Validation of the PRUCISION Instruments in Pediatric Patients with Progressive Familial Intrahepatic Cholestasis

Chad Gwaltney (iD) · Cristina Ivanescu · Lisa Karlsson ·
Natalie Warholic · Lise Kjems · Patrick Horn

ABSTRACT

Introduction: Patients with cholestatic liver disease, including progressive familial intrahepatic cholestasis (PFIC) and Alagille syndrome, may have debilitating pruritus, and reducing pruritus is a key therapeutic goal. However, few instruments are available that adequately measure pruritus in pediatric patients with cholestatic liver disease. The objectives of the current study were to establish the measurement properties of the novel PRUCISION patient-reported outcome (PRO) and observer-reported outcome (ObsRO) instruments and to estimate a threshold for clinically meaningful change in pruritus score.
Methods: The PRO/ObsRO instruments are completed twice daily via electronic diary and include 5-point pictorial responses to assess pruritus. Sleep disturbance and tiredness were quantified using 5-point pictorial responses,

Natalie Warholic, Lise Kjems, and Patrick Horn:
Employees of Albireo Pharma at the time of the study.

C. Gwaltney (✉)
Gwaltney Consulting, 1 Bucks Trail, Westerly,
RI 02891, USA
e-mail: cgwaltney@gwaltneyconsulting.com

C. Ivanescu
IQVIA, New York, NY, USA

L. Karlsson · N. Warholic · L. Kjems · P. Horn
Albireo Pharma, Inc., Boston, MA, USA

yes/no responses, and numerical ratings. Data from PEDFIC 1 (NCT03566238), a phase 3 study evaluating odevixibat efficacy and safety in children with PFIC, were used to assess the psychometric properties of these instruments. Quantitative assessments included evaluation of test–retest reliability, determination of construct validity via convergent and known-group validity analyses, and characterization of sensitivity to change. A threshold for within-patient meaningful change from baseline to week 24 was determined using blinded data from PEDFIC 1 and distribution- and anchor-based analyses.
Results: Because the majority of patients in PEDFIC 1 were aged < 8 years ($n = 52/62$) and thus too young to complete the PRO instrument, which was intended for patients aged ≥ 8 years, the small sample size of patients who completed the PRO precluded a full psychometric analysis of the PRO instrument. The ObsRO was completed by a caregiver of every patient in PEDFIC 1. The ObsRO instrument had acceptable test–retest reliability based on intraclass correlation values (most > 0.75). Convergent validity analyses revealed moderate-to-strong correlations ($r \geq 0.3$) between baseline ObsRO pruritus scores and baseline Global Impression of Symptoms (GIS) items. In known-groups validity analyses, there were significant differences between baseline groups defined by the GIS for ObsRO pruritus scores and for some sleep disturbance scores. Week 24

ObsRO scores were in the expected direction in groups defined by the Global Impression of Change scale (i.e., improved or not improved); many mean differences between these groups were significant. Sensitivity to change for the ObsRO PRUCISION instrument was also demonstrated by moderate-to-strong Pearson correlations between change from baseline to weeks 21–24 in ObsRO scores and GIS items ($r \geq 0.3$). Based on these analyses, a within-patient change of -1.00 from baseline in ObsRO pruritus score was determined to be clinically meaningful.

Conclusion: The PRUCISION ObsRO instrument is reliable, valid, and sensitive to change, supporting its use as a tool to measure pruritus and sleep disturbance in patients with PFIC and other pediatric cholestatic liver diseases.

PLAIN LANGUAGE SUMMARY

Progressive familial intrahepatic cholestasis (PFIC) is a collection of liver diseases that typically affects very young children. A problematic symptom of PFIC is extremely itchy skin, or pruritus, that can keep patients and their families up at night. The PRUCISION questionnaire was developed to measure the severity of a patient's pruritus and sleep disturbance from the perspective of the patient's caregiver. The current study had two primary goals: (1) to assess whether PRUCISION could reliably measure these symptoms and detect changes over time relative to other established rating scales that assess related concepts, and (2) to identify what score change on PRUCISION could be considered clinically meaningful. To do this, data from a clinical study, called PEDFIC 1, in patients with PFIC were used: patient's scores on PRUCISION from their caregiver's perspective were compared with scores on other established scales, first before any treatment was given in PEDFIC 1, and then again after 24 weeks of treatment with a drug called odevixibat. In general, there was good agreement between PRUCISION scores and scores on other scales. For example, when PRUCISION scores indicated that symptoms improved, this tended

to correlate with improvement on other measures. Additionally, these analyses indicated that if the PRUCISION score drops by 1 point or more, that can be considered a clinically important change. Overall, this study found that the caregiver-reported PRUCISION questionnaire is valid for assessing changes in pruritus and sleep symptoms in patients with PFIC, which may benefit patients as new treatments are developed.

Key Summary Points

Prior to this study, there was a lack of publicly available instruments that adequately measured pruritus in pediatric patients with cholestatic liver diseases from the perspective of patients or their caregivers. Here, we assessed the measurement properties of newly developed tools, a patient-reported outcome (PRO) and an observer-reported outcome (ObsRO), called PRUCISION, for examining pruritus and sleep disturbance in such patients using data from a phase 3 study of odevixibat in patients with progressive familial intrahepatic cholestasis (PFIC).

Examination of the baseline distributions of the PRO and ObsRO data indicated that patients in PEDFIC 1 experienced significant pruritus and related sleep disturbance.

Psychometric analyses found that the ObsRO instrument reliably measures pruritus and sleep disturbance in children and adolescents with PFIC and represents a promising tool for assessing these symptoms over time from the perspective of their caregivers.

INTRODUCTION

Progressive familial intrahepatic cholestasis (PFIC) is a group of rare, autosomal-recessive liver diseases estimated to occur in approximately 1 in 75,000 children worldwide [1]. PFIC is characterized by impaired bile acid secretion and transport and the accumulation of bile components, such as bilirubin and bile acids, in the liver [2, 3]. As hepatic levels of these components increase, they can be excreted into the systemic circulation, where they may be associated with the development of jaundice and pruritus, respectively [3–5].

Pruritus is a problematic symptom for cholestatic patients that can severely reduce quality of life, limit activities of daily living, and cause significant sleep deprivation [6–8]. The initial development of the PRUCISION patient-reported outcome (PRO) and observer-reported outcome (ObsRO) pruritus instruments included a review of the literature, discussion with expert clinicians, and interviews with pediatric patients with cholestatic liver diseases and their caregivers. These are reported in a companion article in this issue. In the initial study, pruritus and associated sleep disturbance were identified as central to the experience of pediatric cholestatic liver disease. Although reducing pruritus is a key treatment objective in these diseases, there was a lack of publicly available instruments that adequately measured pruritus in pediatric patients from a patient or caregiver perspective prior to this study [5]. The PRUCISION PRO and ObsRO pruritus instruments were developed to address this need.

The PRUCISION instruments were developed according to regulatory guidelines for establishing measurement properties (e.g., reliability, validity, and ability to detect change) of clinical outcomes assessments [9, 10]. Based on these guidelines, reliability measures may include test–retest and inter- or intra-rater assessments [9]. Approaches for determining instrument construct validity, or experimentally demonstrating that an instrument measures core constructs, may include assessing convergent validity, or how well constructs that theoretically should be related to each other are

observed to be related, and known-groups validity, or the degree to which an instrument could distinguish between clinically distinct groups [10]. The guidelines define the ability to detect change as the ability of an instrument to identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept [9].

The objectives of this study were to assess such measurement properties (reliability, validity, and sensitivity to change) of the newly developed PRUCISION PRO and ObsRO instruments and to estimate a threshold for clinically meaningful change in pruritus score.

METHODS

PRO and ObsRO Instruments

The PRO/ObsRO instruments focus on key symptoms of pediatric cholestatic liver disease: pruritus, sleep disturbance, and associated tiredness (Fig. 1a and b). Patients and/or caregivers captured details of these symptoms twice daily (each morning and evening) using an electronic diary (eDiary). The morning diary entry was used to record nighttime itching and scratching severity, aspects of sleep disturbance, and tiredness upon waking. The evening/bedtime diary entry was completed just before the patient went to bed and recorded the patient's itching and scratching severity and tiredness during the day.

The PRO consists of seven questions and uses two different rating scales (Fig. 1a). The PRO morning diary questions 1, 2, 3, and 5 and the evening diary questions 1 and 2 are scored on a 5-point pictorial response scale, where higher scores indicate worse symptoms. The PRO morning diary question 4 has a "yes" or "no" (binary) response format.

