achieving a ≥ 1-Point Reduction in Nighttime Scratching Severity (AM Score) from Baseline to Week 22: Maralixibat vs Odevixibat



CI: confidence interval

A post-hoc analysis of pruritus control, defined as the proportion of days with an ItchRO(Obs) AM score (representing nighttime scratching) of 0 or 1 in any given week, was conducted for maralixibat (EPAR, Section 2.6.5.2). This analysis was also conducted for odevixibat based on data from Study A4250-012. At Week 22, pruritus control was reported for a higher percent of days for patients who were receiving odevixibat in Study A4250-012 (mean [95% CI] 68.2% [53.1%, 83.4%]) compared to patients who were receiving maralixibat in Study LUM001-304 (mean 41.7%).

The COMP concluded that although a trend for higher efficacy of Bylvay could be seen, no clear conclusions could be drawn based on the available data. The indirect comparison was hampered by the comparator product data (small sample size, different trial design). The Committee believed that even if the sponsor was to provide additional information on the MAIC (e.g. on effective sample sizes), it would unlikely change the conclusion that the efficacy seems to be similar for both products.

Comparison of safety profiles to support a clinically relevant advantage

In clinical development, odevixibat and maralixibat exhibited qualitatively similar safety profiles. However, the tolerability profile related to gastrointestinal adverse events was better for odevixibat relative to maralixibat as outlined in Table 8. During treatment with maralixibat, gastrointestinal

adverse events were noted in 71% of the 31 patients during the 18-week open-label period, including 42% of patients with diarrhoea, 39% with abdominal pain, and 36% with vomiting, even using a dose titration scheme to reach the planned dose of 400 µg/kg/day. This is in contrast to the odevixibat study where, even over a considerably longer duration of treatment of 24 weeks, the overall incidence of gastrointestinal adverse events was much lower at 34% as was the incidence of diarrhoea (29%), abdominal pain (11%), and vomiting (6%).

Additionally, there were no discontinuations due to adverse events during the 24-week treatment period in the odevixibat Study A4250-012, whereas in the maralixibat Study LUM001-304, discontinuations due to adverse events were reported in 2 patients in the 18-week run-in period, 1 patient after the randomised withdrawal period, and 2 additional patients in the long-term extension [Gonzales 2021]. Following completion of Study A4250-012, 50 of 52 (96%) patients elected to enroll in the odevixibat long-term extension Study A4250-015, whereas 23 of 28 (82%) patients entered the long-term extension in the maralixibat Study LUM001-304.

Table 8. Naïve Unweighted Indirect Comparison: Treatment-emergent Gastrointestinal Disorders (Safety Analysis Set, Study A4250-012 and Study LUM001-304)

	Study A4250-012 (24-week study)		**Study LUM001-304 (through Week 18)**
MedDRA SOC Preferred Term	**Placebo (N=17) n (%)**	**Odevixibat (N=35) n (%)**	**Maralixibat (N=31) n (%)**
Gastrointestinal disorders	**2 (11.8)**	**12 (34.3)**	**22 (71.0)**
Diarrhoea	1 (5.9)	10 (28.6)	13 (41.9)
Abdominal pain	1 (5.9)	4 (11.4)	12 (38.7)
Vomiting	1 (5.9)	2 (5.7)	11 (35.5)

MedDRA: Medical Dictionary for Regulatory Activities; SOC: system organ class

Of note, odevixibat (invented name Bylvay) was authorised on 16 July 2021 for the treatment of progressive familial intrahepatic cholestasis (PFIC) in patients aged 6 months or older. In total, approximately 640 patients with cholestatic liver disease have been exposed to odevixibat to date during clinical development and post-marketing. No new safety signals have been detected and the positive benefit-risk balance has been confirmed in the last Periodic Safety Update Report (PSUR) reporting period. Thus, the overall treatment experience of odevixibat is broader than that of maralixibat. The COMP noted that this additional post marketing data was not discussed in the comparison to maralixibat.

The sponsor provided a **naïve side-by-side comparison** showing some AEs between placebo vs. Bylvav for the Bylvay trial until week 24 versus AEs for Livmarli during the 18-week run-in period (i.e. before initiation of the RWD and without a PBO control for this timeframe).

This being a naïve comparison, there is no attempt to adjust for differences in trial populations and the COMP considered that the observed differences could be caused by differences in trial populations or conduct. The lack of a placebo control for the chosen comparison from the Livmarli trial makes it difficult to interpret. In addition, there is the limitation of the relatively small sample sizes.

The treatment duration for Bylvay is longer (24 vs 18 weeks) which in relation to the AEs observed in this timeframe should bias results in favour of Livmarli as AEs would only accrue with a longer exposure period. Setting aside all the afore mentioned limitations, the percentage of gastrointestinal AEs observed for Bylvay as compared to Livmarli looks smaller (34% vs 71%). However, due to the

very small number of patients this cannot be considered as convincing data in support of the significant benefit at this point in time. The COMP considered that additional safety data such as that in the PSUR the sponsor refers to should be included in the comparison to understand if indeed there is a clinically relevant difference between the two products regarding gastrointestinal adverse (GI) events. The sponsor should also not only look at the GI but a full comparison of all safety aspects of the two products will be required.

Significant benefit based on a major contribution to patient care

Improvement in sleep disturbance and daytime tiredness, and impact on quality of life

Baseline data in the pivotal odevixibat study, A4250-012, showed that 61% of caregivers reported that disturbance in their child's sleep was severe or very severe. As further evidence for the impact of sleep on QoL for patients with cholestatic liver disease, data collected during development of the Albireo clinical outcome assessment tools, the ObsRO and PRO instruments, sleep disturbance and fatigue were reported as the most bothersome features most commonly experienced by patients with ALGS and other cholestatic liver diseases. Sleep disturbance was mentioned by 92% of respondents and the average disturbance rating was 8.3 on a 0 to 10-point scale, where 10 indicates high levels of disturbance. Fatigue was reported in 69% of the interviews and the average disturbance rating was 7.8. Sleep disturbance and fatigue were the only impacts reported by more than 50% of the respondents. Improving sleep deficits and the resulting daytime tiredness is a clinically important treatment goal in patients with ALGS.

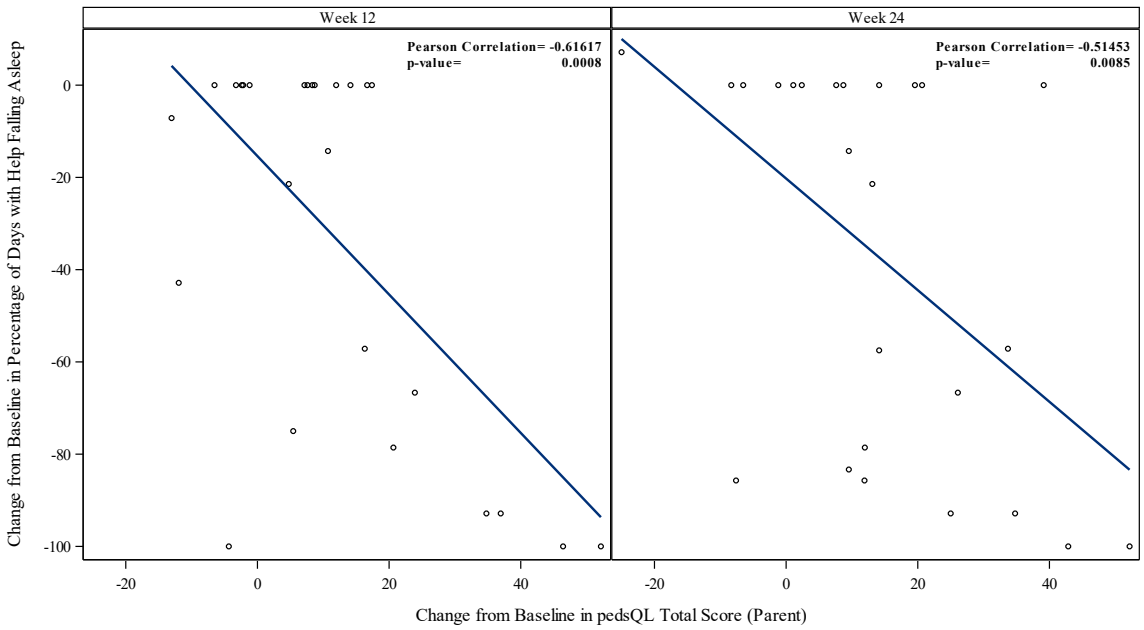
As highlighted in the results the odevixibat trial showed statistically significant improvements over placebo in sleep disturbance and tiredness; these improvements continued through the 24-week study and during longer-term treatment in the extension study. The improvement in sleep was confirmed by additional analyses on global symptom relief specifically related to sleep.

To assess the relationship between improvements in sleep disturbance and QoL as assessed in the pivotal odevixibat study A4250-012, post hoc correlation analyses were conducted. Specifically, the changes from baseline to Weeks 12 and 24 in the caregiver-reported sleep parameters and in changes in the PedsQL total score were evaluated.

Figure 4 presents the results of the correlation analysis for changes to Weeks 12 and 24 in the PedsQL total score and the sleep parameters, percent of days with help falling asleep. As shown, improvements from baseline in the 2 parameters were moderately correlated with R = 0.6162 for Week 12 and R = 0.5145 for Week 24. Similar correlations were observed with PedsQL total score and other sleep parameters.

The results from Study A4250-012 show that treatment with odevixibat led to statistically significant improvements over placebo in sleep disturbance and tiredness, and that sleep is correlated to the patients' QoL. Sleep was not assessed as an endpoint in the maralixibat programme. For maralixibat, QoL was noted to improve during open-label treatment (mean [95% CI] change from baseline in PedsQL total score: 11 [4, 17]) but then declined considerably during the 4week period once the patients and caregivers were blinded to treatment (mean [95% CI] change during the RWD in PedsQL total score: 8 [17, 1]) [Gonzales 2021]. The sponsor considered that odevixibat showed a major contribution to patient care over maralixibat.

