Received: 5 November 2019 | Revised: 18 December 2019 | Accepted: 20 December 2019

DOI: 10.1111/ejh.13376

Check for updates

ORIGINAL ARTICLE

European Journal of Haematology WILEY

Development of the pyruvate kinase deficiency diary and pyruvate kinase deficiency impact assessment: Disease-specific assessments

Sam Salek[1] | Audra N. Boscoe[2] | Sarah Piantedosi[2] | Shayna Egan[3] | Christopher J Evans[3] | Ted Wells[3] | Jennifer Cohen[4] | Robert J. Klaassen[5] | Rachael Grace[6] | Michael Storm[2]

[1]School of Life and Medical Sciences, University of Hertfordshire, Hatfield, UK

[2]Agios Pharmaceuticals Inc., Cambridge, MA, USA

[3]Endpoint Outcomes, Boston, MA, USA

[4]Patient Advocate, Wrentham, MA, USA

[5]Division of Hematology/Oncology, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada

[6]Dana-Farber/Boston Children's Cancer and Blood Disorders Program, Boston, MA, USA

Correspondence
Shayna Egan, Endpoint Outcomes, 200 Pine Ave, Suite 200, Long Beach, CA 90802, USA.
Email: Shayna.egan@endpointoutcomes.com

Funding information
Agios Pharmaceuticals Inc.

Abstract

Introduction: Currently recommended patient-reported outcome (PRO) measures for patients with pyruvate kinase (PK) deficiency are non-disease-specific. The PK Deficiency Diary (PKDD) and PK Deficiency Impact Assessment (PKDIA) were developed to be more targeted measures for capturing the symptoms and impacts of interest to this patient population.

Methods: The instruments were developed based on concept elicitation interviews with 21 adults and modified based on 20 cognitive interviews. The domain structure and item concepts of the PKDD and PKDIA were compared with currently recommended measures, the EORTC QLQ-C30 and the SF-36v2®.

Results: The PKDD is a seven-item measure of the core signs and symptoms of PK deficiency. The PKDIA is a 14-item measure of the impacts of PK deficiency on patients' health-related quality of life (HRQoL). Minimal similarities were found between the new measures and the EORTC QLQ-C30 (eg, 43% of concepts were similar to the PKDD; 42% were similar to the PKDIA) and SF-36v2® (57% of concepts were similar to the PKDD; 17% were similar to the PKDIA).

Conclusions: The PKDD and PKDIA fill a gap in the existing outcomes measurement strategy for PK deficiency. Future work includes psychometric evaluation of these newly developed measures.

KEYWORDS

anemia, hemolytic anemia, patient-reported outcome, pyruvate kinase deficiency

1 | INTRODUCTION

Pyruvate kinase (PK) deficiency is an autosomal recessive red blood cell disorder and the most common cause of congenital non-spherocytic hemolytic anemia. PK deficiency causes a glycolytic defect that shortens the red cell lifespan.[1] The true prevalence of PK deficiency is unknown, but it is a rare condition with an estimated diagnosed prevalence of 3.2 to 8.5 per million individuals

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2019 Agios Pharmaceuticals Inc. *European Journal of Haematology* Published by John Wiley & Sons Ltd

in western populations, and a higher frequency found in certain subgroups, including the Pennsylvania Amish, due to a founder effect.[1-5] Patients with PK deficiency are homozygous or compound heterozygous for over 300 identified mutations of the *PKLR* gene.[6] Patients experience a wide range of complications, including iron overload, gallstones, aplastic crises, osteoporosis/bone fragility, extramedullary hematopoiesis, and pulmonary hypertension. Many patients undergo splenectomy to partially improve their hemolysis but this can lead to complications including postsplenectomy infections and thrombosis.[7,8] Patients experience a range of transfusion requirements, with some undergoing intermittent or regular transfusions and with the need for transfusions often increasing when patients experience infections, stress, or pregnancy.[7]

Treatment largely consists of supportive therapy; the only curative treatment is hematopoietic allogeneic stem cell transplantation (HSCT), which for this condition has only been reported in a small number of patients.[1,8,9]

Given the rarity of PK deficiency and the difficulty in finding patients to participate in large-scale research, there is limited insight into the symptoms and impacts of the disease on patients. Patient-reported outcome (PRO) data indicate a burdensome condition where notable signs (eg, jaundice) and symptoms, particularly energy-related concepts (eg, tiredness), cause a wide range of impacts on health-related quality of life (HRQoL).[10] An understanding of these issues is important for optimal disease management and for determining how to measure the effects of interventions on the HRQoL of patients with PK deficiency.