The ObsRO instrument consists of eight questions and uses three response formats (Fig. 1b). The ObsRO morning diary question 1 and evening diary questions 1 and 2 use 5-point response scales similar to those on the PRO. ObsRO morning diary questions 2, 3, 4, and 5 have "yes" or "no" responses, and morning question 6 allows for a numeric response.

a

PRO Morning Diary (to be completed shortly after waking each morning)

Please answer the questions on the following screens. There are no right or wrong answers. Please think about the time <u>since you went to bed last night</u> (beginning when you started trying to fall asleep).

1. How bad was your worst itching since you went to bed last night?

0 NO ITCHING | 1 A LITTLE ITCHING | 2 MEDIUM ITCHING | 3 A LOT OF ITCHING | 4 THE WORST ITCHING

2. How hard was it to <u>fall</u> asleep last night because of your itching?

0 NOT HARD AT ALL | 1 A LITTLE HARD | 2 MEDIUM HARD | 3 VERY HARD | 4 VERY, VERY HARD

3. How hard was it to <u>stay</u> asleep last night because of your itching?

0 NOT HARD AT ALL | 1 A LITTLE HARD | 2 MEDIUM HARD | 3 VERY HARD | 4 VERY, VERY HARD

4. Did you wake up last night because of itching?

No   Yes

5. How tired do you feel this morning?

0 NOT TIRED AT ALL | 1 A LITTLE TIRED | 2 MEDIUM TIRED | 3 VERY TIRED | 4 VERY, VERY TIRED

PRO Bedtime Diary (to be completed when going to bed each night)

Please answer the questions on the following screens. There are no right or wrong answers. Please think about the time <u>since you woke up this morning</u>.

1. How bad was your worst itching since you woke up this morning?

0 NO ITCHING | 1 A LITTLE ITCHING | 2 MEDIUM ITCHING | 3 A LOT OF ITCHING | 4 THE WORST ITCHING

2. How tired were you since you woke up this morning?

0 NOT TIRED AT ALL | 1 A LITTLE TIRED | 2 MEDIUM TIRED | 3 VERY TIRED | 4 VERY, VERY TIRED

Fig. 1 PRUCISION PRO (**a**) and ObsRO (**b**) instruments used in the PEDFIC 1 study. *ObsRO* observer-reported outcome, *PRO* patient-reported outcome.

b

ObsRO Morning Diary (to be completed shortly after child wakes up each morning)

Please answer the questions on the following screens. There are no right or wrong answers. Please think about the time <u>since your child went to bed last night</u> (beginning when your child started trying to fall asleep). As a reminder, when we use the term "scratching," we would like you to think about all of the different types of scratching behaviors that you can see.

1. How bad was your child's worst scratching since he/she went to bed last night?	**0** NO SCRATCHING **1** A LITTLE SCRATCHING **2** MEDIUM SCRATCHING **3** A LOT OF SCRATCHING **4** WORST POSSIBLE SCRATCHING
2. Since your child went to bed last night, did you see blood due to scratching?	**No** **Yes**
3. Did your child need a caregiver to help him/her fall asleep last night due to his/her itching?	**No** **Yes**
4. Did your child need a caregiver to soothe him/her at some point during the night last night due to his/her itching?	**No** **Yes**
5. Did your child need a caregiver to sleep with him/her at some point during the night last night due to his/her itching?	**No** **Yes**
6. How many times did you notice that your child woke up last night?	*Enter a number from 0 to 99*

ObsRO Bedtime Diary (to be completed when child is going to bed each night)

Please answer the questions on the following screens. There are no right or wrong answers. Please think about the time <u>since your child woke up this morning</u>. As a reminder, when we use the term "scratching," we would like you to think about all of the different types of scratching behaviors that you can see.

1. How bad was your child's worst scratching since he/she woke up this morning?	**0** NO SCRATCHING **1** A LITTLE SCRATCHING **2** MEDIUM SCRATCHING **3** A LOT OF SCRATCHING **4** WORST POSSIBLE SCRATCHING
2. How tired did your child seem to be today?	**0** NOT TIRED AT ALL **1** A LITTLE TIRED **2** MEDIUM TIRED **3** VERY TIRED **4** VERY, VERY TIRED

Fig. 1 continued

PEDFIC 1 Clinical Trial

Overview

The psychometric measurement properties (i.e., reliability, construct validity, sensitivity to change) of the PRO/ObsRO instruments were examined through analysis of data from the randomized, double-blind, placebo-controlled PEDFIC 1 study [11] (for a study flowchart, see Fig. 2). This phase 3 study enrolled children aged 6 months to 18 years with PFIC1 or PFIC2. The primary aim of the study was to assess the efficacy of the ileal bile acid transporter inhibitor odevixibat, as determined by improvements in pruritus and reductions in serum bile acids. The study also assessed safety and tolerability.

The PEDFIC 1 study consisted of a screening period with a duration of 5–8 weeks, a 24-week treatment period, and a 4-week follow-up period. The PEDFIC 1 study materials received central or local ethics committee (CEC and LEC, respectively) approval from all 45 study sites in the United States, Canada, Europe, Australia, and the Middle East that intended to enroll patients (i.e., from 10 CECs representing 21 sites and from 22 LECs representing 24 sites); the full list of ethics committees providing approval is available in the appendix of the PEDFIC 1 article [11]. In addition, the study was conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonization Good Clinical Practice guidelines; all patients or their caregivers provided written consent prior to study participation.

Assessments

Patients and/or their caregivers were provided with an eDiary at the first screening visit for recording itching (PRO), observed scratching (ObsRO), and sleep disturbance (PRO and ObsRO; Fig. 1a and b) throughout the study. The PRO instrument was used in patients aged ≥ 8 years, and the ObsRO instrument was completed by patients' caregivers, regardless of patient age.

Patients aged ≥ 8 years, caregivers, and clinicians completed the Global Impression of Change (GIC; called PGIC, CaGIC, CGIC, for patient-, caregiver-, and clinician-reported versions, respectively) and Global Impression of Symptoms (GIS; called PGIS, CaGIS, CGIS, for patient-, caregiver-, and clinician-reported versions, respectively) at randomization (GIS only) and at weeks 4, 12, and 24 of PEDFIC 1. The GIC assesses change in itching or scratching and sleep since starting the study drug using a 7-point scale that ranges from 1 ("very much better") to 7 ("very much worse"). The GIS assesses itching (PGIS) or scratching (CaGIS and CGIS) and sleep (all versions) in the past week using a 5-point scale that ranges from 1 ("none"/no problems) to 5 ("very bad/very severe" problems).

During PEDFIC 1, caregivers and patients completed the Pediatric Quality of Life Inventory (PedsQL; version 4.0), an instrument designed to assess quality of life in children and adolescents [12]. The PedsQL examines functioning of the patient via four domains: physical, emotional, social, and school. The family impact module component of the PedsQL measures the impact of pediatric chronic health conditions on parents and the family; this assessment has eight domains: physical, emotional, social, and cognitive functioning, as well as communication, worry, daily activities, and family relationships. The PedsQL and family impact module were administered at randomization and week 24 in PEDFIC 1.

Analyses

Completion rates for PRO and ObsRO instruments were calculated as follows: the total number of actual completed eDiary entries was divided by the number of eDiary entries expected to be completed for a given time period (i.e., 14 days for baseline and 28 days for all 4-week interval time points).