Figure 4. Correlation Between Change from Baseline to Weeks 12 and 24 in Caregiver-Reported PedsQL Total Score and Percent of Days with Help Falling Asleep (Full Analysis Set, Study A4250-012)

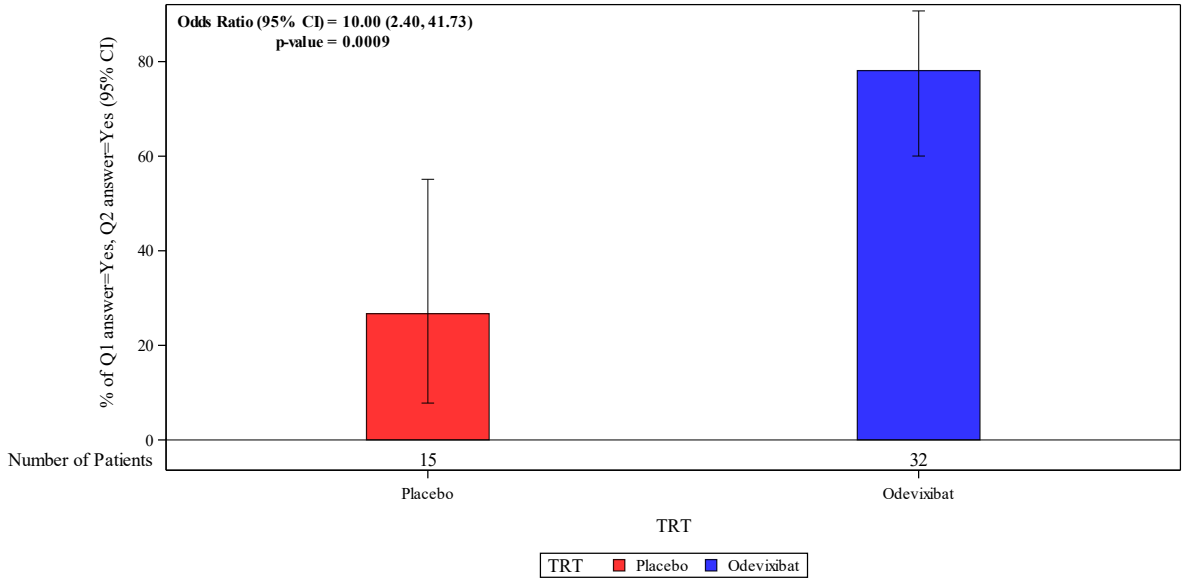


PedQL: Pediatric Quality of Life

Patient impact

An exit interview was conducted with caregivers to assess the impact of treatment on scratching and sleep. Results of this analysis showed that a statistically significantly higher proportion of caregivers of patients who received odevixibat reported meaningful improvements in the patient since the start of treatment (78.1%) compared to caregivers of patients who received placebo (26.7%) (p = 0.0009) (Figure 5).

Figure 5. Exit Survey Results: Caregiver Impression of Treatment Effect at Week 24 (Full Analysis Set, Study A4250-012)



CI: confidence interval
Note: p-value based on Cochran-Mantel-Haenszel test adjusting for baseline age stratification factor

According to the EPAR – public assessment report for Livmarli: "Improvement in the pruritus was accompanied with improvements in the quality of life, sleep and fatigue (mostly over the open-label long-term treatment period), which are also considered important favourable effects". The fact that a particular endpoint or symptom was not captured in the Livmarli trial does not automatically mean that Livmarli would not have an effect on them.

The COMP considered that the sponsor should further discuss the correlation between sleep quality and pruritus in the condition and justify the major contribution to patient care over Livmarli.

In conclusion

Based on the data submitted and the MACI, the COMP concluded that the Bylvay and Livmari appeared to have similar efficacy. Safety considerations require more safety data for both products as the naïve comparison was based on very small sample sizes form both trials. A comparison of the full safety profile is needed in order to discuss if there is a clinically relevant advantage based on any safety aspects. Finally, the COMP considered that additional discussion on the correlation between pruritus and sleep quality in the condition will be needed in order to justify a major contribution to patient care.

4. COMP list of issues

Significant benefit

It is known that establishing significant benefit based on safety generally requires convincing data from a sufficient number of patients and long follow-up period. The sponsor is requested to provide a comparison of the entire safety profile of Bylvay vs Livmarli and elaborate on the reasons to consider safety improvements as a clinically relevant advantage. Post-marketing surveillance data might be used in this comparison.

If the sponsor cannot establish a clinically relevant advantage, then major contribution to patient care may be claimed. The claim of improvement on sleep as a major contribution to patient care should be further elaborated. It is noted that there is no comparator to assess the impact of this in the clinical setting on pruritus and sleep quality.

Comments on sponsor's response to the COMP list of issues

The sponsor has provided a written response and presented their position at an oral explanation. The first point to be discussed was data to support the basis of a clinically relevant advantage. A safety comparison was submitted and presented. The methodology used for comparison of safety data focused on a statistical comparison of the incidence of adverse events between the Livmarli 18-week open label run-in phase and the Bylvay 24-week, randomised, double-blind, placebo-controlled study (Table 9) as well as post marketing surveillance data (Table 10) with a focus on gastrointestinal disorders.

Table 9. Incidence of adverse events between the Livmarli 18-week open label run-in phase and the Bylvay 24-week, randomised, double-blind, placebo-controlled study

	Study A4250-012		Study LUM001-304	Risk Difference
Adverse Events	Placebo (N=17) N (%)	Bylvay (N=35) N (%)	Livmarli (N=31) N (%)	Bylvay-Livmarli (95% CI)
OVERALL TEAE	**12 (70.6)**	**26 (74.3)**	**30 (96.8)**	**-22.5 (-40.25, -5.35)**
Discontinuations due to TEAE	0	0	2	-6.5 (-21.42,4.65)
Gastrointestinal disorders SOC	**2 (11.8)**	**12 (34.3)**	**22 (71.0)**	**-36.7 (-57.89, -11.02)**
Diarrhoea	1 (5.9)	10 (28.6)	13 (41.9)	-13.4 (-36.51, 10.71)
Abdominal pain	1 (5.9)	4 (11.4)	12 (38.7)	-27.3 (-48.22, -5.82)
Vomiting	1 (5.9)	2 (5.7)	11 (35.5)	-29.8 (-49.35, -8.93)

Table 10. Post marketing surveillance data with a focus on gastrointestinal disorders

Reaction	Bylvay (PFIC) (N=423)		Livmarli (ALGS) (N=600)	
	Number of Cases	Percent of Patients	Number of Cases	Percent of Patients
Overall AEs	**65**	**15.4%**	**464**	**77.3%**
Diarrhoea	9	2.13%	51	8.50%
Vomiting	7	1.65%	23	3.83%
Abdominal Pain Upper	3	0.71%	18	3.00%
Abdominal Pain	4	0.95%	15	2.50%
Abdominal Discomfort	0	0.00%	11	1.83%
Nausea	2	0.47%	8	1.33%

The sponsor has claimed in their presentation that there was a two-fold higher incidence of gastro-intestinal adverse events for Livmarli when comparing the trials as well as a lower incidence of overall adverse-events for Bylvay. Additionally, the sponsor indicates that dose titrations were required for Livmarli due to lack of GI tolerability. It was highlighted that diarrhoea and vomiting were more important with Livmarli and that in paediatric patients this can lead to rapid dehydration and electrolyte imbalance.

It was noted by the COMP that they have supplemented their descriptive initial naïve side-by-side comparisons with estimated risk differences and confidence intervals. However, those estimates remain naïve risk differences without an attempt to adjust for differences between trial populations. Hence, this is adding to the general uncertainties around the comparability of the safety events due to differences in study design – for Maralixibat coming from an open-label uncontrolled study phase, whereas for Odevixibat from a double-blind RCT. The Livmarli trial was not placebo controlled and had the added bias of including disease-specific GI effects. In addition, in Kamath B et al 2023 it was published that the majority of gastrointestinal adverse events occurred within the first 4 weeks of treatment and lasted less than a week, the majority of diarrhea and abdominal pain were mild to

moderate in severity and transient in nature. There were no discontinuations for diarrhea with up to greater than 5 years of treatment with Livmarli.

It was also noted that in the Livmarli SmPC for Alagille Syndrome that dose escalations were recommended. The starting dose is 40mcg/kg/day and it is stated that: *Improvement in pruritus and reduction of serum bile acid levels may occur gradually in some patients after initiating odevixibat therapy. If an adequate clinical response has not been achieved after 3 months of continuous therapy, the dose may be increased to 120mcg/kg/day (see section 4.4).* So, a level of dosing flexibility is evident and the sponsor's argument regarding lowering the dose purely due to important adverse gastro-intestinal events appears to be negated by the recommendations on dose escalation.

Post-marketing data has been also submitted and summarized above in Table 10. The sponsor claims that the postmarketing surveillance data shows substantially lower overall AEs and GI AEs reported as compared to Livmarli.

The COMP noted that the post-marketing data on safety appeared to have been pooled from all available post-authorisation data on AEs which could be found. It was noted however that the big limitation and fully outside of the sponsor's control was that the comparator product was approved very recently and there is little data. In addition, it was noted that due to the conditions being different PFIC for Bylvay and Alagille Syndrome for Livmarli the difference in dosing could have had an effect on the reporting rates.

The COMP described in the table below some of the challenges regarding the dosing and collection of adverse events.

Table 11. Posology of Bylvay for Alagille Syndrome

Dose escalation

Improvement in pruritus and reduction of serum bile acid levels may occur gradually in some patients after initiating odevixibat therapy. If an adequate clinical response has not been achieved after 3 months of continuous therapy, the dose may be increased to 120 mcg/kg/day (see section 4.4.).

Table 2 shows the strength and number of capsules that should be administered daily based on body weight to approximate a 120 mcg/kg/day dose, with a maximum daily dose of 7 200 mcg per day.

Table 2: Number of Bylvay capsules needed to achieve the nominal dose of 120 mcg/kg/day

Body weight (kg)	Number of 600 mcg capsules		Number of 1 200 mcg capsules
4 to < 7.5	**1**	or	N/A
7.5 to < 12.5	**2**	or	1
12.5 to < 17.5	**3**	or	N/A
17.5 to < 25.5	**4**	or	2
25.5 to < 35.5	6	or	**3**
35.5 to < 45.5	8	or	**4**
45.5 to < 55.5	10	or	**5**
≥ 55.5	12	or	**6**

Capsule strength/number in **bold** is recommended based on predicted ease of administration.

The COMP noted the difference in the number of Bylvay capsules needed to achieve the recommended dose of Livmarli and requested that the sponsor further elaborate on how the safety data compared at the time of collection noting the incidence of GI AEs in patients taking 120mcg/kg/day had not been addressed. The sponsor was unable to provide an adequate response to the question. It was considered that in the post marketing data there was therefore an important dose bias, i.e. the recommended dose for PFIC patients is 40mcg/kg/day (with the possibility of dose escalation up to

120mcg/kg/day, if an adequate clinical response has not been achieved after 3 months of continuous therapy), while the recommended dose for Alagile syndrome is 120mcg/kg/day (with the option of dose reduction to 40 mcg/kg/day if tolerability issues occur in the absence of other causes). This could be driven by the different indications for the two products.

It was therefore concluded that the argument on an improved safety profile does not appear to be strong enough as the indirect comparison between the two studies are not matched and that the post marketing data has not been considered sufficiently compelling due to potential differences in doses used of Bylvay and Livmarli making comparability difficult to ascertain (i.e. the equipotent dosing is unknown). It was also noted that diarrhoea is associated with the condition and is associated with the bile acid secretion. No clinically relevant advantage could be established based on safety.