A published evaluation of existing PRO measures appropriate for use in clinical trials in PK deficiency recommended the European Organisation for Research and Treatment of Cancer Quality-of-life Questionnaire Core 30 (EORTC QLQ-C30) and Short Form 36-item Health Survey Version 2 (SF-36v2®).[11] As the SF-36v2® is a generic measure, and the EORTC QLQ-C30 was developed for patients receiving treatment for cancer, there may be symptoms and impacts relevant to PK deficiency that are not covered in these instruments. Furthermore, the EORTC QLQ-C30 may include symptoms and impacts that are not applicable to patients with PK deficiency.

Given the shortcomings of current approaches to measuring outcomes in PK deficiency, disease-specific tools could provide a more robust assessment of the symptoms and impacts that are important to patients. The aims of this research were to (a) describe the development of the patient-reported PK deficiency daily diary (ie, symptom assessment) and the PK deficiency-specific impact assessment and (b) compare the new instruments to the existing measures currently recommended for use in this disease.

2 | METHODS

The PK Deficiency Diary (PKDD) and PK Deficiency Impact Assessment (PKDIA) have been developed in accordance with the United States (US) Food and Drug Administration's (FDA) PRO guidance (Figure 1).

Novelty Statements

- The new aspect of this work is the development of de novo patient-reported outcome measures of symptoms and impacts in PK deficiency.
- The central finding of this work is that the de novo disease-specific measures are more appropriate for assessing symptoms and impacts of PK deficiency than generic quality of life measures (EORTC QLQ-C30 and SF-36v2®) currently recommended for use in this patient population.
- The specific clinical relevance of this work could be improved assessment of the symptoms and impacts of PK deficiency.

2.1 | Literature review and conceptual framework

A targeted review of literature and other materials (eg, market research and feedback from patient advisory boards) compiled by the study sponsor was conducted to inform the development of a hypothesized conceptual framework of signs, symptoms, and impacts commonly experienced by patients with PK deficiency.[8]

2.2 | Concept elicitation interviews and item generation

The preliminary conceptual framework development helped to inform the direction and content of interviews with patients. Participants in the United States were recruited by investigators distributing a recruitment flyer to participants in a Pyruvate Kinase Natural History Study (NHS), through a patient advisory board, a patient-led website for PK deficiency, and through a social media support group; interested participants reached out to be screened for eligibility and to schedule an interview. Participants in Germany and the Netherlands were contacted, screened, and consented by local NHS investigators. Interviews were 60 minutes and conducted in person or via telephone using a semi-structured interview guide in the participants' native language. Interviews were audio recorded and transcribed, except in the case of interviews with Amish participants where cultural considerations only allowed for a notetaker. Additional information such as eligibility criteria and the process for qualitative data analysis has been described more specifically elsewhere.[10]

Following analysis of the concept elicitation interviews, an item generation meeting was held with clinical and PRO experts to reach consensus on the overall preliminary structure and format of the instruments, selection of concepts, provisional item wording, and

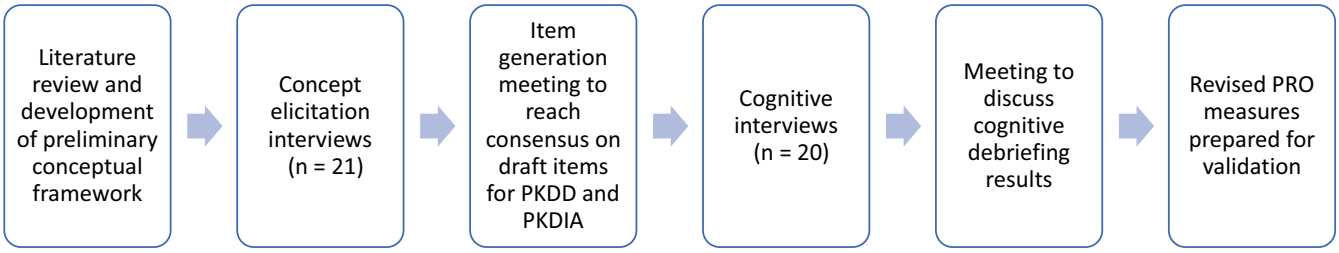

FIGURE 1 Overview of the process for developing the PKDD and PKDIA [Colour figure can be viewed at wileyonlinelibrary.com]

agreement on the final list of items and corresponding instructions. Determination of concepts included in the PKDD and PKDIA was primarily guided by frequency of report, spontaneity of response during the interviews (ie, reporting a concept without probing or prompting from the interviewer), average bothersome (for both signs/symptoms and impacts) and severity ratings for signs/symptoms only (when available) from subjects, and input from clinical experts. Clinical experts reviewed the results of patient interviews and convened in person or via teleconference to discuss and provide counsel on item drafting and revision based on their experience treating this patient population.