A blinded analysis of ObsRO pruritus data from PEDFIC 1 was conducted to estimate a meaningful within-patient change threshold. The analysis included all available eDiary data captured as of June 29, 2020, which represented data from 48 patients. The identified threshold for meaningful change was then incorporated into the main PEDFIC 1 analysis plan prior to database lock and unblinding. All other analyses were based on data from the final locked
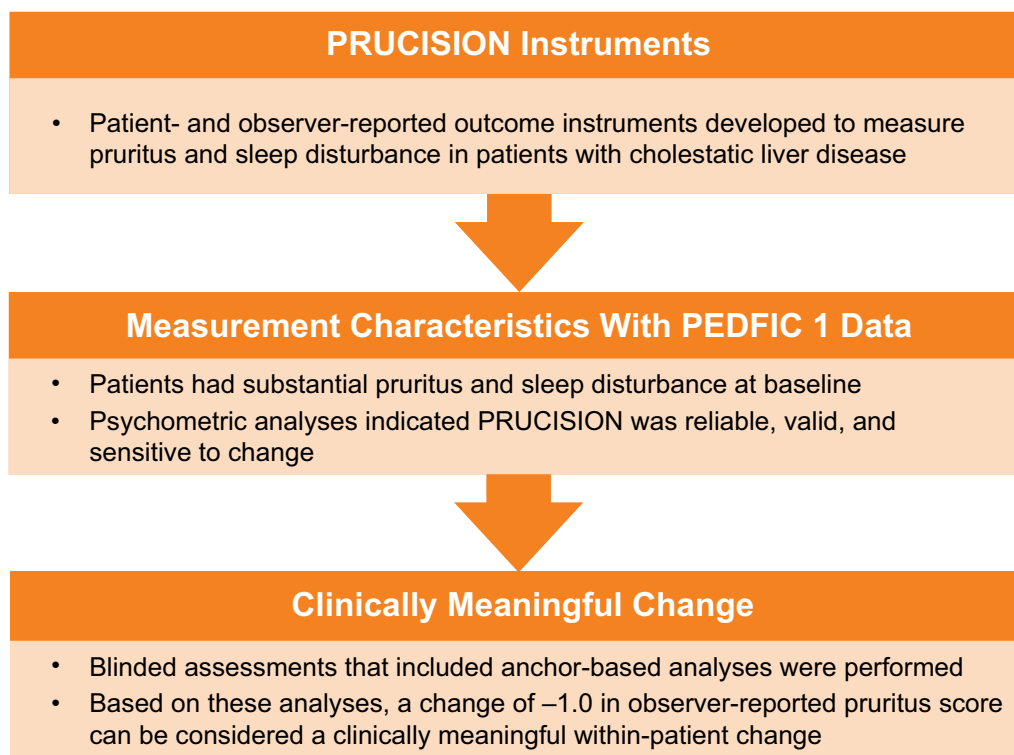
PRUCISION Instruments

- Patient- and observer-reported outcome instruments developed to measure pruritus and sleep disturbance in patients with cholestatic liver disease

Measurement Characteristics With PEDFIC 1 Data

- Patients had substantial pruritus and sleep disturbance at baseline
- Psychometric analyses indicated PRUCISION was reliable, valid, and sensitive to change

Clinically Meaningful Change

- Blinded assessments that included anchor-based analyses were performed
- Based on these analyses, a change of −1.0 in observer-reported pruritus score can be considered a clinically meaningful within-patient change

Fig. 2 Study flow chart

database, with the analysis population comprising all randomized patients who received ≥ 1 dose of study treatment, and psychometric analyses performed on pooled treatment arms.

Psychometric Properties of PRUCISION Instruments

Descriptive Statistics at Baseline
Descriptive statistics for PRO/ObsRO instrument items were summarized at baseline, including mean (standard deviation [SD]) values and the distribution of scores to assess potential floor and ceiling effects.

Reliability
Reliability was assessed by calculating inter-item correlations between morning and bedtime PRO/ObsRO pruritus scores. In addition, test–retest reliability was assessed by calculating intraclass correlations (ICCs), where ICC values of 0.50–0.75 indicate moderate reliability, values of 0.75–0.90 indicate good reliability, and

values > 0.90 indicate excellent reliability [13]. ICCs were calculated using two approaches, as follows: (1) by comparing weekly PRO/ObsRO scores from the first week (days −14 to −8) of the baseline interval ("test") with those from the second week (days −7 to −1) of the baseline interval ("retest"); and (2) by comparing scores in stable patients (i.e., those with no change on the GIC at week 4 or those with the same responses at baseline and week 4 on the GIS) during the baseline interval (average from day −14 to day −1) with those at week 4 (using the monthly score).

Construct Validity
Construct validity was tested in two ways, through convergent validity and known-groups validity assessments. Convergent validity was assessed using Spearman correlations to compare baseline PRO/ObsRO scores and baseline PGIS, CaGIS, and CGIS item scores and PedsQL and PedsQL family impact module scores. Spearman correlations between 0.30 and 0.49 are considered moderate, and those ≥ 0.50 are

considered strong [14]. Known-groups validity was evaluated by partitioning baseline responses from the GIS into two groups based on symptom severity as follows: responses of "none," "mild," and "moderate" were grouped and compared with responses of "severe"/"very severe." A Student t test was used to test differences in PRO/ObsRO scores between these response groups.

Sensitivity to Change

Two approaches were also used to evaluate the instruments' sensitivity to change. First, sensitivity to change was assessed by comparing mean changes from baseline to weeks 21–24 in PRO/ObsRO scores between "improved" (participants who answered "a little better," "much better," or "very much better" at week 24 on GIC scales) and "not improved" (participants who answered "no change," "a little worse," "much worse," or "very much worse" at week 24 on GIC scales); mean differences in PRO/ObsRO scores between these groups were evaluated using analysis of covariance models. Second, Pearson correlations were calculated between the change from baseline to weeks 21−24 in the PRO/ObsRO scores and the GIC and GIS at week 24.

Threshold for Clinically Meaningful Change

Distribution- and anchor-based analyses were used to estimate a threshold for within-patient meaningful change from baseline to weeks 12 and 24 for the ObsRO pruritus instrument. Using distribution-based analyses, the 0.5 SD and 1 standard error of the mean (SEM) values for baseline pruritus scores were calculated. Anchor-based analyses involved examining the degree of change in ObsRO pruritus score from baseline to week 12 or 24 among patients who experienced pruritus improvement in PEDFIC 1 based on GIC and GIS anchors. Each of the anchors included several response options. The responses were dichotomized in different ways to facilitate selection of the threshold that best differentiated responders and nonresponders. There were three dichotomized categories for both the GIC and GIS scales (Table 1). For both scales, definition 1 reflected the maximum amount of change, definition 3 reflected the smallest amount of change,

and definition 2 fell between these two extremes. Correlations between each dichotomized level of the anchor and the ObsRO pruritus change score were calculated, with increased weight given to anchors with a correlation of 0.30 or greater. The anchor most highly correlated with the pruritus measure was used as the primary anchor. The 95% confidence interval (CI) for the mean change from baseline in pruritus value was also calculated for patients who were stable according to the GIC or GIS anchor. The lower bound of the 95% CI for this stable group was used as a comparison value in examining the meaningful change estimates from the other anchor categories.

The anchor-based analyses were used as the primary determinant for estimating a meaningful change threshold, with receiver operating characteristic (ROC) analyses and empirical cumulative distribution function (eCDF) plots used as complementary approaches. The threshold was evaluated as follows: (1) a primary anchor was selected based on correlations between the anchor and the pruritus measure; (2) the smallest median value for the primary anchor that exceeded the values from the distribution-based analyses AND the lower bound of the 95% CI from the stable anchor category was selected as a candidate threshold value; (3) values were tabulated to evaluate the consistency between the primary and complementary threshold estimates; and (4) the final meaningful change estimate was rounded to the nearest 0.5 value on the 0–4 scale to increase the interpretability of the threshold value. Because the results for the 12- and 24-week intervals were similar, the remainder of this manuscript focuses on week 24 data (week 12 data are not shown).

RESULTS

PEDFIC 1 Clinical Trial

Patient Demographics and Baseline Characteristics

Demographics and baseline characteristics of patients randomized in PEDFIC 1 are summarized in Table 2. Overall, 84% of patients ($n = 52/62$) were aged < 8 years, with 37% and

Table 1 GIC and GIS anchors and dichotomized response categories

Definition name	Responses indicating improvement (anchor)
GIC items	
Definition 1: improvement of ≥ 2 categories	• "Very much better" or "much better" (PGIC) • "Very much better" or "moderately better" (CGIC; CaGIC)
Definition 2: improvement of ≥ 1 category	• "Very much better," "much better," or "a little better" (PGIC) • "Very much better," "moderately better," or "a little better" (CGIC; CaGIC)
Definition 3: improvement of 1 or 2 categories	• "Much better" or "A little better" (PGIC) • "Moderately better" or "A little better" (CGIC; CaGIC)
GIS items	
Definition 1: improvement of ≥ 2 categories	• Improvement of ≥ 2 categories on GIS scale from baseline to week 24; for example, a change from "severe" to "mild" (CaGIS; CGIS)
Definition 2: improvement of ≥ 1 category	• Improvement of ≥ 1 category in GIS scale from baseline to week 24; for example, a change from "moderate" to "mild" (CaGIS; CGIS)

Table 1 continued

Definition name	Responses indicating improvement (anchor)
Definition 3: improvement of 1 or 2 categories	• Improvement of 1 or 2 categories on GIS scale from baseline to week 24; for example, a change from "severe" to "moderate" or "severe" to "mild" (CaGIS; CGIS)

GIC items ask about the degree of change since starting the study medication. GIS items ask about symptom severity in the week before the assessment

CaGIC Caregiver Global Impression of Change, *CaGIS* Caregiver Global Impression of Symptoms, *CGIC* Clinician Global Impression of Change, *CGIS* Clinician Global Impression of Symptoms, *GIC* Global Impression of Change, *GIS* Global Impression of Symptoms, *PGIC* Patient Global Impression of Change

Table 2 PEDFIC 1 patient demographics and baseline characteristics

Characteristic	Participants (*N* = 62)
Age (years)	
Mean (SD)	4.2 (3.9)
Median (range)	3.2 (1–16)
Female, *n* (%)	31 (50)
Race, *n* (%)	
White	52 (84)
Black/African American	2 (3)
Asian	2 (3)
Other	6 (10)
PFIC type, *n* (%)	
PFIC1	17 (27)
PFIC2	45 (73)

PFIC progressive familial intrahepatic cholestasis, *SD* standard deviation

31% aged < 2 years and 2–4 years, respectively; these patients did not contribute PRO data to the analysis.