As a clinically relevant advantage could not be established by the sponsor and efficacy and safety between the products was considered equivalent, major contribution to patient care was discussed.

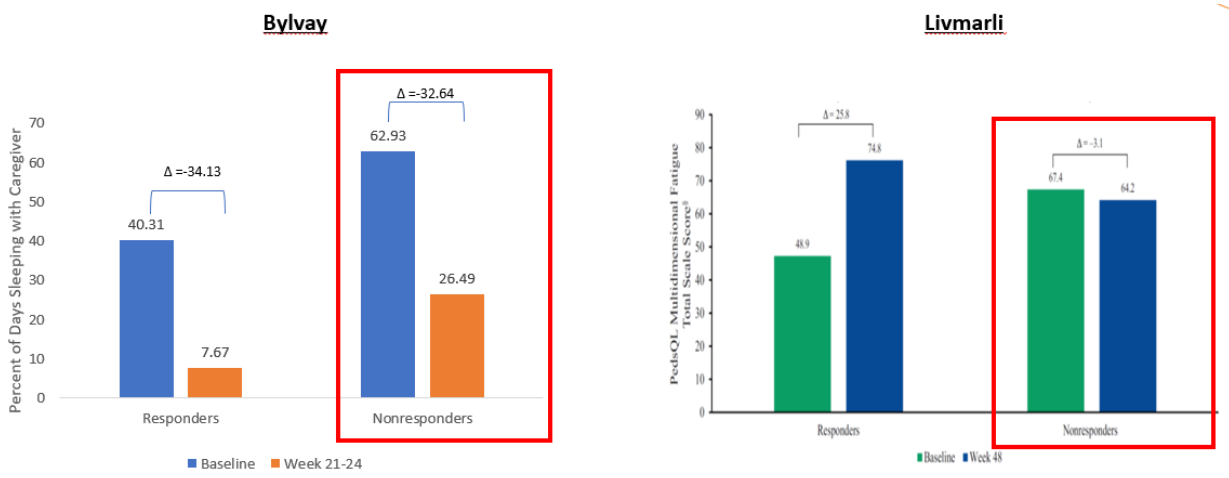
The sponsor has again compared the Bylvay Phase III study where sleep was a pre-specified endpoint, sleep disturbance and tiredness were evaluated with a validated PRO instrument and sleep was evaluated twice daily for the entire 24-week trial period. It is highlighted that Livmarli trial did not include sleep as a pre-specified endpoint, no sleep-specific instrument was used and only PedsQL generic score & Fatigue scale items were collected. They argue that in this instrument fatigue is not a direct measure of sleep disturbance and it is measured at baseline and 3 other timepoints over 48weeks.

Table 12. Bylvay Phase III study: evaluation of sleep disturbance and tiredness with dedicated PRO/ObsRO instrument

Sleep Parameter	Change from baseline to Weeks 21-24 Least square mean (SE)		One-sided p-value
	Placebo n=17	Odevixibat n=35	
Percent of days with help falling asleep	-10.1 (9.2)	-43.4 (6.7)	**0.0016**
Percent of days with soothing	-6.3 (7.9)	-46.7 (5.8)	**<0.0001**
Percent of days sleeping with caregiver	-8.3 (7.2)	-34.5 (5.4)	**0.0017**
Tiredness	-0.5 (0.2)	-1.1 (0.1)	**0.0062**

They also propose that the effect of Bylvay on sleep is not only due to its effect on pruritus. It is claimed that there is a lack of strong correlation between improvements in sleep disturbance and pruritus and pruritus responders and nonresponders to Bylvay have improvements in sleep. It is further claimed that the mechanism of sleep disturbance in chronic liver disease is not fully understood and is likely multifactorial (Shah, 2020).

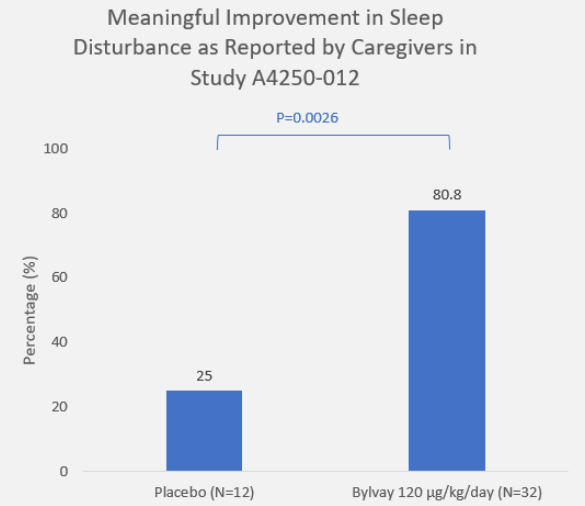
Figure 6. Bylvay Improved Sleep in Both Pruritus Responders and Nonresponders Livmarli improved fatigue in pruritus responders only



Change from Baseline to Month 6 in Percent Days Sleeping With Caregiver by Pruritus Responder

Change from Baseline in PedsQL Fatigue Score at Week 48 for Patients who Received Livmarli by Pruritus Responder

Figure 7. Patient and Caregiver Response in Exit Interviews of Bylvay Study Demonstrate that Improvement in Sleep Improves Quality of Life



Sample responses:

- "Sleeping through the night and falling asleep faster; more energy during day and wanting to participate in activities."
- "Reading has improved causing all of his grades to go up. He can focus on hobbies & is being nicer to his sibling because he feels better."
- "Sleeps through the night, falls asleep right away, wakes up with a lot of energy; better sleep, less itchy, better skin and hair, overall attitude is better."
- "Well rested and functions the same as other children of the same age."

The sponsor concludes that treatment with Bylvay led to a statistically significant and clinically meaningful improvement in sleep disturbance and tiredness because it was measured as an independent assessment with a validated instrument, improvements were observed in both pruritis responders and nonresponders and improvements in sleep disturbance correlated with improvement in QoL. They note that sleep was not assessed as an endpoint in the Livmarli study, no evidence that treatment with Livmarli would have a clinically meaningful and statistically significant impact on sleep; Fatigue is not equal to sleep. They also argue that extrapolation of improvement of sleep based on effects on pruritus is not supported.

The COMP noted that a comparison between two trials which are not matched has been made. This leads to methodological issues already highlighted to the sponsor in the initial evaluation regarding this unbalanced and unmatched indirect comparison which have not been addressed adequately. It was noted that the assumption that the comparison between responders and nonresponders in the Bylvay

trial was statistically robust was questionable as the nonresponder group consisted of only 7 patients while the responders had 28.

Table 13. Change from Baseline in Sleep Parameters (ObsRO) at Weeks 21-24 for Patients who Received Odevixibat by Pruritus Responder Status (Full Analysis Set, Study A4250-012)

SLEEP PARAMETER ASSESSMENT	ODEVIXIBAT (N=35)	
	Pruritus Responder (N=28) Mean (SD)	Pruritus Non-Responder (N=7) Mean (SD)
Percent of Days with Help Falling Asleep		
Baseline	58.90 (8.310)	80.27 (13.635)
Change to Week 21-24	-44.25 (7.916)	-21.83 (15.229)
Percent of Days with Soothing		
Baseline	67.55 (7.550)	83.50 (13.986)
Change to Week 21-24	-52.72 (7.554)	-25.43 (16.118)
Percent of Days Sleeping with Caregiver		
Baseline	40.31 (8.602)	62.93 (16.439)
Change to Week 21-24	-34.13 (8.200)	-32.64 (14.681)
Tiredness		
Baseline	2.22 (0.175)	1.83 (0.210)
Change to Week 21-24	-1.27 (0.197)	-0.31 (0.278)
Number of Awakenings		
Baseline	2.96 (0.474)	2.64 (0.593)
Change to Week 21-24	-2.21 (0.493)	-0.58 (0.500)

A more detailed analysis of the Livmarli QoL data revealed a similar trend.

Table 14. Change from baseline to week 48 HRQoL item scores, overall and stratified by ItchRO(Obs)* treatment response

Table VII. Change from baseline to week 48 HRQoL item scores, overall and stratified by ItchRO(Obs)* treatment response

		ItchRO(Obs)* treatment response at wk 48*		
	Overall (n = 27)	Responders (n = 20)	Nonresponders (n = 7)	*P* value
PedsQL Generic Core Scale				
Emotional: Trouble sleeping	40.74 ± 34.07	52.50 ± 27.98	7.14 ± 27.82	.001†
Physical: Problem participating in active play or exercise	12.96 ± 40.65	12.50 ± 38.47	14.29 ± 49.70	.92
School: Missing school/daycare to go to the doctor or hospital	13.16 ± 25.51	12.50 ± 27.39	16.67 ± 14.43	.80
Social: Problem playing with other children	0.00 ± 31.08	2.63 ± 33.22	−12.50 ± 14.43	.39
Social: Not able to do things that other children his/her age can do	6.52 ± 33.89	7.89 ± 35.41	0.00 ± 28.87	.68
PedsQL Family Impact Scale				
Communication: Hard to talk about child's health with others	14.42 ± 33.30	19.74 ± 32.89	0.00 ± 32.27	.19
Emotional: Feeling anxious	15.38 ± 42.47	17.11 ± 39.13	10.71 ± 53.73	.74
Physical: Feeling tired during the day	27.88 ± 31.88	32.89 ± 33.39	14.29 ± 24.40	.19
Social: Isolated from others	3.85 ± 34.42	6.58 ± 36.17	−3.57 ± 30.37	.52
Social: Hard to find time for social activities	13.46 ± 35.52	14.47 ± 39.37	10.71 ± 24.40	.82
Social: Not enough energy for social activities	13.46 ± 36.22	14.47 ± 41.93	10.71 ± 13.36	.82
Worry: Worried about child's future	15.38 ± 29.22	14.47 ± 31.53	17.86 ± 23.78	.80
PedsQL Multidimensional Fatigue Scale				
Cognitive: Difficulty keeping his/her attention on things	28.57 ± 34.72	32.35 ± 35.09	12.50 ± 32.27	.32
General: Feeling tired	26.19 ± 33.98	33.82 ± 26.43	−6.25 ± 47.32	.03†
Sleep/Rest: Sleeping a lot	7.14 ± 41.94	17.65 ± 35.09	−37.50 ± 43.30	.014†
Sleep/Rest: Difficulty sleeping through the night	42.86 ± 34.59	52.94 ± 30.47	0.00 ± 0.00	.003†
Sleep/Rest: Feeling tired when he/she wakes up in the morning	27.38 ± 42.50	41.18 ± 27.87	−31.25 ± 47.32	< .001†
Sleep/Rest: Resting a lot	15.48 ± 44.35	23.53 ± 39.99	−18.75 ± 51.54	.09
Sleep/Rest: Taking a lot of naps	20.24 ± 32.23	27.94 ± 30.47	−12.50 ± 14.43	.02†

Means and SDs are shown for continuous characteristics. *P* values are for the comparison of mean scores according to treatment response status and were calculated using t-test or ANOVA for continuous variables.
*ItchRO(Obs) is a clinical outcome assessment measure of pruritus, reported by the caregiver.
†Statistical significance.