2.3 | Cognitive interviews and item revision

Following item generation, cognitive interviews were conducted with participants in the same manner as concept elicitation interviews to establish evidence of face and content validity of the draft instruments to better understand the relevance, language clarity, and ease of understanding of the items.[12,13] Participants were asked to complete the draft PKDD and PKDIA in a think-aloud method, whereby they were encouraged to verbalize their thoughts while completing the instruments.[14] The interview guide specifically probed on the clarity, interpretability, relevance, comprehensiveness, and appropriateness of the items, response scales, and recall periods used in the draft measures.[12,13]

Lastly, the updated PKDD and PKDIA measures were presented to the FDA for review and comment as part of a larger Investigational New Drug (IND) application. Further updates were made to the measures to reflect feedback received from the FDA.

2.4 | Comparison to EORTC QLQ-C30 and SF-36v2®

The items in the PKDD and PKDIA were then compared with the domain structure and item concepts included in the EORTC QLQ-C30 and SF-36v2® to determine the degree of conceptual overlap and differences between the newly developed measures and recommended existing measures.[15] Co-authors reviewed copies of the measures and compared specific attributes, including face validity (ie, conceptual coverage and inclusion of proximal symptoms and/or impacts) and measurement characteristics (ie, item wording, recall, and response options).

3 | RESULTS

3.1 | Literature review and hypothesized conceptual framework

The signs and symptoms framework included four hypothesized domains: anemia symptoms (weakness, dizziness/fainting, lack of/low energy, fatigue, tiredness, exhaustion, lethargy, loss of appetite, shortness of breath), appearance signs (jaundice, pallor), gallstone symptoms (abdominal pain), and other signs and symptoms (body pain, bone pain).

The impacts framework included eight hypothesized domains: activities of daily living (difficulty caring for family, changing daily routine to avoid infection or over-exertion, difficulty driving, difficulty performing household activities), appearance (concealing appearance/skin tone), cognition (difficulty concentrating, memory loss), emotional (angry/frustrated, depressed, self-conscious, low self-esteem, fear of disease progression, feeling guilty), social (avoid/limit activities with family/friends, negatively impacted relationships with family/friends, social isolation), physical (unable to exercise or play sports), work/school (missed days of work/school, negatively affects performance), and burden of disease/treatment (inconvenience of transfusions, time spent in hospital).

3.2 | Concept elicitation interviews

Twenty-one participants were interviewed for the concept elicitation phase. Detailed demographic and health information is included in Table 1.

Thirty-eight signs and symptoms and 59 unique impacts of PK deficiency emerged and were divided into nine categories (ie, activities of daily living, appearance, cognitive, emotional, leisure activity, physical, sleep, social, and work or school). A more in-depth description of the most frequently reported signs, symptoms and impacts has previously been published.[10]

3.3 | Cognitive Interviews

A total of 20 participants were interviewed for the cognitive debriefing phase; 18 (90.0%) had previously participated in the concept elicitation phase.