Instrument Completion Rates

At baseline, a total of 10 patients aged ≥ 8 years were eligible to complete the PRO; however, one eligible patient was unable to complete this measure due to disability. A total of 62 caregivers were asked to complete the ObsRO. Overall, the eDiary completion rates at baseline, indicating the percentage of scheduled daytime and nighttime scores that were recorded, were high (PRO: ≥ 79%; ObsRO: ≥ 96%). The completion rates for the PRO and ObsRO at week 24 were ≥ 76% and ≥ 93%, respectively.

Descriptive Analysis at Baseline

Baseline PRO itching and ObsRO scratching scores in PEDFIC 1 indicated that patients experienced substantial levels of itching/scratching, as most itching/scratching baseline scores were between 2 and 4 (Table 3). For PRO sleep disturbance and tiredness items, at least half the patients had average daily scores of 3–4 at baseline (Table 3). Approximately 60% of caregiver respondents reported that their child needed help falling asleep as a result of itching, needed soothing, or slept with the caregiver (i.e., co-sleep) all days during the baseline interval. For the itching/scratching and sleep disturbance PRO/ObsRO items, no floor effects were observed; however, ceiling effects (i.e., ≥ 25% of patients with average daily scores between 3 and 4) were observed.

Psychometric Measurement Properties— PRO Instrument

The small sample size of patients who completed the PRO precluded a full psychometric validation of the PRO instrument.

Psychometric Measurement Properties— ObsRO Instrument

ObsRO Reliability The mean (SD) ObsRO morning and bedtime pruritus scores at baseline (n = 62, each) were 2.84 (0.66) and 3.01 (0.57), respectively, and a strong inter-item correlation was found between these scores (r = 0.81).

Table 3 Descriptive statistics for PRO and ObsRO daily pruritus scores at baseline

Baseline score[a]	n	Mean (SD)
Pruritus daily score[b]		
Itching (PRO)	9	2.8 (0.5)
Scratching (ObsRO)	62	2.9 (0.6)
Sleep/tiredness (PRO)		
Sleep disturbance—fall asleep[b,c]	9	2.9 (0.7)
Sleep disturbance—stay asleep[b,d]	9	2.8 (1.0)
Sleep disturbance—wake up[e,f]	9	91 (24)
Tiredness upon waking[b,g]	9	3.1 (0.7)
Daytime tiredness[b,h]	9	2.9 (0.6)
Sleep/tiredness (ObsRO)		
Sleep disturbance—blood[e,i]	62	44 (37)
Sleep disturbance—fall asleep[e,j]	62	79 (37)
Sleep disturbance—soothe[e,k]	62	80 (36)
Sleep disturbance—co-sleep[e,l]	62	68 (42)
Sleep disturbance—awakenings[m,n]	62	9.4 (17)
Daytime tiredness[b,o]	62	2.4 (0.9)

ObsRO observer-reported outcome, *PRO* patient-reported outcome, *SD* standard deviation

[a]Baseline scores were derived using the last 14 days prior to the first dose of study medication

[b]Possible scores ranged from 0 to 4, with higher scores indicating worse symptoms

[c]Corresponds to PRO morning diary question 2

[d]Corresponds to PRO morning diary question 3

[e]Possible responses were "yes" or "no"; percentage of days participant answered "yes" was summarized

[f]Corresponds to PRO morning diary question 4

[g]Corresponds to PRO morning diary question 5

[h]Corresponds to PRO evening diary question 2

[i]Corresponds to ObsRO morning diary question 2

[j]Corresponds to ObsRO morning diary question 3

[k]Corresponds to ObsRO morning diary question 4

[l]Corresponds to ObsRO morning diary question 5

[m]Corresponds to ObsRO morning diary question 6

[n]Possible responses ranged from 0 to 99

[o]Corresponds to ObsRO evening diary question 2

On the ObsRO, moderate-to-strong correlations were found between daytime tiredness and all three scratching scores (r values ranged from 0.58 to 0.65), and strong correlations were observed between sleep disturbance items of needing help falling asleep, needing soothing, and sleeping with a caregiver (r values ranged from 0.68 to 0.87).

Comparisons between weekly ObsRO pruritus and sleep disturbance scores from the first week ("test") to the second week ("retest") of the baseline interval yielded ICCs $\geq$ 0.75, indicating acceptable reliability (Table 4) [13]. Comparisons between baseline and week 4 ObsRO pruritus and sleep disturbance scores in stable patients identified by the CaGIC yielded ICC values $\geq$ 0.69; these values were $\geq$ 0.54 when stable patients were identified by the CaGIS (Table 4).

ObsRO Construct Validity Moderate-to-strong correlations ($r \geq 0.39$) between ObsRO scratching and the GIS were observed (Table 5),

demonstrating convergent validity. Moderate-to-strong correlations were also observed for several ObsRO sleep disturbance items and PGIS scratch scores (Table 5). In addition, strong correlations were observed between the ObsRO sleep disturbance items and PedsQL self-reported total scores (Table 6). Some ObsRO sleep disturbance items at baseline had moderate correlations with baseline PedsQL caregiver-reported school domain scores (Table 6).

For known-groups validity analyses, group differences changed in the expected direction (e.g., ObsRO pruritus scores increased as scores on the GIS indicated increased severity of pruritus and sleep disturbance) for all pruritus scores. These differences were significant for the ObsRO scratching and some sleep disturbance items when groups were based on CaGIS and CGIS scores (Table 7).

ObsRO Sensitivity to Change Mean differences in ObsRO scratching, sleep disturbance, and tiredness scores between "improved" and

Table 4 Test–retest reliability of ObsRO scores

ObsRO score	First (days −14 to −8) vs. second (days −7 to −1) week of baseline			Baseline (average of days −14 to −1) vs. week 4 in stable[a] patients					
				CaGIC			CaGIS		
	n	ICC	95% CI	*n*	ICC	95% CI	*n*	ICC	95% CI
Daily scratching	62	0.87	0.79, 0.92	20	0.80	0.56, 0.91	21	0.66	0.18, 0.86
Daytime scratching	62	0.88	0.82, 0.93	20	0.79	0.54, 0.91	21	0.54	0.12, 0.79
Nighttime scratching	62	0.77	0.64, 0.85	20	0.75	0.48, 0.89	21	0.74	0.29, 0.90
Sleep disturbance—blood	62	0.81	0.71, 0.88	24	0.69	0.41, 0.85	16	0.83	0.49, 0.94
Sleep disturbance—fall asleep	62	0.95	0.91, 0.97	24	0.90	0.78, 0.95	16	0.95	0.86, 0.98
Sleep disturbance—soothe	62	0.98	0.97, 0.99	24	0.90	0.78, 0.95	16	0.86	0.62, 0.95
Sleep disturbance—co-sleep[b]	62	0.96	0.94, 0.98	24	0.85	0.68, 0.93	16	0.88	0.64, 0.96
Sleep disturbance—awakenings	62	0.95	0.92, 0.97	24	0.72	0.46, 0.87	16	0.97	0.92, 0.99
Daytime tiredness	62	0.84	0.75, 0.90	24	0.77	0.42, 0.91	16	0.91	0.61, 0.97

CaGIC Caregiver Global Impression of Change, *CaGIS* Caregiver Global Impression of Symptoms, *ICC* intraclass correlation coefficient, *ObsRO* observer-reported outcome
[a]Stable patients were defined as those with responses of "no change" on the GIC at baseline and week 4 or the same score on the GIS at baseline and week 4
[b]Refers to patient needing to sleep with caregiver

Table 5 Convergent validity: correlations between baseline ObsRO scores and baseline GIS items