The COMP however noted that a similar impact was reported in the public domain for Livmarli as presented in the graph below from a publication by Kamath et al J Pediatr 2023;252:68-75. The follow-up time was double that in the Bylvay study. (48weeks to 24weeks) The COMP questioned if the comparison was enough to establish major contribution to patient care.

The sponsor also just focuses on the naïve comparison of the two trials without considering what is published in the public domain.

It was also noted that the EPAR and OMAR for Livmarli discussed the improvements in quality of life, sleep and fatigue. In the OMAR a statistically significant reduction (improvement) in morning EDQ(Obs) sleep disturbance score between baseline and the average of weeks 15-26 and in PedsQL (Parent) multidimensional fatigue scale score between baseline and the average of weeks 18, 22 and 26 in the maralixibat group compared to placebo was reported.

The COMP noted the small number of patients in the non-responder group (7) in the Bylvay group and questioned the significance of the comparison with the responders (28) from study A4250-012 regarding the quality of the sleep. In addition, it was noted that the comparison between the major contribution to patient care regarding sleep trended more towards similarity for the two products although the methods used for measurement were different. The COMP could therefore not establish if the impact on sleep for Bylvay was indeed a significant major contribution to patient care when compared to Livmarli.

The COMP concluded they could not recommend maintaining the orphan designation and thus the 10-year market exclusivity for Bylvay in Alagille Syndrome.

5. COMP position adopted on 31 July 2023

The COMP concluded that:

- the proposed therapeutic indication falls entirely within the scope of the orphan condition of the designated Orphan Medicinal Product;
- the prevalence of Alagille syndrome (hereinafter referred to as "the condition") was estimated to remain below 5 in 10,000 and was concluded to be 0.3 in 10,000 persons in the European Union, at the time of the review of the designation criteria;
- the condition is life-threatening and chronically debilitating due to severe pruritus, cholestasis, liver failure, and congenital cardiac defects. Life expectancy is in most cases around 20 years and death is associated with blood vessel abnormalities, cardiac failure and end-stage liver disease;
- in view of the fact that a satisfactory method for the treatment of the condition has been authorised in the European Union (Livmarli), the existence of significant benefit over the authorised method of treatment should be established at the stage of the granting of marketing authorisation;
- the sponsor's claim that Bylvay is of significant benefit to those affected by the orphan condition does not hold. Significant benefit over Livmarli was claimed on the grounds of a clinically relevant advantage as well as a major contribution to patient care;
- the sponsor did not provide enough data to establish a clinically relevant advantage over Livmarli:
  - the products are deemed to have similar efficacy profiles, and differences noted by sponsor are not significant. Furthermore, sponsor's conclusions are hampered by MAIC issues;
  - in regard to safety, the products have shown similar safety profiles, and claims by the sponsor on significant differences are likewise hampered by the lack of robust data, biases and further relevant differences (e.g. dosing), etc;
- the sponsor did not provide enough data to demonstrate a MCPC in comparison to Livmarli. While the sponsor claims improvement of sleep, the same sponsor recognises that it has not provided data for Livmarli for the comparison, even though available data shows Livmarli is also beneficial for sleep. Furthermore, QoL data suggest similar effects, without any significant differences.

The COMP, having considered the information submitted by the sponsor and on the basis of Article 5(12)(b) of Regulation (EC) No 141/2000, is of the opinion that:

- the criteria for designation as set out in the first paragraph of Article 3(1)(a) are satisfied;
- the criteria for designation as set out in Article 3(1)(b) are not satisfied.

The Committee for Orphan Medicinal Products has recommended that Bylvay, (2S)-2-{[(2R)-2-[({[3,3-dibutyl-7-(methylthio)-1,1-dioxido-5-phenyl-2,3,4,5-tetrahydro- 1,2,5-benzothiadiazepin-8-yl]oxy}acetyl)amino]-2-(4-hydroxyphenyl)acetyl]amino}butanoic acid, odevixibat for treatment of Alagille syndrome (EU/3/12/1040) is removed from the Community Register of Orphan Medicinal Products.

6. Appeal to the negative opinion adopted on 31 July 2023

Grounds for appeal

The sponsor presented detailed grounds for appeal (EMA/OD/0000152080) on 15 September 2023.

Please refer to the sponsor's appeal documents in the case *Input from Industry* folder.

The detailed grounds for appeal were further addressed by the sponsor at an oral explanation before the COMP on 3 October 2023.

Comments on the grounds of appeal

The appeal grounds pertain to significant benefit claims of a clinically relevant advantage (i.e. improved efficacy and improved safety), as well as to a claim on a major contribution to patient care (MCPC). While these general claims have already been made during the initial submission, the sponsor presents new analyses as well as clarifications on certain aspects. The sponsor emphasizes that although the designs of the maralixibat and odevixibat registrational studies were different, a valid comparison was made by considering and adjusting for differences in baseline characteristics and other parameters that could bias the interpretation of the data. In addition, new analyses with longer term odevixibat treatment data are presented, including from the uncontrolled open-label extension study A4250-015.

In order to evaluate the significant benefit of odevixibat compared to maralixibat, a comparison was made of the publicly available results from Study LUM001-304, the primary study supporting the authorisation of maralixibat in ALGS, and the pivotal odevixibat study and its extension study, A4250-012 and Study A4250-015, respectively. Additional evidence was derived from a comparison of the safety profile of the 2 products from the placebo-controlled maralixibat Studies LUM001-301 and LUM001-303 with the placebo-controlled odevixibat Study A4250-012 and long-term safety data from maralixibat extension studies (LUM001-303 and LUM-001-305) and odevixibat study A4250-015. A comparison of post-marketing safety data from the FDA from both products is also included.

For the unanchored indirect adjusted and naïve comparisons presented by the sponsor, updated long-term exposure data for odevixibat have been obtained through a data cutoff date of 17 July 2023. Based on the new data cutoff date, 42 patients have now received odevixibat for ≥48 weeks through the pivotal study A4250-012 and the open-label extension study A4250-015. The sponsor considers that the re-assessment of the baseline characteristics using the larger sample size for odevixibat based on the updated data confirmed that the patient populations were similar across the studies. This argument is based on descriptively comparing the baseline characteristics of all patients from the pivotal and supportive odevixibat and maralixibit studies (Table 15).

Table 15. Comparison of demographic and baseline characteristics for relevant odevixibat studies and maralixibat studies

	ODEVIXIBAT			**MARALIXIBAT**	
	STUDY A4250-012		**POOLED STUDIES A4250-012/015 ODEVIXIBAT (N=52)**	**STUDY LUM001-304 MARALIXIBAT (N=31)**	**STUDIES LUM001-301 AND 303 MARALIXIBAT (N=57)**
CHARACTERISTICS[A]	**ODEVIXIBAT (N=35)**	**PLACEBO (N=17)**			
Mean (SD) Age, years	6.7 (3.78)	5.4 (4.41)	6.5 (3.98)	5.4 (4.2)	6.5 (4.6)
Male, n (%)	21 (60.0)	6 (35.3)	27 (51.9)	19 (61.3)	31 (54.4)
Use of UDCA, n (%)	30 (85.7)	16 (94.1)	46 (88.5)	25 (81)	N/A
Mean (SD) Weekly AM Average Pruritus Score	2.88 (0.535)	3.04 (0.610)	2.65 (0.761)	2.9 (0.5)	2.9 (0.7)a
Mean (SD) Serum Bile Acids, µmol/L	237.4 (114.88)	246.1 (120.53)	246.4 (130.39)	283 (211)	234.7 (202.3)
Mean (SD) PedsQL Total Score	76.6 (10.60)	67.1 (13.49)	70.1 (15.35)	59.4 (17.0)	62.7 (19.7)

PedsQL: Pediatric Quality of Life inventory; SD: standard deviation; UDCA: ursodeoxycholic acid. [a]The publication does not clarify what ItchRO(Obs) score this is, AM, PM or AM+PM. Data for Study LUM001-304 from Gonzales 2021 and Kamath 2022 and for Studies LUM001-301 and LUM001-303 from Shneider 2022.

The sponsor considered that both studies included clinically relevant endpoints (pruritus severity, serum bile acids, QoL, and AEs) assessed at similar timepoints allowing for meaningful comparisons to be conducted.

Table 16. Comparison of clinically relevant endpoints: odevixibat study A4250-012/A4250-015 and maralixibat study LUM001-304

ENDPOINT	STUDY A4250-012/A4250-015 ODEVIXIBAT	STUDY LUM001-304 MARALIXIBAT
Pruritus/ scratching	ObsRO COA: eDIARY assessment of scratching by caregiver twice daily (morning and evening) through Week 48 5-point scale from 0 (no scratching) to 4 (worst possible scratching)	ItchRo(Obs) COA: eDIARY assessment of scratching by caregiver twice daily (morning and evening) through Week 48 5-point scale from 0 (none observed or reported) to 4 (very severe)
Serum bile acids	Measureda at baseline and Weeks 4, 12, 20, 24, 36, and 48	Measuredb at baseline and Weeks 18, 22 and 48
PedsQL total score and the single item, trouble sleeping	Measureda at baseline and Weeks 12, 24, and 48	Measuredb at baseline and Weeks 18, 22 and 48
Adverse events	Throughout treatment	Throughout treatment

COA: clinical outcome assessment; ItchRo(Obs) itch reported outcome (observer); ObsRO: observer-reported outcome.[a]Additional assessments were conducted; those presented in the table are consistent from the start of odevixibat treatment across Studies A4250-012 and A4250-015.[b]Data available in the public domain; for Week 22 changes are from start of the randomized withdrawal period (end of Week 18), not baseline.

The sponsor further clarified that pharmacodynamic analyses confirmed that the doses studied in the registrational maralixibat (400 µg/kg/day) and odevixibat (120 µg/kg/day) studies were therapeutically equivalent, and that odevixibat is 4 to 10 times more potent than maralixibat and appears to be more efficient at reaching the IBAT target in the distal ileum, resulting in a lower drug burden.

6.1. Efficacy

The sponsor supports their claim for improved efficacy of Bylvay vis a vis Livmarli by the totality of data on QoL, pruritus severity, and serum bile acids from new analyses of patients who received at least 48 weeks of treatment with maralixibat or odevixibat. A regression-adjusted matching approach was chosen for the unanchored indirect comparisons between the pooled data from the odevixibat Studies A4250-012 and A4250-015 and the publicly available maralixibat data [Gonzales 2019, Gonzales 2021, Kamath 2023, Shneider 2018, Shneider 2022, maralixibat EPAR, the public version of the maralixibat OMAR, maralixibat FDA integrated review, clinicaltrials.gov, FDAERS].