TABLE 1 Demographic and health summary of interview participants

Characteristic	Concept elicitation participants (N = 21)	Cognitive interview participants (N = 20)
Country of residence		
United States	10 (47.6%)	10 (50.0%)
Netherlands	7 (33.3%)	7 (35.0%)
Germany	4 (19.0%)	3 (15.0%)
Age (y)		
Mean (standard deviation)	38.9 (11.8)	43.3 (13.6)
Min-Max	19.4-58.4	21-78
Gender		
Female	11 (52.4%)	11 (55.0%)
Male	10 (47.6%)	9 (45.0%)
Race		
Data not collected[a]	11 (52.4%)	10 (50.0%)
White	10 (47.6%)	10 (50.0%)
Ethnicity		
Data not collected[a]	11 (52.4%)	10 (50.0%)
Not Hispanic/Latino	10 (47.6%)	10 (50.0%)
Community		
Data not collected[a]	11 (52.4%)	10 (50.0%)
Not Amish	6 (28.6%)	4 (20.0%)
Amish	4 (19.0%)	5 (25.0%)
Data missing	0 (0.0%)	1 (5.0%)[b]
Highest level of education		
Currently in high school	1 (4.8%)	0 (0.0%)
High school (no degree) or less[c]	5 (23.8%)	5 (25.0%)
High school graduate (or equivalent)[f]	2 (9.5%)	4 (20.0%)
Some college (no degree)	3 (14.3%)	1 (5.0%)
Associate's degree	2 (9.5%)	0 (0.0%)
Bachelor's degree[e]	3 (14.3%)	5 (25.0%)
Master's degree[d]	4 (19.0%)	4 (20.0%)
Professional degree	1 (4.8%)	0 (0.0%)
Doctoral degree	0 (0.0%)	1 (5.0%)
Work status[g]		
Working full-time	9 (42.9%)	9 (45.0%)
Student[h]	4 (19.0%)	3 (15.0%)
Homemaker	4 (19.0%)	3 (15.0%)
Working part-time	2 (9.5%)	4 (20.0%)
On disability	1 (4.8%)	1 (5.0%)
Other	1 (4.8%)[i]	1 (5.0%)[j]
Splenectomy status		
Splenectomized[i]	18 (85.7%)	18 (90.0%)
Not splenectomized[l]	3 (14.3%)	2 (10.0%)
Transfusion status[k]		
Transfusion independent	16 (76.2%)	12 (60.0%)
Transfusion dependent	5 (23.8%)	8 (40.0%)

[a]Race/ethnicity data not collected in certain countries due to local privacy laws or because it was not relevant (ie, Amish status).

[b]One participant felt uncomfortable providing this information and declined to answer.

[c]In the Netherlands, this is equivalent to lower or pre-vocational education (standard education until the age of 12-16).

[d]In the Netherlands, this includes a Master's degree and higher (age ≥ 18).

[e]In the Netherlands, this is equivalent to higher vocational education (age ≥ 18).

[f]In the Netherlands, this is equivalent to secondary vocational education (ages 14-18) or higher secondary education (ages 12-18).

[g]Participants could select more than one work status.

[h]In Germany, this option was also inclusive of scholar, visiting a professional school, education and training.

[i]Other response was 'medical leave'.

[j]Other response was 'self-employed'.

[k]Participant did not routinely require red blood cell (RBC) transfusions, but may occasionally require transfusion(s) for anemia as a result of a medical event (eg, viral infection, pregnancy). Typically defined as receiving ≤ 3 RBC units over prior 12 months.

[l]Participant required ongoing regular (or fairly regular) RBC transfusions to manage anemia, typically defined as receiving >4-5 transfusions within a 12-month period.

The results of the cognitive interviews revealed that both the PKDD and PKDIA were well understood by study participants, with all participants (n = 20, 100.0%) demonstrating the ability to understand the instruments' instructions. For the PKDD, 75.0% (n = 15) interpreted each item as intended. The most notable issues related to interpretation pertained to relevance (ie, concept was never experienced) and attribution (ie, participant attributed the symptom to something else unrelated to PK deficiency).

To address issues with relevance, the 'difficulty concentrating' item was removed from the PKDD and added to the PKDIA. The items for bone pain and paleness were found to be less relevant and experienced by fewer participants than other concepts, and based on expert feedback that pale skin was not a hallmark sign of PK deficiency, only bone pain was retained for further exploration during planned psychometric validation.

The participants did not consistently understand the item on need for additional rest or sleep. This item was clarified to 'describe how much additional rest or sleep you feel you needed'. In addition, the item was adjusted to a five-point scale ranging from 'no additional

rest or sleep' to 'a lot of additional rest or sleep' to better direct participants to consider only how much additional rest or sleep they felt they needed. Furthermore, this item was moved to the PKDIA.

Study participants understood and appropriately used the recall period, 'over the past day (from the time you woke up to the time you are completing this questionnaire)'; however, the recall period was changed to 'today', as 'today' is more appropriate for use in clinical trials and for international translation.

Based on feedback from the FDA, an item measuring overall tiredness at its worst was also included, given the pervasiveness of this concept during concept elicitation interviews. The FDA also suggested measuring jaundice using a 5-point verbal descriptor severity scale. Lastly, based on FDA feedback, items measuring the concepts 'difficulty starting things you wanted to get done' and 'difficulty finishing things you wanted to get done' were moved from the PKDD to the PKDIA, given they are better suited for measurement as impacts of low energy levels vs proximal symptoms. Thus, the second version of the PKDD, following cognitive debriefing consisted of seven items measuring seven concepts.