ObsRO score[a]	Scratch item			Sleep item		
	PGIS	CaGIS	CGIS	PGIS	CaGIS	CGIS
Daily scratching	0.55	0.55	0.46	0.48	0.56	0.47
Daytime scratching	0.39	0.50	0.42	0.44	0.45	0.44
Nighttime scratching	0.64	0.60	0.51	0.58	0.66	0.51
Sleep disturbance—blood	0.47	0.30	0.37	0.66	0.38	0.23
Sleep disturbance—fall asleep	0.60	−0.01	0.02	0.06	0.26	0.32
Sleep disturbance—soothe	0.63	0.05	0.16	0.13	0.29	0.41
Sleep disturbance—co-sleep[b]	0.53	−0.03	0.08	−0.05	0.34	0.30
Sleep disturbance—awakenings	0.44	0.23	0.12	0.07	0.52	0.21
Daytime tiredness	0.57	0.21	0.21	0.34	0.30	0.23

Table reports Spearman correlation coefficients. Moderate correlation (orange fill): $0.3 \leq |r| < 0.5$; strong correlation (green fill): $|r| \geq 0.5$

CaGIS Caregiver Global Impression of Symptoms, *CGIS* Clinician Global Impression of Symptoms, *GIS* Global Impression of Symptoms, *ObsRO* observer-reported outcome, *PGIS* Patient Global Impression of Symptoms

[a]Number of patients in this analysis: PGIS, *n* = 8; CaGIS, *n* = 58; CGIS, *n* = 55

[b]Refers to patient needing to sleep with caregiver

Table 6 Convergent validity: correlations between baseline ObsRO and baseline PedsQL self-reported or caregiver-reported scores

ObsRO score[a]	PedsQL self-reported scores					PedsQL caregiver-reported scores				
	Total score	Physical	Emotional	Social	School	Total score	Physical	Emotional	Social	School
Daily scratching	−0.27	−0.07	−0.06	−0.50	−0.18	−0.22	−0.22	−0.16	−0.08	−0.24
Daytime scratching	−0.15	0.08	0.01	−0.43	−0.07	−0.13	−0.17	−0.08	−0.01	−0.20
Nighttime scratching	−0.18	−0.02	−0.04	−0.60	−0.19	−0.30	−0.25	−0.30	−0.12	−0.40
Sleep disturbance—blood	−0.09	−0.26	−0.15	0.29	−0.11	0.01	0.18	−0.14	−0.12	−0.27
Sleep disturbance—fall asleep	−0.79	−0.75	−0.77	−0.59	−0.32	−0.19	0.01	−0.18	−0.20	−0.44
Sleep disturbance—soothe	−0.79	−0.73	−0.80	−0.57	−0.31	−0.21	0.01	−0.20	−0.19	−0.45
Sleep disturbance—co-sleep[b]	−0.83	−0.83	−0.81	−0.52	−0.38	−0.21	0.02	−0.22	−0.24	−0.48
Sleep disturbance—awakenings	−0.24	−0.10	−0.17	−0.32	−0.41	−0.20	−0.11	−0.25	0.08	−0.18
Daytime tiredness	−0.29	−0.06	0.02	−0.38	−0.30	−0.13	−0.25	−0.05	0.07	−0.10

Table reports Spearman correlation coefficients. Moderate correlation (orange fill): $0.3 \leq |r| < 0.5$; strong correlation (green fill): $|r| \geq 0.5$

ObsRO observer-reported outcome, *PedsQL* Pediatric Quality of Life Inventory

[a]Number of patients in this analysis: self-report, *n* = 10; caregiver report, *n* = 38, except school functioning, where *n* = 25

[b]Refers to patient needing to sleep with caregiver

Table 7 Known-groups validity: mean differences in baseline ObsRO scratching scores between groups defined by the GIS

ObsRO score	Groups defined by the CaGIS				P value[a]	Groups defined by the CGIS				P value[a]
	None, mild, or moderate		Severe/very severe			None, mild, or moderate		Severe/very severe		
	n	Mean (SD)	n	Mean (SD)		n	Mean (SD)	n	Mean (SD)	
Daily scratching[b]	12	2.4 (0.5)	46	3.1 (0.5)	**< 0.001**	13	2.5 (0.6)	42	3.1 (0.6)	**0.005**
Daytime scratching[b]	12	2.3 (0.6)	46	3.0 (0.6)	**0.003**	13	2.4 (0.7)	42	3.0 (0.6)	**0.016**
Nighttime scratching[b]	12	2.5 (0.4)	46	3.2 (0.5)	**< 0.001**	13	2.6 (0.5)	42	3.2 (0.5)	**0.001**
Sleep disturbance—blood[c]	13	27.8 (27)	45	47.7 (38)	**0.041**	17	38.3 (32)	38	48.2 (38)	0.323
Sleep disturbance—fall asleep[c]	13	57.5 (48)	45	86.1 (29)	0.061	17	61.1 (47)	38	86.4 (28)	0.053
Sleep disturbance—soothe[c]	13	57.1 (49)	45	87.3 (28)	**0.050**	17	60.7 (48)	38	87.9 (27)	**0.040**
Sleep disturbance—co-sleep[c,d]	13	45.1 (48)	45	76.5 (37)	**0.043**	17	56.0 (49)	38	77.2 (36)	0.121
Sleep disturbance—awakenings[c]	13	4.1 (4.8)	45	11.5 (19)	**0.021**	17	6.6 (9.8)	38	11.7 (20)	0.208
Daytime tiredness[c]	13	2.2 (0.8)	45	2.6 (0.9)	0.235	17	2.3 (0.7)	38	2.5 (0.9)	0.256

P values ≤ 0.05 are bolded and indicate statistical significance

CaGIS Caregiver Global Impression of Symptoms, *CGIS* Clinician Global Impression of Symptoms, *GIS* Global Impression of Symptoms, *ObsRO* observer-reported outcome, *SD* standard deviation

[a]P values are based on a Welch *t* test

[b]GIS item: scratch

[c]GIS item: sleep

[d]Refers to patient needing to sleep with caregiver

"not improved" groups from the GIC were significant and in the expected direction (Table 8). For example, there were greater mean reductions in ObsRO scratching scores in the "improved" group per the CaGIC versus the "not improved" group at week 24. Similar results were found when groups were identified using the CGIC. Additionally, moderate-to-strong correlations ($r \geq 0.3$) were found between ObsRO scratching scores and the GIC and GIS as reported by patients, caregivers, and clinicians (Table 9). In general, moderate-to-strong correlations were also identified between ObsRO tiredness and sleep disturbance items and the GIC and GIS (Table 9). Strong correlations were also identified between ObsRO scratching and some sleep disturbance items and PedsQL self-reported scores for physical functioning

(Table 10). Additionally, moderate-to-strong correlations were observed between ObsRO scratching and some sleep disturbance items and PedsQL caregiver-reported scores, particularly in the school domain. ObsRO scratching and daytime tiredness items were strongly correlated with the PedsQL family impact module total scores and domain scores for physical functioning and daily activities (Table 11).

ObsRO Threshold for Clinically Meaningful Change in Pruritus Distribution-based analyses showed that the 0.5 SD of the baseline daily ObsRO scratching score was 0.30, and 1 SEM at baseline was 0.21. These values served as a lower limit for the meaningful change threshold. Anchor-based analyses supported the use of CaGIS as an anchor for establishing a

Table 8 Sensitivity to change analyses: mean change from baseline to weeks 21–24 in ObsRO pruritus scores by GIC response category at week 24

ObsRO score	Item	Response categories				P value[c]
		Not improved[a]		Improved[b]		
		n	LS Mean (SE)	n	LS Mean (SE)	
Daily scratching	CaGIC (Scratch)	11	0.19 (0.29)	29	−1.18 (0.18)	**< 0.001**
Daytime scratching	CaGIC (Scratch)	11	0.14 (0.29)	29	−1.14 (0.18)	**< 0.001**
Nighttime scratching	CaGIC (Scratch)	11	0.25 (0.29)	29	−1.23 (0.18)	**< 0.001**
Sleep disturbance—blood	CaGIC (Sleep)	13	0.91 (6.6)	27	−36.3 (4.6)	**< 0.001**
Sleep disturbance—fall asleep	CaGIC (Sleep)	13	2.10 (10.5)	27	−49.1 (7.2)	**< 0.001**
Sleep disturbance—soothe	CaGIC (Sleep)	13	0.98 (9.9)	27	−52.7 (6.8)	**< 0.001**
Sleep disturbance—co-sleep[d]	CaGIC (Sleep)	13	–2.78 (9.6)	27	−49.7 (6.6)	**< 0.001**
Sleep disturbance—awakenings	CaGIC (Sleep)	13	0.53 (4.6)	27	0.65 (3.17)	0.984
Daytime tiredness	CaGIC (Sleep)	13	0.19 (0.25)	27	−1.30 (0.17)	**< 0.001**
Daily scratching	CGIC (Scratch)	14	0.13 (0.25)	32	−1.24 (0.17)	**< 0.001**
Daytime scratching	CGIC (Scratch)	14	0.12 (0.25)	32	−1.21 (0.16)	**< 0.001**
Nighttime scratching	CGIC (Scratch)	14	0.16 (0.26)	32	−1.27 (0.17)	**< 0.001**
Sleep disturbance—blood	CGIC (Sleep)	16	–4.83 (6.5)	30	−35.4 (4.7)	**< 0.001**
Sleep disturbance—fall asleep	CGIC (Sleep)	16	0.95 (9.5)	30	−48.6 (6.8)	**< 0.001**
Sleep disturbance—soothed	CGIC (Sleep)	16	−0.49 (9.1)	30	−51.2 (6.5)	**< 0.001**
Sleep disturbance—co-sleep[d]	CGIC (Sleep)	16	–4.47 (8.7)	30	−48.6 (6.3)	**< 0.001**
Sleep disturbance—awakenings	CGIC (Sleep)	16	–1.56 (3.8)	30	0.42 (2.80)	0.680
Daytime tiredness	CGIC (Sleep)	16	0.11 (0.22)	30	−1.35 (0.16)	**< 0.001**