6.1.1. Improvement in quality of life

6.1.1.1. Results of the indirect weighted adjusted comparisons

The sponsor conducted unanchored indirect comparisons of QoL based on the PedsQL total score for patients who received at least 48 weeks of treatment with maralixibat in Study LUM001-304 or odevixibat in Studies A4250-012/A4250-015. A regression-adjusted weighted estimator method was used for comparison of these data. The sponsor had initially considered performing a MAIC including baseline PedsQL in the weighting which however resulted in a very small effective sample size (ESS) with many odevixibat patients assigned a small weight for all outcomes. Consequently, for the QoL analysis, 4 baseline covariates were included in the MAIC (age, sex, AM scratching score, and serum

bile acid levels) and for the regression analysis, baseline PedsQL total score was additionally included as a covariate.

Table 17 provides the ESS and the unadjusted and adjusted baseline characteristics following propensity score weighting (MAIC) for QoL analyses.

Table 17. Effect of matching on baseline characteristics for final QoL analyses

BASELINE CHARACTERISTIC	**MARALIXIBAT LUM001-304**	**ODEVIXIBAT A4250-012/A4250-015 THROUGH 48 WEEKSA**	
		PEDSQL OUTCOMES	
		UNADJUSTED	**WEIGHTED[B]**
NUMBER OF PATIENTS	27	N = 30	ESS = 23.58
Mean (SD) age, years	5.7 (4.3)	6.77 (3.42)	5.70 (2.77)
Proportion of males (%)	66.7	50.0	66.7
Mean (SD) weekly AM scratching scorec	2.90 (0.56)	2.84 (0.79)	2.90 (0.74)
Mean (SD) serum bile acid, µmol/L	266.05 (213.86)	237.00 (120.42)	266.05 (133.65)

ESS: effective sample size; PedsQL: Pediatric Quality of Life instrument; QoL: quality of life; SD: standard deviation.aIncludes patients who received odevixibat for ≥48 weeks across Studies A4250-012 and A4250-015. bPropensity score adjusted for age, sex, race, baseline nighttime scratching score and baseline serum bile acid level. cBased on the ItchRO(Obs) for maralixibat and the ObsRO for odevixibat.

A summary of the unanchored regression-adjusted matched comparisons of changes from baseline to Week 48 in the QoL parameters is provided in Table 18. As shown, treatment with odevixibat led to a statistically significantly greater improvement in the PedsQL total score after 48 weeks of treatment compared to maralixibat. The mean weighted adjusted difference was 14.47 and the lower bound of the 95% CI was greater than zero (5.56, 23.38).

Table 18. Regression-adjusted matched comparisons: changes from baseline to week 48 in PedsQL total score (study LUM001-304 and studies A4250-012/A4250-015)

OUTCOME	**PARAMETER**	**POPULATION ADJUSTMENT APPLIED**	**ODEVIXIBAT VS. MARALIXIBAT [95% CI]**
QoL Endpoints			
Change from baseline to week 48 of PedsQL total score	Mean difference	Regression-adjusted MAIC[a]	**14.47 [5.56, 23.38]***

CI: confidence interval; MAIC: matching adjusted indirect comparison PedsQL: Pediatric Quality of Life instrument. *Statistically significant; 95% CI excludes 0.aPropensity scores based on age, sex, and baseline pruritus score and serum bile acid level; regression using age, sex, baseline pruritus score and serum bile acid level, and baseline PedsQL total score.

The statistically significant improvement in the PedsQL total score is deemed clinically relevant by the sponsor. Given that the minimal clinically important difference (MCID) for the PedsQL total score is 4.5 points [Varni 2003], this difference of 14.5 points in favour of odevixibat is >3x the magnitude of change that patients perceive as beneficial and the lower bound of the 95% CI exceeded the MCID.

The sponsor also emphasizes that indirect naïve comparisons over time showed improvements in QoL throughout both the blinded (A4250-012) and open-label (A4250-015) treatment with odevixibat while for maralixibat the initial improvement in QoL during the 18-week open-label treatment period in Study

LUM001-304, QoL declined substantially once patients were transitioned to blinded treatment in the short 4-week randomised withdrawal (RWD) period. QoL then improved through Week 48 upon return to open-label treatment with maralixibat.

6.1.2. Improvement in pruritus and serum bile acids

6.1.2.1. Results of the indirect weighted adjusted comparisons

Indirect weighted adjusted analyses for the key efficacy results for the pruritus and serum bile acid endpoints from the maralixibat LUM001-304 study with comparable data from the odevixibat A4250-012/A4250-015 studies were conducted for patients who received at least 48 weeks of treatment with maralixibat or odevixibat. The sponsor had initially considered performing a MAIC including baseline total bilirubin in the weighting which however resulted in a very small ESS with many odevixibat patients assigned a small weight for all outcomes. Consequently, a regression-adjusted matching method was used for comparison of these data (Table 19). For the pruritus analyses, 4 baseline covariates were included in both the MAIC and the regression analysis (age, sex, AM scratching score, and serum bile acid levels). For serum bile acids, the same 4 factors were included in the MAIC and the regression analysis; baseline total bilirubin was additionally included as a covariate in the regression analysis.

Table 19. Effect of matching on baseline characteristics for final pruritus and serum bile acid analyses

BASELINE CHARACTERISTIC	**MARALIXIBAT LUM001-304**	**ODEVIXIBAT A4250-012/A4250-015 THROUGH 48 WEEKS[A]**			
		PRURITUS OUTCOMES		**SERUM BILE ACID OUTCOME**	
		UNADJUSTED	**WEIGHTED[B]**	**UNADJUSTED**	**WEIGHTED[B]**
NUMBER OF PATIENTS	27	36	ESS = 27.87	41	ESS = 33.57
Mean (SD) age, years	5.7 (4.3)	5.95 (3.72)	5.70 (3.74)	6.05 (3.76)	5.70 (3.55)
Proportion of males (%)	66.7	50.0	66.7	51.2	66.7
Mean (SD) weekly AM scratching scorec	2.90 (0.56)	2.83 (0.72)	2.90 (0.57)	2.84 (0.73)	2.90 (0.67)
Mean (SD) serum bile acid, µmol/L	266.05 (213.86)	229.28 (107.59)	266.05 (118.92)	230.60 (112.57)	266.05 (129.24)

ESS: effective sample size; ObsRO: observer-reported outcome; SD: standard deviation.aIncludes patients who received odevixibat for ≥48 weeks across Studies A4250-012 and A4250-015.bPropensity score adjusted for age, sex, baseline nighttime scratching score and baseline serum bile acid level.cBased on the ItchRO(Obs) for maralixibat and the ObsRO for odevixibat.

A summary of the unanchored regression-adjusted matched analyses for comparison of changes from baseline to Week 48 in the key efficacy parameters related to pruritus and serum bile acids is provided in Table 20. Although not reaching statistical significance, all adjusted comparisons pointed at consistently stronger effects in improvement in pruritus severity and serum bile acid levels for odevixibat as compared to maralixibat.

Table 20. Regression-adjusted matched comparisons: changes from baseline to week 48 in pruritus severity score and serum bile acid levels (study LUM001-304 and studies A4250-012/A4250-015)

OUTCOME	PARAMETER	POPULATION ADJUSTMENT APPLIED	ODEVIXIBAT VS. MARALIXIBAT [95% CI]
Pruritus Endpoints			
Change from baseline to Week 48 in weekly nighttime scratching score	Mean difference	Regression-adjusted MAIC[a]	-0.53 [-1.09, 0.02]
Proportion of patients achieving a decrease in weekly nighttime scratching score of ≥1 point at Week 48	Odds ratio	Regression-adjusted MAIC[a]	10.13 [0.62, 165.14]
Proportion of patients achieving a weekly nighttime scratching score of ≤ 1 at Week 48	Odds ratio	Regression-adjusted MAIC[a]	2.65 [0.81, 8.62]
Serum Bile Acid Endpoint			
Change from baseline to Week 48 of serum bile acid levels	Mean difference	Regression-adjusted MAIC[b]	-16.78 [-87.00, 53.45]
Proportion of patients achieving a decrease in serum bile acids to ≤ 102 µmol/L at Week 48[c]	Odds ratio	Regression-adjusted MAIC[b]	1.26 [0.19, 8.38]

CI: confidence interval; MAIC: matching adjusted indirect comparison.[a]Propensity scores based on age, sex, and baseline pruritus score and sBA; regression using same factors.[b]Propensity scores based on age, sex, and baseline pruritus score and sBA; regression using age, sex, baseline pruritus score, baseline sBA and baseline total bilirubin. [c]Analysis based on patients with sBA > 102 µmol/L at baseline.

6.1.3. Efficacy Conclusions MAH

Treatment with odevixibat led to a statistically significantly greater improvement in PedsQL total score after 48 weeks of treatment compared to maralixibat. The substantially greater mean difference favouring odevixibat over maralixibat represents a significant benefit for patients with ALGS as the PedsQL total score assesses multiple aspects of a patient's life and overall well-being.

All pruritus and serum bile acid analyses conducted for patients who received at least 48 weeks of treatment with odevixibat demonstrated consistently stronger effects for improvement in these endpoints as compared to maralixibat. The sponsor states that the probability that all tested parameters favour odevixibat over maralixibat by chance is extremely low, and that small sample sizes are often a reality in rare diseases due to the limited patient population.

6.1.4. Efficacy Conclusions COMP

To substantiate their claim of significant benefit, the sponsor is presenting results from unanchored regression-adjusted matched comparisons. It is acknowledged that the unanchored comparisons are motivated by the substantial differences in study designs underlying the odevixibat and the maralixibat applications: odevixibat was investigated in a placebo (PBO) controlled RCT for 24 weeks before transitioning to the open-label phase; maralixibat was investigated in an uncontrolled study with a 4-week PBO controlled randomised withdrawal phase. Nevertheless, the validity of these unanchored comparisons relies on the strong assumption that all relevant prognostic and predictive variables are accounted for in the model. This limitation is also acknowledged by the sponsor. Consequently, results

would need to be particularly compelling (and reliable) to robustly establish that observed differences are truly caused by a difference in QoL/efficacy (or safety) and pharmacodynamics of the underlying treatment and not e.g. by differences in study populations or study designs/treatment regimens.

The comparison relies on pooled data across odevixibat studies A4250-012/-015, i.e. with a fraction of patients starting odevixibat treatment within the pivotal study A4250-012 and a fraction of patients starting odevixibat treatment within the open-label extension study A4250-015. In their position the sponsor did not provide information about the number of patients starting odevixibat treatment per study which is adding to uncertainty about the origin of the data contributing to this comparison.