With the PKDIA, 85.0% of participants (n = 17) interpreted each item as intended. The only notable interpretation issue pertained to the item on unwanted attention, which was addressed by relocating the item to appear after the 'bothered by appearance' item to better suggest to subjects that they consider their appearance when thinking about receiving unwanted attention from others.

The remaining issue requiring modification was related to the response options for the work/school skip pattern item (ie, an item where the response triggers an additional question about work/school). To help clarify, the phrase 'for reasons unrelated to my PK deficiency' was added to the former response option. Thus, the second version of the PKDIA, following cognitive debriefing, consists of 14 items measuring 12 concepts. Two of the 14 items (ie, items 9a and 11a) serve as skip patterns for whether or not the respondent is asked about the impact on physical activity and school.

The conceptual frameworks for the PKDD and PKDIA were revised after the cognitive interviews (Figure 2). The revised conceptual frameworks include three hypothesized sign and symptom domains, as the gallstone symptoms domain included in the original hypothesized conceptual framework was not supported by patient interviews or clinical input, and seven hypothesized impact domains, as concepts included in the original domains for emotional impacts and burden of disease/treatment were not frequently reported or applicable to participants in subsequent interviews.

3.4 | Comparison to EORTC QLQ-C30 and SF-36v2®

The EORTC-QLQ-C30 contains nine multi-item scales and six single-item measures designed to assess HRQoL in cancer patients.[16,17] Of the seven concepts in the PKDD, 3 (43%) were common, and one was related but not an exact match (ie, bone pain) to the EORTC QLQ-C30. Of the 12 distinct concepts in the PKDIA, 5 (42%) were

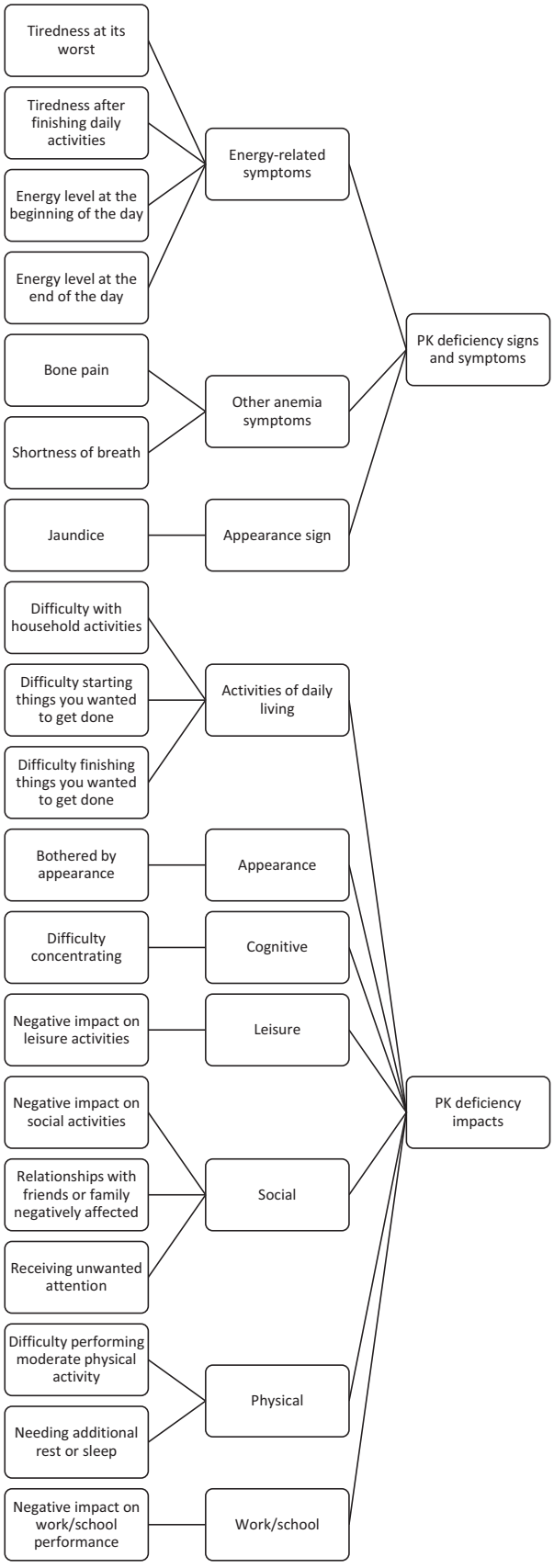

FIGURE 2 Revised conceptual framework for the PKDD and PKDIA based on the results of cognitive interviews