P values ≤ 0.05 are bolded and indicate statistical significance

ANCOVA analysis of covariance, *CaGIC* Caregiver Global Impression of Change, *CGIC* Clinician Global Impression of Change, *GIC* Global Impression of Change, *LS* least squares, *ObsRO* observer-reported outcome, *PGIC* Patient Global Impression of Change, *SE* standard error

[a]"Not improved" includes all respondents who answered "no change," "a little worse," "much worse," or "very much worse" at week 24 on PGIC, CaGIC, or CGIC

[b]"Improved" includes all respondents who answered "a little better," "much better," or "very much better" at week 24 on PGIC, CaGIC, or CGIC

[c]*P* value is based on an ANCOVA model comparing "not improved" and "improved" groups

[d]Refers to patient needing to sleep with caregiver

meaningful change in the ObsRO scratching score (Table 12). The smallest median ObsRO scratching change values for the CaGIS that exceeded the values from the distribution-based analyses and the lower bound of the 95% CI from the stable anchor category were −0.95 and −0.96, for the monthly and biweekly ObsRO pruritus scores, respectively, at week 24 (Table 13). These and other anchor-based and ROC analyses are included in Table 13.

The eCDF curves for change in ObsRO pruritus score shifted left with increasing

Table 9 Sensitivity to change analyses: correlation of change from baseline to weeks 21–24 in ObsRO scores and GIC or GIS items at week 24

ObsRO score	Pearson correlation coefficient (n)					
	PGIC	CaGIC	CGIC	PGIS	CaGIS	CGIS
Daily scratching[a]	0.57 (5)	0.59 (36)	0.62 (42)	0.83 (5)	0.71 (35)	0.52 (38)
Daytime scratching[a]	0.48 (5)	0.53 (36)	0.61 (42)	0.72 (5)	0.63 (35)	0.48 (38)
Nighttime scratching[a]	0.63 (5)	0.63 (36)	0.61 (42)	0.92 (5)	0.77 (35)	0.54 (38)
Sleep disturbance—blood[b]	0.22 (7)	0.52 (40)	0.44 (46)	0.43 (7)	0.59 (39)	0.32 (42)
Sleep disturbance—fall asleep[b]	0.78 (7)	0.58 (40)	0.54 (46)	0.76 (7)	0.61 (39)	0.47 (42)
Sleep disturbance—soothe[b]	0.77 (7)	0.64 (40)	0.62 (46)	0.76 (7)	0.60 (39)	0.53 (42)
Sleep disturbance—co-sleep[b,c]	0.62 (7)	0.59 (40)	0.59 (46)	0.65 (7)	0.57 (39)	0.48 (42)
Sleep disturbance—awakenings[b]	−0.27 (7)	−0.00 (40)	−0.12 (46)	−0.49 (7)	0.14 (39)	−0.04 (42)
Daytime tiredness[b]	0.88 (7)	0.57 (40)	0.59 (46)	0.93 (7)	0.66 (39)	0.48 (42)

Moderate correlation (orange fill): $0.3 \leq |r| < 0.5$; strong correlation (green fill): $|r| \geq 0.5$

CaGIC Caregiver Global Impression of Change, *CaGIS* Caregiver Global Impression of Symptoms, *CGIC* Clinician Global Impression of Change, *CGIS* Caregiver Global Impression of Symptoms, *GIC* Global Impression of Change, *GIS* Global Impression of Symptoms, *ObsRO* observer-reported outcome, *PGIC* Patient Global Impression of Change, *PGIS* Patient Global Impression of Symptoms

[a]GIC/GIS item: itch
[b]GIC/GIS item: sleep
[c]Refers to patient needing to sleep with caregiver

improvement in CaGIS score (Fig. 3). Based on these analyses, a change of −1.00 from baseline to week 24 was selected as the final meaningful change threshold, indicating that a decrease in the average caregiver scratching score of 1 or more points can be considered a clinically meaningful improvement in pruritus.

DISCUSSION

The results of this study indicate that a novel clinical outcome assessment, the PRUCISION ObsRO instrument, is a valid and reliable measure able to capture longitudinal changes in the severity of pruritus and sleep disturbance in patients with PFIC.

Using data from the phase 3 PEDFIC 1 study in patients with PFIC, the measurement characteristics of the ObsRO PRUCISION instrument, as well as the PRO PRUCISION instrument, were evaluated. PRO/ObsRO scores at PEDFIC 1 baseline indicated that these patients experienced significant pruritus and associated sleep disturbance prior to treatment intervention. Examination of baseline distribution data from PEDFIC 1 indicated that the PRUCISION instruments had ceiling effects (i.e., they may not be as sensitive to worsening over time), but that they were capable of detecting improvement as a result of treatment. Further analysis of the psychometric properties of the PRO was limited due to the small number of patients who completed the instrument during the study. Psychometric analysis of the PRUCISION ObsRO instrument demonstrated that the instrument had acceptable test–retest reliability overall. Further analyses supported the construct validity and sensitivity to change of the ObsRO instrument. The anchor-based analyses with ObsRO scores performed here indicated that a 1-point reduction corresponds to a meaningful improvement in pruritus.

Table 10 Sensitivity to change analyses: correlation between change from baseline to weeks 21–24 in ObsRO score and PedsQL self-reported and caregiver-reported scores at week 24

ObsRO score[a]	PedsQL self-reported scores					PedsQL caregiver-reported scores				
	Total score	Physical	Emotional	Social	School	Total score	Physical	Emotional	Social	School
Daily scratching	−0.48	−0.56	−0.45	−0.44	−0.24	−0.45	−0.36	−0.32	−0.36	−0.29
Daytime scratching	−0.49	−0.58	−0.46	−0.50	−0.21	−0.44	−0.38	−0.34	−0.30	−0.29
Nighttime scratching	−0.45	−0.52	−0.42	−0.37	−0.27	−0.43	−0.32	−0.28	−0.40	−0.27
Sleep disturbance—blood	−0.28	−0.42	−0.23	−0.16	−0.13	−0.40	−0.20	−0.36	−0.50	−0.22
Sleep disturbance—fall asleep	−0.51	−0.57	−0.62	−0.24	−0.36	−0.29	−0.16	−0.17	−0.30	−0.30
Sleep disturbance—soothe	−0.53	−0.58	−0.64	−0.26	−0.35	−0.29	−0.15	−0.21	−0.34	−0.23
Sleep disturbance—co-sleep[b]	−0.53	−0.60	−0.67	−0.31	−0.26	−0.33	−0.21	−0.23	−0.39	−0.21
Sleep disturbance—awakenings	−0.00	0.13	0.10	0.16	−0.38	−0.52	−0.43	−0.57	−0.39	−0.20
Daytime tiredness	−0.21	−0.32	−0.35	−0.12	0.09	−0.19	−0.12	−0.31	−0.16	0.03

Pearson correlation coefficient. Moderate correlation (orange fill): $0.3 \leq |r| < 0.5$; strong correlation (green fill): $|r| \geq 0.5$

ObsRO observer-reported outcome, *PedsQL* Pediatric Quality of Life Inventory

[a]Number of patients in this analysis: self-report, $n = 7$; caregiver report, $n = 23$, except school functioning, where $n = 16$

[b]Refers to patient needing to sleep with caregiver

The PRUCISION PRO/ObsRO instruments were developed to address a lack of adequate measurement tools for quantifying pruritus in pediatric patients with cholestatic liver disease. The analyses used for developing and validating the PRUCISION PRO/ObsRO instruments followed best practice guidelines to assess reliability, construct validity, and sensitivity to change [10, 15, 16]. The findings described here support the use of the ObsRO PRUCISION instrument to measure potential treatment benefits of odevixibat in patients with PFIC. In addition, because the PRUCISION instruments were developed with input from caregivers and patients with a range of cholestatic liver disorders (i.e., Alagille syndrome, biliary atresia, and primary sclerosing cholangitis; see companion article in this issue), these instruments can be applied to measure symptoms of pruritus and sleep disturbance in children with other cholestatic liver diseases, such as those associated with significant pruritus [5]. In fact, the primary outcome of ASSERT (NCT04674761), an interventional study in patients with Alagille syndrome that was initiated in 2021, is based on ObsRO PRUCISION scores. The PRUCISION instruments may also be useful in monitoring symptoms over time in patients who receive no or other treatment options.