In addition, the sponsor presents again indirect naïve comparisons on the PedsQL total score over time including the data for maralixibat reported in the Gonzales 2021 paper. *QoL data (PedsQL total score):*

Besides pruritus and sBAs, the sponsor now also includes QoL data (PedsQL total score) to support their claim of significant benefit due to improved efficacy.

Further to the above, the COMP noted that it is very challenging to compare QoL outcomes across studies, as external factors such as study design and other study related factors can largely influence the outcomes of this important, but subjective, outcome parameter. Performing a MAIC including baseline PedsQL total score in the weighting resulted in a very small ESS with many odevixibat patients assigned a small weight for all outcomes, which could also be attributed to differences between the studies.

Importantly, the COMP pointed out that patients treated with Bylvay had similar improvement from baseline in the PedsQL total score as patients treated with placebo in the randomised pivotal study A4250-012. No statistically significant treatment effect was demonstrated, and the study failed to meet this particular endpoint. A positive trend towards an improved PedsQL total score is not reflected in the SmPC for Bylvay either. This questions whether conclusions could be drawn regarding relative efficacy versus other studies and products.

For Livmarli, after an initial improvement in the open-label titration phase, the randomised withdrawal phase had a negative impact both on PedsQL in the active treatment and placebo group. Thus, the design of the study did impact how patients and their caregivers perceive QoL, both for placebo and active treatment. This raises additional concerns over the validity of such unanchored cross study comparisons of this particular endpoint. There is also a risk for bias for the PedsQL comparison at week 48 due to the randomised withdrawal phase (see also discussion below).

Among the 52 patients enrolled in the pivotal odevixibat trial, 42 patients received at least 48 weeks of treatment with odevixibat across the pivotal and open-label extension studies (A4250-012/-015). Among those 42 patients QoL data was available for 30 patients only (Table 17). For maralixibat, QoL data at week 48 was available for 27 out of 31 patients enrolled in the pivotal phase 2 trial (Table 17). The COMP further pointed out that this represents a "loss" of 27% of patients in QoL data for odevixibat while 13%, half of the loss, was presented for maralixibat. The sponsor did not list reasons for why QoL data was missing for either product, which poses a limitation to the interpretation of the results. In particular, the results could be biased in case QoL data are not missing at random. This point could also not be sufficiently clarified during the oral hearing. Reference is made to the EMA Guideline on Missing Data in Confirmatory Clinical Trials (available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf).

Furthermore, it is understood that the 27 patients with QoL outcome data from the pivotal maralixibat study LUM001-304 comprises patients who constantly received maralixibat for 48 weeks as well as patients who received PBO for 4 weeks during the randomised withdrawal period. There is a risk that

the patients who received PBO treatment for 4 weeks will have worse QoL outcomes as compared to patients who were constantly on active treatment, which would unduly favour odevixibat over maralixibat in the indirect comparison. The sponsor does not comment on this risk for bias.

Considering all the above, the COMP did not consider the entirety of QoL data as presented by the sponsor as evidence sufficient for establishing a claim for improved efficacy of odevixibat over maralixibat.

Pruritus and Serum Bile Acids data:

While the COMP noted that in general the missing data in these analyses are less severe, the general issue as regards the high level of uncertainty due to the unanchored cross study comparisons still remains (see discussion above). Furthermore, the committee questioned the clinical relevance of some of the observed differences, e.g. the reported numeric difference of -16 µmol/L in sBA levels, with broad confidence intervals also indicating that the "true" effect may be the opposite.

The observed effects in mean changes from baseline were not of statistical significance and, in addition, there was uncertainty about the quantification of the effect size due to risk for bias and wide confidence intervals (small precision).

Moreover, Alagille syndrome is a heterogenous disease, and it is not certain from the submitted analyses whether all prognostic factors were adjusted for.

Considering the above, the COMP did not consider that the data presented by the sponsor (including, in their totality) could demonstrate an improved efficacy of Bylvay vis-à-vis Livmarli.

6.2. Safety

The following sections provide a comparison of the safety profiles of odevixibat and maralixibat based on data from clinical trials and available post-marketing information.

6.2.1. Clinical trial data

6.2.1.1. Safety analyses

The sponsor presents data from new safety analyses, an indirect weighted adjusted analyses as well as naïve comparisons. The sponsor expects that the new safety analyses address the concerns raised previously by the COMP, i.e. differences in treatment duration (18 weeks with maralixibat and 24 weeks with odevixibat) as well as potential reporting bias in the comparison of open-label data (maralixibat) with placebo-controlled data (odevixibat).

6.2.1.2. Results of the regression-adjusted matched indirect comparisons for safety

Regression-adjusted matched indirect comparisons for safety analyses were conducted based on 2 sets of data during open-label treatment as there are only limited (4-week) placebo-controlled data for maralixibat at the approved therapeutic dose, precluding meaningful comparisons. The comparisons were conducted based on: 1) open-label treatment during the initial 18 weeks of treatment with each compound, and 2) after the 4-week placebo-controlled RWD period (Weeks 23 to 48) in Study LUM001-304 for maralixibat and after the 24-week placebo-controlled period in Study A4250-012 for odevixibat (i.e. during Study A4250-015).

Table 21 provides the ESS and the unadjusted and adjusted baseline characteristics following propensity score weighting (MAIC) for sex, AM scratching score, and serum bile acid levels for the 18-week analysis. Results were consistent for the Weeks 23/24-48 analysis (data not shown).

Table 21. Effect of matching on baseline characteristics for safety analyses

CHARACTERISTIC	MARALIXIBAT LUM001-304	ODEVIXIBAT STUDY A4250-015	
		WEEKS 0-18	
		UNADJUSTED	ADJUSTED[A]
Number of patients	31	16	ESS = 10.05
Mean (SD) age, years	5.4 (4.2)	5.33 (3.90)	3.95 (3.24)
Proportion of males, %	61.3	37.5	61.3
Mean (SD) scratching weekly AM score	2.9 (0.5)	2.37 (1.01)	2.90 (0.96)
Mean (SD) serum bile acid, µmol/L	283 (211)	254.94 (159.55)	283.00 (166.02)

ESS: effective sample size; SD: standard deviation; MAIC: matched-adjusted indirect comparison.[a]Propensity scores based on sex, baseline pruritus score and sBA. Inclusion of age reduced ESS to an uninformative level.

According to the sponsor's analysis of GI TEAEs and SAEs over the initial 18 weeks of open-label treatment with maralixibat and odevixibat, all regression-adjusted matched comparisons for GI disorders, including the overall incidence as well as the incidence of diarrhoea, abdominal pain, and vomiting, were in favour of odevixibat showing lower incidence of these AEs (Table 22). Results were similar for the Week 23/24 to Week 48 analyses (data not shown).

Table 22. Regression-adjusted matched comparisons: changes during the first 18 weeks of treatment with maralixibat and odevixibat in the incidence of gastrointestinal AEs and SAEs (study LUM001-304 and study A4250-015)

PROPORTION OF PATIENTS WITH THE SPEFICIED AE	PARAMETER	POPULATION ADJUSTMENT APPLIED	RISK DIFFERENCE ODEVIXIBAT - MARALIXIBAT [95% CI]
Gastrointestinal disorder TEAEs	Risk Difference	Regression-adjusted MAIC[a]	**-48.67 [-83.32, -14.02]***
Diarrhoea	Risk Difference	Regression-adjusted MAIC[b]	-24.69 [-55.06, 5.67]
Vomiting	Risk Difference	Regression-adjusted MAIC[a]	**-35.48 [-52.33, -18.64]***
Abdominal pain	Risk Difference	Regression-adjusted MAIC[a]	**-38.71 [-55.86, -21.56]***
SAEs	Risk Difference	Regression-adjusted MAIC[a]	-3.51 [-20.53, 13.51]

AE, adverse event; CI, confidence interval; MAIC, matching adjusted indirect comparison; SAE: serious adverse event; sBA, serum bile acid; TEAE: treatment-emergent adverse event. *Statistically significant in favour of odevixibat.[a]Propensity scores based on sex, baseline pruritus score and sBA; regression using age, sex, baseline pruritus score, baseline sBA and baseline total bilirubin.[b]Propensity scores based on sex, baseline pruritus score and sBA; regression using age, baseline sBA and baseline total bilirubin. Non-convergent regression model with all 5 factors included.

6.2.1.3. Results of the naïve unadjusted indirect comparison for safety

Naïve unadjusted indirect comparison of AE incidence over 13 weeks from placebo-controlled data for maralixibat in Studies LUM001-301 and LUM001-302 and over 13-weeks from placebo-controlled data

for odevixibat in Study A4250-012 were also done. The sponsor emphasizes that even with lower (subtherapeutic) doses of maralixibat (≤280 µg/kg/day) than used in Study LUM001-304 (400 µg/kg/day), the incidence of GI side effects, including diarrhoea (43.6% vs 22.9%), abdominal pain (25.6% vs 5.7%), and vomiting (10.3% vs 5.7%), were 2 to 4 times higher during 13 weeks of treatment with maralixibat compared to 13 weeks of treatment with odevixibat at 120 µg/kg/day (Table 23).

It is also noted that bleeding events were reported in these short-term placebo-controlled studies with maralixibat, including epistaxis in 2 patients and haematemesis in 1 patient; these types of events were not reported during the initial 13-weeks of treatment with odevixibat in Study A4250-012. Further, there has been a higher reported rate of serious bleeding events during long-term treatment with maralixibat compared to odevixibat.

The overall results and the incidence of common GI TEAEs for the above analysis were compared statistically; results are provided in Table 23. The overall rate of TEAEs and the rate of abdominal pain were statistically significantly lower for patients who received blinded odevixibat for 13 weeks compared to patients who received blinded maralixibat for 13 weeks, as evidenced by the 95% CIs excluding 0.

Table 23. Statistical comparison of placebo-controlled adverse event incidence for 13-weeks of treatment with maralixibat and with odevixibat (studies LUM001-301 and LUM001-302 and study A4250-012)

MedDRA SOC Preferred Term	**Studies LUM001-301 and LUM001-302 Overall**	**Study A4250-012 Through Week 13**	**Risk Difference (Odevixibat-Maralixibat) [95% CI]**
	Maralixibat N=39	**Odevixibat N=35**	
Overall TEAEs	*35 (89.7)*	*23 (65.7)*	***-24.0 [-43.26, -4.81]****
Gastrointestinal disorders			
Diarrhoea	17 (43.6)	8 (22.9)	-20.7 [-41.52, 1.70]
Abdominal pain	10 (25.6)	2 (5.7)	**-19.9 [-37.05, -2.71]***
Abdominal pain upper	4 (10.3)	1 (2.9)	-7.4 [-21.73, 5.89]
Vomiting	4 (10.3)	2 (5.7)	-4.5 [-19.26, 10.30]
Overall SAEs	*1 (2.6)*	*2 (5.7)*	*3.2 [-8.52, 16.73]*

SAE: serious adverse events; TEAE: treatment-emergent adverse events. Studies LUM001-301 and LUM001-302 evaluated maralixibat doses < 280 µg/kg/day; Study A4250-012 evaluated odevixibat dose of 120 µg/kg/day. *Statistically significant in favour of lower rates for odevixibat. [a]Data are reported for the first 13 weeks of the odevixibat treatment in Study A4250-012. Source for maralixibat: clinicaltrials.gov

Comparison of SAEs, including clinically significant SAEs of haemorrhages and fractures, during long-term treatment with maralixibat and with odevixibat.