TABLE 2 Comparison of conceptual coverage of PKDD and PKDIA to EORTC QLQ-C30 and SF-36v2®

Measure	Domain	Concept	Included in EORTC QLQ-C30	Included in SF-36v2®
PKDD	Energy-related symptoms	Tiredness at its worst	Yes	Yes
		Tired after finishing daily activities	Yes	Yes
		Energy level at beginning of the day	No	Yes
		Energy level at end of the day	No	Yes
	Other anemia symptoms	Bone pain	Related concept	Related concept
		Shortness of breath	Yes	No
	Appearance sign	Jaundice	No	No
PKDIA	Activities of daily living	Household activities	Yes	No
		Starting things you wanted to get done	No	Related concept
		Finishing things you wanted to get done	No	Related concept
	Appearance	Bothered by appearance	No	No
	Cognitive	Difficulty concentrating	No	No
	Leisure	Negative impact on leisure activities	No	No
	Social	Negative impact on social activities	Yes	Yes
		Relationships with friends or family negatively affected	Yes	No
		Receiving unwanted attention	No	No
	Physical	Difficulty performing moderate (eg, walking on an incline or up stairs) physical activity	Related concept	Related concept
		Needing additional rest or sleep	Yes	No
	Work/school	Work/school performance	Yes	Yes

common to the EORTC QLQ-C30, and one was related but not an exact match (ie, difficulty performing moderate physical activity) (Table 2). Further, the EORTC QLQ-C30 uses both a different recall period (ie, 'during the past week') and response scale (ie, a four-point verbal descriptor response scale) than the PKDD and PKDIA.

The SF-36v2® contains eight multi-item scales intended to assess health status in any population.[18] Of the seven concepts in the PKDD, 4 (57%) were common to the SF-36v2® and one was related but not an exact match (ie, bone pain). Of the 12 distinct concepts in the PKDIA, 2 (17%) were common to the SF-36v2®, and three were related but not an exact match (ie, starting things you wanted to get done, finishing things you wanted to get done, and difficulty performing moderate physical activity) (Table 2). Further, as expected with generic HRQoL instruments, several SF-36v2® domains (ie, social functioning and emotional) ask the respondent to broadly consider their physical health or emotional problems when responding, while the PKDD or PKDIA ask the respondent to consider signs, symptoms, and impacts in the context of or due to their PK deficiency, as they are intended to be disease-specific. The SF-36v2® uses both a different recall period (ie, 'the past 4 weeks') and response scale (ie, a five-point verbal descriptor response scale) than the PKDD and PKDIA.

4 | DISCUSSION

Valid, reliable, and responsive tools are needed to track the issues that patients identify as important in routine clinical care and in clinical trials.[10,11] Overall, the core signs, symptoms, and impacts identified through this qualitative research are in line with those included in the hypothesized conceptual framework generated from a literature review, as well as a previous physio-psychosocial model for PK deficiency developed during an evaluation of PRO measures appropriate for use in clinical trials in PK deficiency.[11] These concepts are also supported by recently published results of a NHS in PK deficiency.[7] The newly developed PKDD and PKDIA instruments reflect these signs, symptoms, and impacts and can be considered novel assessments with appropriate patient-centric development history per the 2009 FDA PRO Guidance.[19] While the European Medicines Agency (EMA) has not issued specific guidelines for PRO development, these measures are consistent with the EMA reflection paper on HRQoL assessment, as well as the role of HRQoL data in drug approval and labeling claims in the United States and Europe.[20-22]

The conceptual coverage of existing recommended measures was previously evaluated against a physio-psychosocial model for PK deficiency.[11] While the EORTC-QLQ-C30 and SF-36v2® were identified as potential candidates for use in PK deficiency trials based on these criteria, the physio-psychosocial model used for comparison does not reflect a patient-centric approach. As such, a major limitation of the recommended use of existing measures, as noted by the authors of that research, is that it is not possible to determine which symptoms or impacts seen in other red cell disorders may actually be experienced by patients with PK deficiency without conducting qualitative research within the PK deficiency population. This research confirms that the EORTC-QLQ-C30 and SF-36v2®

lack the appropriate conceptual coverage of disease-specific signs, symptoms, and impacts most relevant and burdensome to patients with PK deficiency. The EORTC-QLQ-C30 and SF-36v2® may underestimate or misrepresent the burden of disease in PK deficiency, as both lack concepts relevant to and include concepts irrelevant to patients with PK deficiency.