Some limitations of this study warrant discussion. First, validation of the PRUCISION PRO instrument was limited by the small sample size of study participants aged ≥ 8 years who could complete the PRO. Additionally, another instrument, the Itch Reported Outcome (ItchRO) tool, was being developed while this study was underway [17]. The ItchRO tool was also intended to measure cholestatic pruritus in children and includes a 5-point response scale that ranges from 0–4 [18]. However, at the time the PEDFIC 1 study was designed, this tool was not publicly available for use or adaptation. Both the PRUCISION PRO/ObsRO instruments and the ItchRO tool contain morning and bedtime pruritus assessments [18], and validation analyses found that a ≥ 1-point reduction

Table 11 Sensitivity to change: correlation between change from baseline to weeks 21–24 in ObsRO score and PedsQL family impact module scores at week 24

ObsRO score[a]	Total score	Physical	Emotional	Social	Cognitive	Communication	Worry	Daily activities	Family relation- ships
Daily scratching	−0.65	−0.60	−0.60	−0.52	−0.48	−0.44	−0.64	−0.65	−0.45
Daytime scratching	−0.61	−0.55	−0.54	−0.46	−0.50	−0.42	−0.60	−0.65	−0.44
Nighttime scratching	−0.67	−0.64	−0.64	−0.56	−0.46	−0.46	−0.66	−0.64	−0.45
Sleep disturbance—blood	−0.58	−0.54	−0.53	−0.52	−0.29	−0.42	−0.63	−0.64	−0.39
Sleep disturbance—fall asleep	−0.44	−0.44	−0.38	−0.36	−0.35	−0.33	−0.37	−0.53	−0.27
Sleep disturbance—soothe	−0.39	−0.41	−0.32	−0.27	−0.33	−0.26	−0.36	−0.49	−0.24
Sleep disturbance—co-sleep[b]	−0.38	−0.43	−0.30	−0.27	−0.31	−0.25	−0.33	−0.41	−0.29
Sleep disturbance—awakenings	−0.43	−0.38	−0.34	−0.32	−0.31	−0.46	−0.25	−0.37	−0.53
Daytime tiredness	−0.58	−0.53	−0.49	−0.45	−0.47	−0.45	−0.50	−0.59	−0.48

Pearson correlation coefficient. Moderate correlation (orange fill): $0.3 \leq |r| < 0.5$; strong correlation (green fill): $|r| \geq 0.5$

ObsRO observer-reported outcome, *PedsQL* Pediatric Quality of Life Inventory

[a]Number of patients in this analysis: $n = 24$

[b]Refers to patient needing to sleep with caregiver

Table 12 Correlations between change from baseline to week 24 in ObsRO scratching measures and anchors

Time point	Week 24		
	PGIS	CaGIS	CGIS
(n) Correlations with Global Impression of Severity anchor			
Monthly	(5) 0.83	**(35) 0.71**	(38) 0.52
Biweekly	(5) 0.86	**(35) 0.72**	(37) 0.54
	PGIC	CaGIC	CGIC
(n) Correlations with Global Impression of Change anchor			
Monthly	(5) 0.57	(36) 0.59	(42) 0.62
Biweekly	(5) 0.62	(36) 0.59	(40) 0.63

Pearson correlation coefficient. Bolded values indicate items selected as primary anchors for analysis

CaGIC Caregiver Global Impression of Change, *CaGIS* Caregiver Global Impression of Symptoms, *CGIC* Clinician Global Impression of Change, *CGIS* Clinician Global Impression of Symptoms, *ObsRO* observer-reported outcome, *PGIC* Patient Global Impression of Change, *PGIS* Patient Global Impression of Symptoms

in observer-reported pruritus scores in both tools can be considered a clinically meaningful change [17]. However, the PRUCISION contains unique questions, and the PRUCISION and ItchRO instruments were primarily validated in different patient populations (i.e., PRUCISION, in patients with PFIC; ItchRO, in patients with Alagille syndrome) [17].

Accumulation of bile acids in the liver and secondary spillover into the systemic circulation is another key feature of PFIC and other cholestatic liver diseases in children [2]. Although not completely understood, the higher the level of serum bile acids, the greater the likelihood that patients may have pruritus [19, 20]. Therefore, future studies that include detailed investigations into the relationship between PRUCISION scores and measurements of serum bile acids would be valuable. In an initial post hoc analysis of pooled data from odevixibat-treated patients in the PEDFIC 1 and open-label extension PEDFIC 2 (NCT03659916) studies up to a data cutoff date of December 2020, a significant correlation was found between mean percentage change from baseline

Table 13 Summary of anchor-based and receiver operating characteristic analyses of daily ObsRO scratching score

Anchor/analysis	Statistic	Monthly score, weeks 21–24	Biweekly score, weeks 23–24
Anchor-based analyses			
CaGIS: improved—definition 1	Mean; median (*n*)	−1.79; −1.76 (11)	−1.81; −1.85 (11)
CaGIS: improved—definition 2	Mean; median (*n*)	−1.23; −1.05 (21)	−1.24; −1.15 (21)
CaGIS: improved—definition 3	Mean; median (*n*)	**−0.72; −0.95 (15)**	**−0.73; −0.96 (15)**
CaGIS: no change	95% CI (*n*)	−0.86; 0.42 (8)	−0.91; 0.36 (8)
CGIS: improved—definition 1	Mean; median (*n*)	−1.58; −1.43 (14)	−1.67; −1.41 (13)
CGIS: improved—definition 2	Mean; median (*n*)	**−0.99; −0.93 (23)**	**−1.02; −0.90 (22)**
CGIS: improved—definition 3	Mean; median (*n*)	−0.60; −0.50 (18)	−0.60; −0.43 (17)
CGIS: no change	95% CI (*n*)	−0.77; −0.18 (14)	−0.71; −0.12 (14)
CaGIC: improved—definition 1	Mean; median (*n*)	−1.45; −1.37 (15)	−1.47; −1.23 (15)
CaGIC: improved—definition 2	Mean; median (*n*)	−1.17; −1.02 (24)	−1.18; −1.08 (24)
CaGIC: improved—definition 3	Mean; median (*n*)	**−0.59; −0.80 (13)**	**−0.60; −0.85 (13)**
CaGIC: no change	95% CI (*n*)	−0.17; 0.76 (6)	−0.11; 0.75 (6)
CGIC: improved—definition 1	Mean; median (*n*)	−1.53; −1.41 (17)	−1.57; −1.34 (17)
CGIC: improved—definition 2	Mean; median (*n*)	−1.24; −1.05 (27)	−1.23; −1.00 (27)
CGIC: improved—definition 3	Mean; median (*n*)	**−0.69; −0.78 (16)**	**−0.66; −0.70 (16)**
CGIC: no change	95% CI (*n*)	−0.15; 0.51 (14)	−0.03; 0.61 (12)
Receiver operating characteristics analysis			
CaGIS		−1.00	−1.00
CGIS		−0.93	−1.00
CaGIC		−0.95	−0.98
CGIC		−0.80	−0.85

Bolded estimates exceed the distribution-based threshold estimates and the lower bound of the 95% CI for the mean change from baseline on the ObsRO scratching score for the stable ("no change") group

CaGIC Caregiver Global Impression of Change, *CaGIS* Caregiver Global Impression of Symptoms, *CGIC* Clinician Global Impression of Change, *CGIS* Clinician Global Impression of Symptoms, *CI* confidence interval, *ObsRO* observer-reported outcome

up to week 72 in serum bile acids and mean change in ObsRO PRUCISION score during the same interval [21]. Longer-term data that could address this relationship further are expected as PEDFIC 2 completes in 2023.