The sponsor also summarises the overall incidence of SAEs with maralixibat than with odevixibat during long term treatment and the incidence of haemorrhage events based on the MedDRA Haemorrhages SMQ and reports of fracture during long-term treatment with odevixibat across Studies A4250-012 and A4250-015 and long-term treatment with maralixibat based on pooled analyses across Studies LUM001-303, LUM001-304, and LUM001-305. This comparison showed more reports of SAEs

of haemorrhage-type events and of fractures during treatment with maralixibat. Overall, SAE Incidence for odevixibat (pooled A4250-012/015; N=50) was 12 (23.1%) and for maralixibat (pooled LUM001-303/304/305; N=84) it was 25 (29.8%). How many of these reported SAEs were judged as being treatment-related by the investigators was not clear. These data are consistent with what has also been observed in review of post-marketing data.

Discontinuation of treatment due to AEs over the course of long-term treatment was higher for maralixibat than for odevixibat

Furthermore, the sponsor compared the discontinuation rates between the 2 treatments. An analysis of long-term treatment with odevixibat across the two phase 3 Studies A4250-012 and A4250-015 (n=50 total) and across the 5 phase 2 Studies LUM001-301, -302, -303, -304, and LUM001-305 (n=84 total) was performed. Overall, a discontinuation rate due to TEAEs of 10.7% was reported for maralixibat and of 2% for odevixibat.

6.2.2. Post-marketing surveillance data

The sponsor presents again the side-by-side comparisons of safety data for maralixibat and odevixibat, based on post-marketing surveillance data derived from a search of the FDA Adverse Event Reporting System (FAERS). In this data-base 600 patients suffering from ALGS treated with maralixibat and 423 patients suffering from PFIC treated with odevixibat were included. The data on the GI related adverse events are the same as compared to the data already submitted by the sponsor and evaluated by the COMP previously (see above Table 10). In addition, the sponsor has now also submitted information on adverse events relating to haemorrhage and fracture suggesting a possibly increased rate for Livmarli/maralixibat compared to Bylvay/odevixibat.

The COMP has previously concluded that the post-marketing data cannot be considered sufficiently compelling due to difficulties in the comparability of the safety profile of odevixibat in patients with PFIC and those with ALGS and due to potential differences in doses used of Bylvay (40-120 mg for the PFIC indication versus 120 mg for the ALGS indication), making comparability difficult to ascertain. It was also noted by the COMP that diarrhoea is associated with the condition and is associated with the bile acid secretion. The sponsor now compared the TEAE rate of odevixibat pivotal studies in PFIC and ALGS, both were 24-week, randomised, double-blind, placebo-controlled studies. Study A4250-005 randomised 62 patients with PFIC to 2 dose levels of odevixibat (40 and 120 µg/kg/day) and placebo. As noted above, Study A4250-012 randomised 52 patients with ALGS to odevixibat 120 µg/kg/day and placebo. The overall incidence of TEAEs was 83.3% for the 42 patients with PFIC who received odevixibat in Study A4250-005 and 74.3% for the 35 patients with ALGS who received odevixibat in Study A4250-012. As well, the overall incidence of TEAEs was similar for PFIC patients who received the 40 and 120 µg/kg/day dose of odevixibat (82.6% and 84.2%, respectively). The sponsor therefore concludes that the safety profile of odevixibat in patients with PFIC and those with ALGS is comparable.

6.2.3. Mechanistic data and impact for patients

In addition, the sponsor now presents detailed arguments on the fact that possible differences in incidence of GI AEs may be explained mechanistically. While Bylvay and Livmarli both act by inhibiting the intestinal bile acid transporter, the impact on individual bile acids might be different. Increased levels of circulating unconjugated bile acids (CDCA, LCA, CA, DCA) were observed after 72 weeks treatment with maralixibat which are known to stimulate gut motility and colonic secretion leading to diarrhoea, abdominal pain and vomiting. This increase in unconjugated bile acids was not observed in patients exposed to odevixibat, as measured at Week 24 (where the levels remained below LoQ), Table

24. The sponsor further emphasized the importance to ALGS patients of a new medicinal product with improved tolerability in view of the fact that many of patients already contend with underlying GI abnormalities and are at risk of cascading effects on overall health/morbidity due to additional effects in this regard.

Table 24. Important Differences in Downstream Effects of Bylvay Provide Foundation for Better Safety Profile

	Bylvay				Livmarli		Reference Range (µmol/L)
	PFIC (24 weeks)		ALGS (24 weeks)		PFIC (72 weeks)		
Bile acid	Before (µM)	After (µM)	Before (µM)	After (µM)	Before (µM)	After (µM)	
CDCA	0.04-0.07	0.06-0.14	<LOQ	0.22	<LOQ	3.37	0.06-0.31
LCA	<LOQ	<LOQ	<LOQ	<LOQ	NA	1.45	0.01-0.03
CA	0.05-0.07	0.05-0.09	0.009	0.062	<LOQ	1.33	0.03-0.16
DCA	<LOQ	0.01	<LOQ	<LOQ	<LOQ	4.35	0.02-0.26

CA: cholic acid, CDCA: chenodeoxycholic acid, DCA: deoxycholic acid, LCA: lithocholic acid

6.2.4. Safety Conclusions MAH

Odevixibat's safety and tolerability profile stands out as superior when juxtaposed with maralixibat, based on several key findings:

1. Lower incidence of AEs: Odevixibat demonstrates a statistically significantly lower incidence of GI side effects and a statistically significantly lower incidence of vomiting and abdominal pain compared to maralixibat in the matched analyses. Such differences in incidence of GI AEs may be explained mechanistically by the very high levels of circulating unconjugated bile acids observed after treatment with maralixibat.
2. Fewer discontinuations indicating better tolerability: The reduced rate of treatment discontinuations due to AEs with odevixibat underscores its better overall tolerability.
3. Post-marketing data consistency: The post-marketing data further corroborates the safety profile of odevixibat, emphasizing its lower overall incidence of AEs, especially GI side effects, compared to maralixibat.

6.2.5. Safety Conclusions COMP

With regard to the newly presented unanchored regression-adjusted MAIC analysis the COMP noted that there were several methodological uncertainties which did not allow to establish the existence of a significant difference in the safety profiles between Bylvay and Livmarli. These uncertainties related to (i) the assumptions underlying the unanchored regression-adjusted comparison, (ii) the small sample size and very small effective sample size (which contributed to age not being included in the propensity score matching and non-convergence of the full regression model for diarrhoea), and (iii) uncertainty over the relevance of the matched parameters for safety comparisons (sex, baseline pruritus score and sBA).

An unanchored comparison relies on the strong assumption that all effect modifiers and prognostic factors are accounted for in the model. The sponsor has not convincingly justified and demonstrated

that observed differences can truly be attributed to a difference in treatment and are not caused by bias.

Furthermore, for establishing drug safety, all available clinical data should be taken into account, and therefore the restriction of the sample size for the purposes of matching is not understood.

The COMP noted that Alagille Syndrome itself is associated with diarrhoea and abdominal pain. This is for example shown by the relatively high rate of diarrhoea events (44.4%) and the rate of abdominal pain (16.7%) in the placebo arms of the 13-week placebo-controlled studies with Livmarli at low-doses (LUM001-301 and LUM001-302). Therefore, the placebo response and background incidence of GI disease symptoms should be taken into consideration as well in the indirect comparisons among study populations.

With regard to the naïve unadjusted indirect comparison of AE incidence over 13 weeks from placebo-controlled data for maralixibat in Studies LUM001-301 and LUM001-302 and for odevixibat in Study A4250-012, the higher rates of diarrhoea and abdominal pain for maralixibat may possibly be explained by differences in the study populations of these trials. Comparing the event rates in the respective placebo arms, diarrhoea events and abdominal pain in the pooled LUM001-301/302 placebo arms were reported to be 44.4% and 16.7% respectively while this was much lower for the placebo arm of study A4250-012 with only 5.9% for each, the diarrhoea events and abdominal pain rate (Livmarli EPAR and Bylvay CHMP AR). In the Livmarli EPAR it is stated that the incidence of events of diarrhoea in the overall maralixibat and placebo groups of studies LUM001-301 and LUM001-302 were similar (43.6% vs. 44.4% of participants, respectively), while events of abdominal pain were slightly reported more rarely in the placebo groups versus maralixibat (16.7% vs 25.6%, respectively). However, the majority of these adverse events were mild to moderate in severity, transient in nature, and resolved with no action taken.

The pharmacological rationale presented to explain the possible difference in gastro-intestinal side effects was based on data sets which were difficult to compare. The analyses of the secondary unconjugated bile acids for Livmarli were performed after 72 weeks of treatment, whereas this was after 24 weeks for Bylvay. Also, for Bylvay some changes from baseline were observed after shorter term drug exposure, albeit minimal. A direct comparison is considered hampered by the differences in drug exposure time. Also, Alagille syndrome is a heterogenous disease which may in itself be associated with gastro-intestinal symptoms.

With regard to the selected serious adverse events (SAE) differences as presented by the sponsor, the COMP noted that fractures and severe haemorrhages due to vascular abnormalities are well-known and common disease complications (*Circulation. 2004;109:1354-1358, Kamath BM et al.,. J Pediatr Gastroenterol Nutr. 2018;67(2):148e156*). As it can also be readily inferred from the EPAR of Livmarli, such SAEs were not considered to be treatment related.

With regard to the reported differences in discontinuations the COMP noted that (also by reference to the EPAR) the reasons for discontinuation for maralixibat were possibly drug related for 6 cases in the pooled safety dataset, mainly due to increments of liver function tests. For both products, however, it was considered (also by the CHMP) that there is a potential risk of hepatotoxicity, to be further investigated post-marketing. There were no discontinuations due to GI AE for Livmarli.

With regard to the post-marketing surveillance data in combination with the additional analyses and conclusions that the overall incidence of TEAEs with odevixibat in patients with PFIC and with ALGS is similar, the COMP pointed out that besides differences in disorders, the dose recommendation of Bylvay is in a lower range for PFIC than for ALGS. This hampers a direct comparison of the databases.