While the EORTC-QLQ-C30 and SF-36v2® include domains that assess overall energy, data from the participants in this research support the inclusion of multiple items in the PKDD designed to distinguish between conceptually different aspects of daily energy levels: PK deficiency patients described tiredness (feeling increasingly tired from daily tasks and activities) and fatigue (feeling chronically tired even after sleeping) as distinct concepts. Given this distinction, the PKDD may be more responsive to detect specific variations in energy levels seen in the PK deficiency population. In addition, since the EORTC-QLQ-C30 was designed for patients receiving treatment for cancer, it includes items such as vomiting and nausea, among others that are not relevant to patients with PK deficiency. Also noticeable is the absence from the EORTC-QLQ-C30 and SF-36v2®of an item assessing jaundice, and the absence from the SF-36v2® of an item assessing shortness of breath. Since jaundice in PK deficiency is caused by increased bilirubin from hemolysis of red blood cells, jaundice could act as a key indicator in evaluating the efficacy of novel pharmacological interventions which decrease hemolysis, and thus should be represented in PK deficiency patient-reported assessments.

Common practice in the field of outcomes assessment suggests that signs and symptoms are best assessed on a daily basis in order to minimize recall bias that occurs with longer recall periods and to ensure a complete understanding of symptom presentation is collected, as symptoms may vary day-to-day. [23] As such, the recall period of 'today' is preferable over 'during the past week' and 'the past 4 weeks', as in the EORTC-QLQ-C30 and SF-36v2®, respectively. Furthermore, accurate daily recall of symptoms is best performed by collecting the worst score for a symptom during a 24-hour recall period.[19] Symptom items in the PKDD have adopted this approach, whereas the EORTC-QLQ-C30 and SF-36v2® do not. Research in several therapeutic areas also supports the use of an 11-point numerical ratings scale (NRS) as the preferred method of assessing symptom severity. [24] There is potential for increased responsiveness on this type of scale that may be more capable of capturing changes over time in a population that has learned to live with and adapt to signs and symptoms associated with their chronic condition.[25]

A potential weakness of this research is that many of the same individuals participated in the concept elicitation and cognitive debriefing stages of this research. Typical practice in instrument development is to have separate populations for the two stages of research so as to avoid any bias that may occur by patients debriefing measures they contributed to developing. However, recent guidance from the ISPOR task force on COAs in rare disease clinical trials does allow for the use of a single population in concept elicitation and cognitive interviews as a pragmatic solution when dealing with rare diseases where patient recruitment is challenging.[26] In addition, the time period between the completion of the concept elicitation and cognitive interviews ranged from 6 to 12 months: A relatively long time period which would tend to minimize any memory of responses from the concept elicitation phase when completing the new measures at the cognitive debriefing phase.

A benefit of the PKDD and PKDIA assessments is the separation of proximal signs and symptoms and distal impacts between two distinct instruments. The EORTC-QLQ-C30 allows for scoring of individual functional and symptom scales, and the SF-36v2® allows the generation of both domain and two overall scores. However, even within these domains, the conceptual mixing of proximal symptoms and impacts is apparent, which may impact measure responsiveness. It should be noted that the expectation is that the PKDD and PKDIA will be administered together, rather than in isolation, to provide a comprehensive assessment of the disease experience. Although the psychometric properties have not yet been evaluated for the PKDD and PKDIA and thus appropriate scoring algorithms have yet to be generated, the conceptual grouping of only specific signs and symptoms included in the PKDD, and only impacts included in the PKDIA, allows for increased flexibility when generating the scoring algorithm of symptoms and impacts.

5 | CONCLUSIONS

This research was used to develop new disease-specific measures: the PKDD and PKDIA consistent with FDA Guidance and EMA literature. A comparison of these measures with existing measures currently recommended for use in this area demonstrates that the PKDD and PKDIA are more relevant and specific to the PK deficiency patient population and may better measure the burden of disease and effect of therapeutic interventions. Planned future work includes the assessment of the psychometric properties of these measures.

ACKNOWLEDGMENTS

This study was funded by Agios Pharmaceuticals Inc Siobhan McDonold, Senior Research Coordinator, Endpoint Outcomes managed recruitment of US participants across NHS sites

ORCID

Shayna Egan https://orcid.org/0000-0003-3170-6682

REFERENCES

1. Grace RF, Zanella A, Neufeld EJ, et al. Erythrocyte pyruvate kinase deficiency: 2015 status report. *Am J Hematol*. 2015;90(9):825-830.
2. Beutler E, Gelbart T. Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population. *Blood*. 2000;95(11):3585-3588.
3. Zanella A, Fermo E, Bianchi P, Chiarelli LR, Valentini G. Pyruvate kinase deficiency: the genotype-phenotype association. *Blood Rev*. 2007;21(4):217-231.
4. Carey PJP, Chandler J, Hendrick A, et al. Prevalence of pyruvate kinase deficiency in a northern European population in the north of England Attention. *Blood*. 2000;96(12):4005-4006.
5. de Medicis E, Ross P, Friedman R, et al. Hereditary nonspherocytic hemolytic anemia due to pyruvate kinase deficiency: a prevalence study in Quebec (Canada). *Hum Hered*. 1992;42(3):179-183.