The results from this study support the validity of the PRUCISION ObsRO instrument for measuring pruritus and sleep disturbance in children with PFIC. In addition, based on the analyses conducted here, a reduction in ObsRO pruritus score of −1.00 can be considered a clinically meaningful change. The ObsRO PRUCISION instrument is appropriate for evaluating the effect of treatment on pruritus and sleep disturbance in PFIC and other pediatric cholestatic liver diseases.
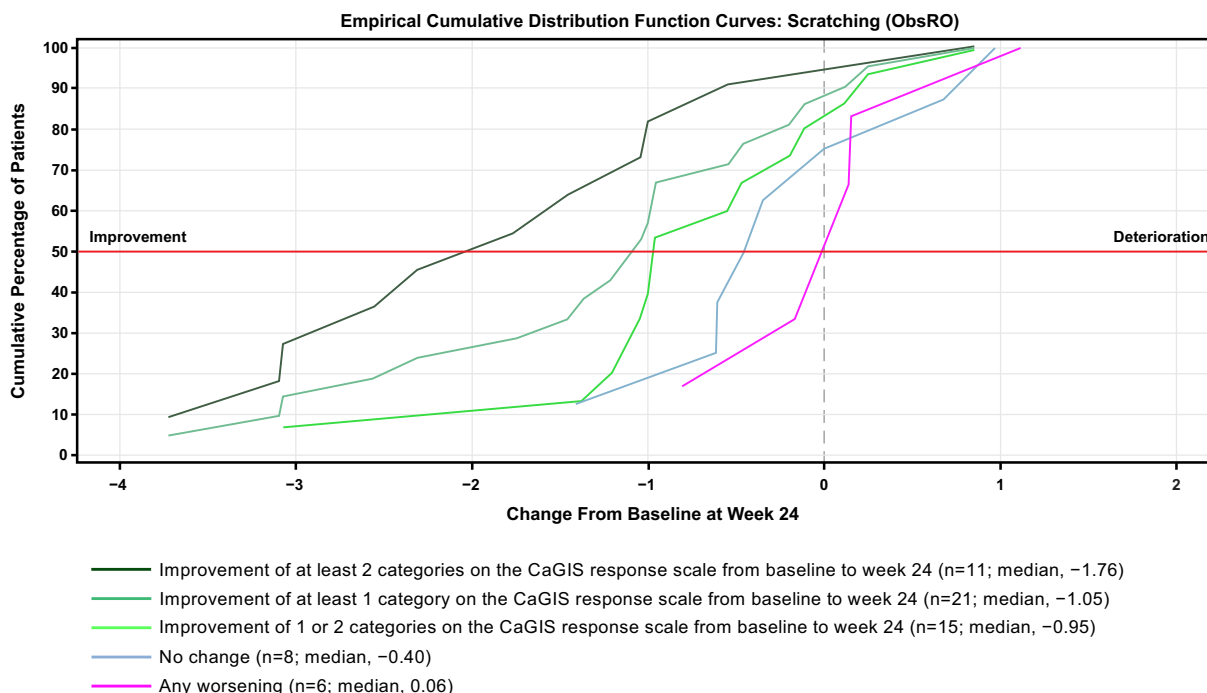
Fig. 3 eCDF plots for change from baseline in daily ObsRO pruritus monthly score by CaGIS at week 24. *CaGIS* Caregiver Global Impression of Symptoms, *eCDF* empirical cumulative distribution function, *ObsRO* observer-reported outcome

ACKNOWLEDGEMENTS

We thank the participants of the study, including the patients, caregivers, and investigators from PEDFIC 1, as well as those who participated in the development and validation of the PRUCISION instruments.

Compliance with Ethics Guidelines. PEDFIC 1 study materials received CEC or LEC approvals from all 45 sites that intended to enroll patients. The study was conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonization Good Clinical Practice guidelines. All patients or their caregivers provided written consent prior to study participation.

Data Availability. Qualified academic investigators and researchers may request additional data and supporting documents

pertaining to this study. For details regarding data availability, instructions for requesting information, and our data disclosure policy, please email medinfo@albireopharma.com.

REFERENCES

1. Data on file. Boston: Albireo Pharma, Inc; 2022.

2. Kamath BM, Stein P, Houwen RHJ, Verkade HJ. Potential of ileal bile acid transporter inhibition as a therapeutic target in Alagille syndrome and progressive familial intrahepatic cholestasis. Liver Int. 2020;40(8):1812–22. https://doi.org/10.1111/liv.14553.

3. Karpen SJ, Kelly D, Mack C, Stein P. Ileal bile acid transporter inhibition as an anticholestatic therapeutic target in biliary atresia and other cholestatic disorders. Hepatol Int. 2020;14(5):677–89. https://doi.org/10.1007/s12072-020-10070-w.

4. Gunaydin M, Bozkurter Cil AT. Progressive familial intrahepatic cholestasis: diagnosis, management, and treatment. Hepatic Med. 2018;10:95–104. https://doi.org/10.2147/hmer.S137209.

5. Thebaut A, Debray D, Gonzales E. An update on the physiopathology and therapeutic management of cholestatic pruritus in children. Clin Res Hepatol Gastroenterol. 2018;42(2):103–9. https://doi.org/10.1016/j.clinre.2017.08.007.

6. Kremer AE, Oude Elferink RP, Beuers U. Pathophysiology and current management of pruritus in liver disease. Clin Res Hepatol Gastroenterol. 2011;35(2):89–97. https://doi.org/10.1016/j.clinre.2010.10.007.

7. Baker A, Kerkar N, Todorova L, Kamath BM, Houwen RHJ. Systematic review of progressive familial intrahepatic cholestasis. Clin Res Hepatol Gastroenterol. 2019;43(1):20–36. https://doi.org/10.1016/j.clinre.2018.07.010.

8. Abetz-Webb L, Kennedy C, Hepburn B, Gauthier M, Johnson N, Madendorp S, et al. The burden of pruritus on patients with Alagille syndrome: results from a qualitative study with pediatric patients and their caregivers [abstract 673]. Hepatology. 2014;60(4 suppl):p526A-7A.

9. U.S. Food and Drug Administration. Patient-Focused Drug Development Guidance Public Workshop: Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for Purpose Clinical Outcomes Assessments. 2018. https://www.fda.gov/media/116281/download. Accessed 14 June 2022.

10. U.S. Department of Health and Human Services. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. Accessed 13 June 2022.

11. Thompson RJ, Arnell H, Artan R, Baumann U, Calvo PL, Czubkowski P, et al. Odevixibat treatment in progressive familial intrahepatic cholestasis: a randomised, placebo-controlled, phase 3 trial. Lancet Gastroenterol Hepatol. 2022;7(9):830–42. https://doi.org/10.1016/S2468-1253(22)00093-0.

12. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care. 2001;39(8):800–12. https://doi.org/10.1097/00005650-200108000-00006.

13. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

14. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York: Lawrence Erlbaum Associates; 1988.

15. Committee for Medicinal Products for Human Use. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. https://www.ema.europa.eu/en/documents/scienti

fic-guideline/reflection-paper-regulatory-guidance-use-healthrelated-quality-life-hrql-measures-evaluation_en.pdf. Accessed 13 June 2022.

16. Nunnally JC, Bernstein IH. Psychometric theory 3E. 3rd ed. New York: Tata McGraw-Hill Education; 1994.

17. Serrano D, Gauthier M, Harrington M, Acevedo L. Psychometric validation of the Itch-Reported Outcome (ItchRO^{TM}) assessment in pediatric patients with Alagille syndrome or progressive familial intrahepatic cholestasis [abstract 571]. Hepatology. 2016;64(1 suppl):284A-A285.

18. Kamath BM, Abetz-Webb L, Kennedy C, Hepburn B, Gauthier M, Johnson N, et al. Development of a novel tool to assess the impact of itching in pediatric cholestasis. Patient. 2018;11(1):69–82. https://doi.org/10.1007/s40271-017-0266-4.

19. van Wessel DBE, Thompson RJ, Gonzales E, Jankowska I, Shneider BL, Sokal E, et al. Impact of genotype, serum bile acids, and surgical biliary diversion on native liver survival in FIC1 deficiency. Hepatology. 2021;74(2):892–906. https://doi.org/10.1002/hep.31787.

20. Ayoub MD, Kamath BM. Alagille syndrome: diagnostic challenges and advances in management. Diagnostics (Basel, Switzerland). 2020;10(11):907. https://doi.org/10.3390/diagnostics10110907.

21. Dalgic B, Verkade H, Lachaux A, Baumann U, D'Antiga L, Czubkowski P, et al. Relationships between decreases in serum bile acids, pruritus, and sleep disturbance scores with up to 72 weeks of odevixibat treatment in patients with progressive familial intrahepatic cholestasis [abstract 69]. J Pediatr Gastroenterol Nutr. 2021;73(suppl 1): S49–50.