Further to the above, the COMP considered that the submitted analyses could not establish a claim of better safety or tolerability of Bylvay over Livmarli.

6.3. Major contribution to patient care (improvement in sleep disturbance and QoL)

In order to substantiate their claim for a major contribution to patient care of Bylvay vis a vis Livmarli, the sponsor provides further analyses to demonstrate that improvement in pruritus was not the sole driver for improvement in sleep during treatment with odevixibat. In this regard the results from two new analyses were presented: 1) a correlation analysis to evaluate the relationship between improvements in pruritus and improvements in sleep and 2) a statistical model (MMRM) that factors out the impact of the change in pruritus in the assessment of changes in sleep.

1) Correlation analyses based on Pearson correlation coefficients were conducted between changes from baseline in the sleep disturbance assessments (tiredness, help falling asleep, awakenings, and days soothing, sleeping with caregiver, and taking medications) and scratching score to Weeks 21-24 in Study A4250-012 and to Weeks 45-48 across the pooled analyses for Studies A4250-012 and A4250-015. Only a modest correlation between sleep and pruritus was seen at both the Weeks 21-24 and 45-48 analyses (correlation coefficient: 0.2838 and 0.2495, respectively).

2) As previously described, in the primary MMRM analysis, odevixibat led to statistically significant improvements in 4 of the key sleep parameters, including percent of days with help falling asleep, with soothing, and with sleeping with the caregiver, as well as daytime tiredness (see above Table 12). After factoring out improvement in pruritus, 3 of the 4 key sleep parameters still showed that odevixibat significantly improved sleep compared to placebo – indicating that although pruritus does play a role in sleep improvement (as evidenced by the smaller LS mean differences between odevixibat and placebo), it is not the sole driver for the results observed (see below Table 25).

Table 25. Effect of treatment with odevixibat on sleep relative to placebo with and without controlling for the relationship between pruritus and sleep in the MMRM

SLEEP PARAMETER	**PRIMARY ANALYSIS[A] WEEKS 21-24**		**ANALYSIS INCLUDING CHANGE IN PRURITUS AS COVARIATE[B] WEEKS 21-24**	
	LS MEAN DIFFERENCE (ODEVIXIBAT-PLACEBO) (95% CI)	**P-VALUE**	**LS MEAN DIFFERENCE (ODEVIXIBAT-PLACEBO) (95% CI)**	**P-VALUE**
Percent of days with help falling asleep	-33.4 (-54.86, -11.85)	**0.0016***	-27.8 (-48.07, -7.56)	**0.0042***
Percent of days with soothing	-40.4 (-58.77, -21.96)	**<0.0001***	-34.4 (-51.18, -17.58)	**0.0001***
Percent of days sleeping with caregiver	-26.3 (-43.27, -9.23)	**0.0017***	-24.5 (-41.45, -7.58)	**0.0028***
Tiredness	-0.60 (-1.065, -0.136)	**0.0062***	-0.09 (-0.477, 0.289)	0.3121

CI: confidence interval; LS: least square; MMRM: mixed model for repeated measures. *Statistically significant result. a The analysis is based on MMRM using a restricted maximum likelihood with baseline sleep score as a covariate, and baseline age stratification, treatment group, time (in months), and treatment-by-time interaction as

fixed effects. b The analysis is based on MMRM using a restricted maximum likelihood with baseline sleep score and change from baseline in scratching score as covariates, and baseline age stratification, treatment group, time (in months), and treatment-by-time interaction as fixed effects.

The sponsor also emphasized again that improvement in sleep disturbance was observed in pruritus responders and in non-responders of odevixibat. In contrast, data with maralixibat did suggest a possible correlation between pruritus and sleep, i.e. data from the PedsQL multidimensional fatigue scale scores and for the trouble sleeping subscore from the emotional domain of the PedsQL reported in the Kamath et al., 2023 publication showed that patients who did not achieve a reduction in scratching score ≥1 had worsening of sleep parameters based on the multidimensional fatigue scale [Table VII, Kamath 2023].

Furthermore, the sponsor refers again the data from the regression-adjusted matched comparison of changes from baseline to Week 48 in the QoL parameters showing a statistically significant greater improvement in the PedsQL total score achieved with Bylvay, as compared to Livmarli. This data has been discussed in detail above and was used by the sponsor to substantiate their claim for improved efficacy.

6.3.1. MCPC Conclusions by MAH

The correlation analyses between sleep improvement and pruritus revealed weak to moderate correlations, suggesting that odevixibat's impact on sleep isn't entirely pruritus dependent. Furthermore, statistical analyses that factored out pruritus improvement still showed odevixibat's significant benefits on sleep. This was further corroborated by the observation that both pruritus responders and non-responders exhibited improvements in sleep parameters when treated with odevixibat.

6.3.2. MCPC Conclusions COMP

The sponsor claims that Bylvay improves sleep as compared to Livmarli. The sponsor also claims that Bylvay may have a beneficial effect on sleep, independent of its effect on pruritus, which is known to disturb sleep. However, apart from general PedQL sub-scores relating to sleep/fatigue, the pivotal clinical trial for Bylvay also employed dedicated sleep assessment instruments which were not used in the pivotal trial of Livmarli (i.e. the Albireo ObsRO/PRO). This hampered the comparison of the effects on sleep of both treatments.

The sponsor argues that the improvement in sleep shown for Bylvay was not caused solely by an improvement in pruritus, and hence one could not automatically conclude that an improvement in sleep would also be seen for Livmarli. The sponsor supports this claim with various analyses. Nevertheless, the COMP considers that even if an effect on pruritus was not the sole driver for improvement in sleep, the presented analyses do not entail that Livmarli does not have a positive effect on sleep. The COMP further noted that Livmarli also had a relevant impact on improvement of fatigue scores. Although not the same as sleep parameters, these are clinically related symptoms.

Furthermore, it could not be established how Bylvay could influence sleep pharmacologically, beyond its effects on pruritus, and whether this would be different from Livmarli, which belongs to the same pharmacological class.

6.4. Overall conclusion COMP

The data presented by the sponsor, including the additional data presented at the re-examination stage, were not deemed sufficient evidence to establish a clinically relevant advantage or a major

contribution to patient care of Bylvay over Livmarli. The criteria for maintaining the orphan designation as set out in Article 3(1)(b) are therefore not considered satisfied.

7. COMP final position on review of criteria for orphan designation adopted on 5 October 2023

The COMP concluded that:

- the proposed therapeutic indication falls entirely within the scope of the orphan condition of the designated Orphan Medicinal Product;
- the prevalence of Alagille syndrome (hereinafter referred to as "the condition") was estimated to remain below 5 in 10,000 and was concluded to be 0.3 in 10,000 persons in the European Union, at the time of the review of the designation criteria;
- the condition is life-threatening and chronically debilitating due to severe pruritus, cholestasis, liver failure, skeletal malformations and fractures, congenital cardiac and vascular defects. Life expectancy is in most cases around 20 years and death is associated with haemorrhages, cardiac failure and end-stage liver disease;
- in view of the fact that a satisfactory method for the treatment of the condition has been authorised in the European Union (Livmarli), the existence of significant benefit over the authorised method of treatment should be established at the stage of the granting of marketing authorisation;
- the sponsor's claim that Bylvay is of significant benefit to those affected by the orphan condition was not established. Significant benefit over Livmarli was claimed on the grounds of a clinically relevant advantage as well as a major contribution to patient care;
- the sponsor did not provide sufficient evidence to establish a clinically relevant advantage over Livmarli:
  - based on the analyses presented by the sponsor, including the additional analyses presented at the re-examination stage, on mean change from baseline to week 48 of pruritus and serum bile acid (sBA) levels a significant difference in the efficacy profiles between Bylvay and Livmarli could not be established. The clinical relevance of some of the estimated effect differences is questioned, combined with uncertainty about the quantification of the effect size due to wide confidence intervals, risk for bias due to the differences between the studies including that the Livmarli arm included patients who received placebo during the randomized withdrawal design (RWD) phase, missing data, as well as insufficient discussion about the assumptions and limitations underlying the unanchored regression-adjusted matched comparison;
  - with regard to the quality of life (QoL) data, an improved effect of Bylvay as compared to Livmarli could not be established. This is due to methodological uncertainties associated with assumptions underlying the unanchored regression-adjusted matched comparison, differences between study designs which could influence QoL scoring, including the fact that the Livmarli arm included patients who received placebo during the RWD phase, missing data in conjunction with the subjectivity of the parameters and the fact that Bylvay could not demonstrate a significant effect over placebo on the exploratory QoL endpoint in its pivotal clinical trial owing to similar improvements in QoL in the placebo treated patients;
  - with regard to safety, a significant difference in the safety profiles between Bylvay and Livmarli could not be established due to methodological uncertainties related to (i) the small sample size and very small effective sample size, (ii) the assumptions underlying the unanchored regression-adjusted comparison and (iii) uncertainty over the relevance of the matched parameters in terms of being prognostic for adverse events. With regard to the naïve

unadjusted indirect comparison and the real-world data, it was considered that, due to their limitations, these data could not suffice to establish the claimed differences in safety. In addition, the pharmacological rationale presented to explain the possible difference in gastro-intestinal side effects was based on data sets which were difficult to compare. Also, Alagille syndrome is a heterogenous disease which may in itself be associated with gastro-intestinal symptoms;

- the sponsor did not provide sufficient data to demonstrate a major contribution to patient care of Bylvay in comparison to Livmarli. In particular, the sponsor claims that Bylvay improves sleep as compared to Livmarli. The sponsor also claims that Bylvay may have a beneficial effect on sleep, independent of its effect on pruritus, which is known to disturb sleep. While both pivotal trials for Bylvay and Livmarli evaluated sleep, different scales were used. This hampered the interpretation of the presented data on sleep. Furthermore, it could not be established how Bylvay could influence sleep pharmacologically, beyond its effects on pruritus, and whether this would be different from Livmarli, which belongs to the same pharmacological class.

The COMP, having considered the information submitted by the sponsor and on the basis of Article 5(12)(b) of Regulation (EC) No 141/2000, is of the opinion that:

- the criteria for designation as set out in the first paragraph of Article 3(1)(a) are satisfied;
- the criteria for designation as set out in Article 3(1)(b) are not satisfied.

The Committee for Orphan Medicinal Products has recommended that Bylvay, (2S)-2-{[(2R)-2-[({[3,3-dibutyl-7-(methylthio)-1,1-dioxido-5-phenyl-2,3,4,5-tetrahydro- 1,2,5-benzothiadiazepin-8-yl]oxy}acetyl)amino]-2-(4-hydroxyphenyl)acetyl]amino}butanoic acid, odevixibat for treatment of Alagille syndrome (EU/3/12/1040) is removed from the Community Register of Orphan Medicinal Products.