6. van wijk R. PKLR gene homepage [Internet]. Leiden Open Variation Database: The Human Variome Project. 2019. Available from: Leiden Open Variation Database.
7. Grace RF, Bianchi P, van Beers EJ, et al. The clinical spectrum of pyruvate kinase deficiency: data from the Pyruvate Kinase Deficiency Natural History Study. *Blood*. 2018;131:2183-2192.
8. van Beers EJ, van Straaten S, Morton DH, et al. Prevalence and management of iron overload in pyruvate kinase deficiency: report from the Pyruvate Kinase Deficiency Natural History Study. *Haematologica*. 2018;104(2):e51-e53.
9. Van SS, Bierings M, Bianchi P, et al. Stem cell transplantation in pyruvate kinase deficiency. Present Eur Hematol Assoc Annu Meet 2017 [Internet]. 2017;(S452).
10. Grace RF, Cohen J, Egan S, et al. The burden of disease in pyruvate kinase deficiency: patients' perception of the impact on health-related quality of life. *Eur J Haematol*. 2018;101:758-765.
11. Salek MS, Ionova T, Johns JR, Oliva EN. Appraisal of patient-reported outcome measures in analogous diseases and recommendations for use in phase II and III clinical trials of pyruvate kinase deficiency. *Qual Life Res*. 2019;28(2):399–410.
12. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 2—assessing respondent understanding. *Value Health*. 2011;14(8):978–988.
13. Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q*. 2007;71(2):287-311.
14. Glaser BG, Strauss AL. The discovery of grounded theory. In: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine de Gruyter; 1967: 1–18.
15. Fayers PPM, Aaronson NKN, Bjordal K, Groenvald M, Curran D, Bottomley A. EORTC QLQ-C30 scoring manual (3rd Edition) [Internet]. European Organisation for Research and Treatment of Cancer. EORTC; 2001.
16. Shih C-L, Chen C-H, Sheu C-F, Lang H-C, Hsieh C-L. Validating and improving the reliability of the EORTC QLQ-C30 using a multidimensional Rasch model. *Value Health*. 2013;16(5):848–854.
17. Reeve BB, Mitchell SA, Dueck AC, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *J Natl Cancer Inst*. 2014;106(7):1-8.
18. Ware JE, Jr., Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care*. 1992;20:473-483.
19. Food and Drug Administration. Guidance for industry use in medical product development to support labeling claims guidance for industry. *Fed Regist*. 2009;74:65132-65133.
20. Marquis P, Caron M, Emery M. The role of health-related quality of life data in the drug approval processes in the US and Europe. *Parmaceutical Med*. 2011;25(3):147-160.
21. Demuro C, Clark M, Doward L, Evans E, Mordin M, Gnanasakthy A. Assessment of PRO label claims granted by the FDA as compared to the EMA (2006–2010). *Value Health*. 2013;16(8):1150-1155.
22. European Medicines Agency. *Reflection Paper on the Regulatory Guidance for the Use of Health- Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products*. London: European Medicines Agency; 2005.
23. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Curr Med Res Opin*. 2009;25(4):929-942.
24. Gries K, Berry P, Harrington M, et al. Literature review to assemble the evidence for response scales used in patient-reported outcome measures. *J Patient Rep Outcomes*. 2018;2(1):41.
25. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J Clin Nurs*. 2005;14(7):798-804.
26. Benjamin K, Vernon MK, Patrick DL, Perfetto E, Nestler-Parr S, Burke L. Patient-reported outcome and observer-reported outcome assessment in rare disease clinical trials: an ISPOR COA emerging good practices task force report. *Value Health*. 2017;20(7):838-855.

How to cite this article: Salek S, Boscoe AN, Piantedosi S, et al. Development of the pyruvate kinase deficiency diary and pyruvate kinase deficiency impact assessment: Disease-specific assessments. *Eur J Haematol*. 2020;104:427–434. https://doi.org/10.1111/ejh.13376